PALATALIZATION IN AUSTIN.
A sociophonetic analysis of sibilants.

By

Wiebke Ahlers

Dissertation

zur Erlangung des Doktorgrades (Dr. phil.)

des Fachbereichs Sprach- und Literaturwissenschaft

der Universität Osnabrück

Osnabrück, März 2020

Deutsche Zusammenfassung

Die vorliegende Arbeit analysiert basierend auf 80 Aufnahmen von Sprechern in Austin den derzeitigen Status und inter-generationellen Verlauf eines Lautwandels, der Sibilanten in bestimmten linguistischen Kontexten betrifft. Sowohl sprach-interne als auch sprach-externe Variablen werden beleuchtet und mit unterschiedlichen statistischen Berechnungen evaluiert.

Bei dem Lautwandel handelt es sich um die Retraktion der Zungenposition in der Aussprache von /s/ im Konsonantenkluster /str/. Diese veränderte Zungenposition hat zur Folge, dass der produzierte Laut eher dem akustischen Frequenzprofil des postalveolaren //ʃ/ entspricht als dem des phonotaktisch alveolaren Sibilanten.

Zur Untersuchung dieses Lautwandels wurden 86 Teilnehmer bei verschiedenen Sprachaktivitäten aufgenommen. Zunächst absolvierten diese eine Bildbenennungsaufgabe, gefolgt vom Vorlesen einer Lotta Geschichte von Astrid Lindgren, der Handlungsnacherzählung selbiger und letztlich einer nachträglichen Aufklärung über die Inhalte der Studie. Auf diese Aufklärung folgte letztlich ein freier Interviewteil, bei dem die Teilnehmer gebeten wurden, den hier untersuchten Lautwandel zu kommentieren.

Die jeweiligen Tokens mit Sibilanten wurden dann händisch in Praat transkribiert und mit Hilfe von FAVE den jeweiligen Phonemen zugeordnet. Diese Zuordnung war Grundvoraussetzung für eine automatisierte Messung des Center of Gravity für alle vorkommenden Sibilanten, welches eine akustische Beurteilung des sprecherspezifischen Sibilantenspektrums und somit die Einordnung von /str/ im vollen Spektrum von alveolar zu post-alveolar ermöglicht. Basierend auf dem Spektrum wurden pro Sprecher alle Werte mit Hilfe von z-scores normalisiert um Vergleichbarkeit zu erreichen.

Die Analyse aller Sibilanten zeigt auf, dass die Verteilung von /str/ sich über das komplette Spektrum erstreckt. In eine Analyse mit Fokus auf /str/-Items wurden verschiedene sprachinterne und externe Faktoren gefunden, die einen Einfluss auf die Retraktion des Clusters haben. Sprachintern sind dies einerseits die Dauer des Sibilanten und andererseits Frequenzeffekte, sodass höher-frequente Wörter generell eher post-alveolar ausgesprochen werden. Sprachexterne Faktoren, die in gemischten Regressionsmodellen einen signifikanten Effekt auf den Lautwandel haben sind Ethnizität des Sprechers und Alter. Dies bestätigt die Vermutung, dass es sich um einen Lautwandel handelt, der sich nach jetzigem Stand durchsetzen wird.

Die qualitative Analyse der Sprecherkommentare zeigt auf, dass bisher kein konkretes soziales Profil für diesen Lautwandel besteht. Präskriptivismus ist hier die vorherrschende Erklärung und Perspektive auf das Phänomen.

Zusammenfassend zeigt diese Arbeit eine neue Perspektive auf einen Lautwandel, der in den 80gern erstmals in der Literatur erwähnt wurde. Sie ergänzt den aktuellen Forschungsstand in besonderem Maße, da der Datensatz nicht nur deutlich mehr Items per Sprecher enthält als bisherige Untersuchungen, sondern auch ein diverseres soziales Profil aufweist. Nur so ist es möglich, sowohl Frequenz als auch soziale Diversität dieses Wandels erfolgreich zu eruieren

Acknowledgements

My own path to writing this dissertation started long before I knew I would write it. I think my parents helped me immensely with the first step by seldomly getting tired of all my questions about the world and by allowing me to express my opinions. Although my father never got tired of reminding me that the discussions we were having were often "Grundsatzdiskussionen", I think his encouragement of this behavior was crucial to writing a book in the end. I have my mother to thank for all the social skills that ended up helping me in all of my stays abroad, but especially in being able to engage with participants for this study.

Once I got to the University of Osnabrück, I enjoyed all of my linguistics classes, both in German and in English. And while this was also due to my interest in the topic, it certainly helped to meet Simon, Siebi, Dennis and Roland, who've not only made my first 5 years in Osnabrück very enjoyable, but who've also supported me throughout the second 5 years- when life got real.

I have Alexander Bergs to thank for his support with both my Bachelor's and Master's thesis. He fostered this linguistic curiosity of mine. However, my curiosity would never have extended this far without him specifically encouraging me to pursue a career in academia. I would have never had this courage on my own.

I feel immensely lucky that Fulbright not only chose me as a candidate for their FLTA program but also funded the research that this dissertation is based on. I further thank Lars Hinrichs for making my research stay in Austin possible in the first place. I will be forever indebted to the many kind souls that supported me in community centers in Austin.

Academic courage was something I've had to work hard for and still am working on today. When I started out, I was intimidated by academia and the two people who've helped me cope with it most are Axel Bohmann and Phillip Meer. I greatly benefitted from working on our projects together. This thesis would not be the same without their continued comments and support.

I also want to thank Anke Dittmer, who helped me to fit all the puzzle pieces of data wrangling and measurements I needed to perform in one coherent script.

Last but not least, I need to thank the friends I've met after university. I am amazed at how Shelby covered the distance and supported me throughout the past 6 years. She kept me sane throughout this project. Dodo has made sure that I was able to go through all the ups and downs from a place that is a home- it's made all the difference.

Table of Contents

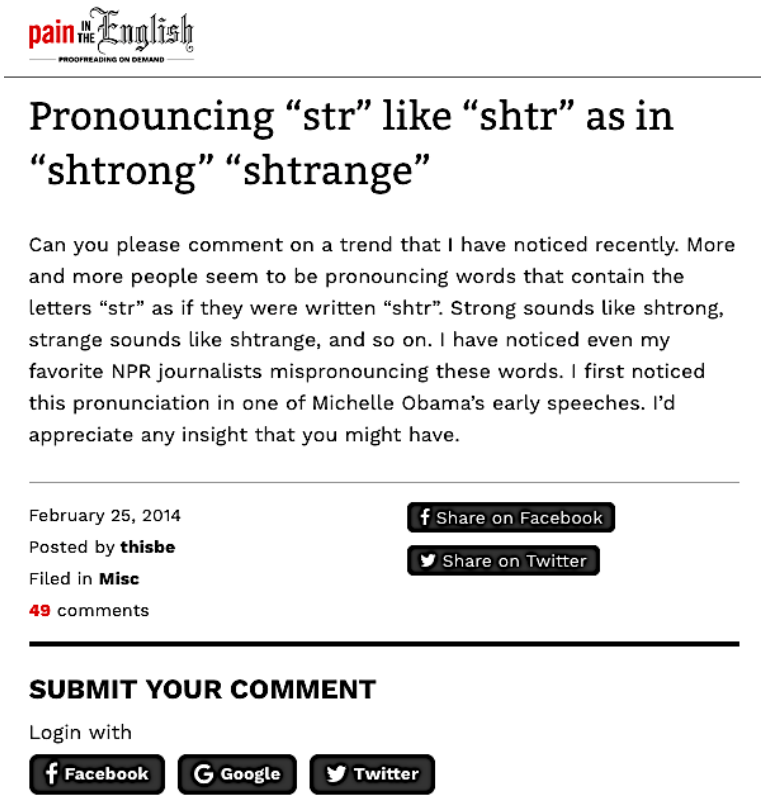
List of Figures

List of Tables

1.Introduction

One man's trash is another man's treasure- or in this case, what a community may consider one of the biggest flaws in the current status of their language is usually a sociolinguist's treasure. This is the case for a phenomenon called /s/-retraction, which a user on "pain in the English" describes in Figure 1.1.

Figure 1.1: Screenshot of forum post on painintheenglish.com



The process described here in folk linguistic terms is the change from the alveolar production of /s/ in sibilant clusters to a backer version that resembles the acoustics of the post-alveolar sibilant. Orthographically, this difference is described as a difference between "str" and "shtr". The author of the comment utters a deep concern that their "favorite NPR journalists" are now "mispronouncing" words such as STRONG and STRANGE. Even the first lady in 2014 is mentioned as an example for this pattern.

This trend is cause for debate not only in this online forum, but has also been negatively evaluated on many other online platforms. In 2014, a contestant was mocked in the British show X Factor for singing the song *Streetlife* by Randy Crawford with post-alveolar production in that first sibilant. After she sings just a few lines of the song lyrics, juror Simon Cowell stops her: "Can I say something? It's straight street. It's not shtreet[1]." (The X Factor UK: 0:14). She replies surprised: "Am I saying shtreet? Oh no let me fix it then" (0:20). She starts singing again, once again producing the retracted version of the

[1] (sh added to emphasize post-alveolar pronunciation)

sibilant, with all jurors immediately starting to laugh. Cowell asks: "Do you know shtop?" (1:00). Her enunciation is addressed further, and though they like her voice, Cowell states "we have a problem with your shshs" (1:07).

Both her inability to produce anything other than the retracted sibilant and the juror's reaction illustrate a variation in many varieties of English that linguists and non-linguists alike have noticed in the last few decades: The rise of a more retracted version of the alveolar sibilant wherever it precedes plosives. Most often, this version is discussed for the sibilant followed by /t/ and /r/, the exact /str/-combination[2] that the contestant on X Factor struggles with in *Streetlife*. From a perception point of view this could be described orthographically as a change from the "regular" pronunciation of the word STREET (with the first phoneme identical to the first phoneme of SEA) to a version that presents with a retracted sibilant orthographically transcribed as SHTREET. This change from a "hissing" to a "hushing sibilant […] with extreme rounding" (Labov 1981, 30) was first noticed by Labov, but has since been examined in many regions in the United States (Durian 2007; Rutter 2011; Gylfadottir 2015; Wilbanks 2017; Phillips 2018; Stuart-Smith, Sonderegger, Macdonald, et al. 2018) and beyond (Nichols & Bailey 2018; Stuart-Smith, Sonderegger, Macdonald, et al. 2018). Although scientific interest in this phenomenon is on the rise and researchers have investigated it in many varieties of English, the sound change is still considered under-researched in numerous ways.

There is discussion as to the phonological basis of this sound change concerning the status of assimilation (to an affricated version of /t/ or "at a distance" to /r/), the involvement of reanalysis in its progression as well as its context structure within the word (Shapiro 1995; Lawrence 2000; Janda & Joseph 2003a; Gylfadottir 2015; Stevens & Harrington 2016; Wilbanks 2017). Both theories of assimilation are evident in the X Factor video described above; the singer only retracts the cluster, but the judge extends the environment to *shtop*, which would indicate an assimilation to /t/.

Phonetically, the sibilant is described as displaying "a peculiar acoustic character" (Shapiro 1995). This character refers to its position on an imagined continuum between alveolar and post-alveolar sibilant production (Baker, Archangeli & Mielke 2011; Gylfadottir 2015). In an attempt to understand this peculiarity, the retracted sibilant's frequency characteristics have been analyzed acoustically both in the durational center of the sibilant (Gylfadottir 2015; Wilbanks 2017; Stuart-Smith, Sonderegger, Macdonald, et al. 2018) and for the entire duration of the sibilant (Stevens & Harrington 2016; Phillips 2018).

[2]There are several ways to transcribe the variable production of the consonant cluster that this dissertation discusses. The most commonly used transcription is the phonemic transcription, which is in line with the understanding that the phonotactic expectation and mental representation of the phoneme in this cluster is generally /s/. Since gradualness is expected in this change, /ʃtr/ would overestimate the current status. Another option would be to differentiate between the instances where I address the phonemic cluster /str/ and those where I address the sociolinguistic variable (str). In order to provide a more coherent reading experience, I side with Stevens & Harrington and Rutter, who do not make this distinction (Rutter 2011; Stevens & Harrington 2016)(Rutter 2014; Stevens & Harrington 2016)

From a sociolinguistic perspective, little other than a preference of retracted /str/ by younger speakers and thus an indication of change in apparent time has shown a correlation with this variant so far. While there are some studies on the effects of auditory perception (Scudieri 2012; Stevens & Harrington 2016), the general notion that it is a change below the level of consciousness can only be guessed from the scarcity of non-linguistic comments on it. The video described above could be interpreted in both directions. The contestant herself is oblivious to her pronunciation and even after being mocked she is unable to produce the alveolar sibilant in this position. However, Cowell is entirely aware of this pronunciation and able to perform both versions.

The present work empirically investigates all of the open questions sketched above. The research questions will be introduced in greater detail in Chapter 4, after I describe the state of the art on /s/-retraction. More broadly speaking, this dissertation uses interview data from Austin to provide an entirely new angle on whether this non-regional sound change exists in Austin in apparent time. In addition to the evaluation of birth year and its effect on the production of /str/-clusters, I will reexamine and expand previous findings within both sociolinguistic and phonological research. The main focus lies on understanding the extent to which the present study can shed light on the social characteristics of this pronunciation variant, more specifically its ethnicity and gender profile. Before these finer grained social factors can be meaningfully evaluated, the language-internal basis of the process needs further elaboration. Previous research has provided strong evidence for a change in apparent time (Baker, Archangeli & Mielke 2011; Gylfadottir 2015; Wilbanks 2017; Stuart-Smith, Sonderegger, Mcauliffe, et al. 2018). However, due to the limited amount of naturally occurring tokens of /str/, previous research has worked with different datasets both in terms of the recency of recordings and the amount of items per speaker. The present study is one of the first to acoustically evaluate the sound change in an experimental framework designed to elicit /str/-clusters. It is therefore suited to further describe the nature of the change and extend the profile by adding intra-speaker variation. One facet of this description is evaluating gradualness in its progression. Lexical gradualness would indicate a change through lexical diffusion, while a lack of word frequency effects would further strengthen the notion of assimilation or coarticulation as a language internal factor in this change. The phonetic gradience of the sibilant will be elaborated to evaluate the "peculiar acoustic character" further. Lastly, the analysis of meta interviews will provide a qualitative first approach to the sociolinguistic salience of /s/-retraction as indicated in both the blog post and video outlined above.

The sound change from alveolar to a more retracted fricative sound in clustered consonant onsets lies at the heart of the present investigation. In their often-cited "Big-Bang Theory" on sound change, Janda and Joseph argue that every one of these changes is trifold, thus having a "sociolinguistic side, a phonetic side and also a phonological side" (Janda & Joseph 2003a: 1). It is evident from the above descriptions that this holds true for /s/-retraction in many ways.

The following chapters will provide the theoretical backdrop for the sound change under investigation in this study in that same trisection. The first chapter introduces the social factors established in the literature as possible factors in sound change. The aim of this chapter is to understand how these social categories were developed theoretically and

how they figured into sociolinguistic analyses. It begins with an introduction of the theoretical framework of variationist sociolinguistics and its understanding of language change. The conception of age that directly links to theories of change and change in apparent time in particular is then portrayed. Descriptions of larger social categories such as gender, ethnicity and education complete the macroscopic lens for the sociolinguistic analysis. Subsequently, the focus is on changes within the individual through a sketch of the development of sociolinguistic theories of style and identity. The chapter ends with a section on social meaning in language, which introduces many of the theories and concepts necessary to hypothesize about the social progression and function of the change.

In the second chapter, phonetic and phonological theories relevant to the core of retraction are elaborated. It starts with a differentiation between the concepts of phoneme cognition and phoneme production. In a second step, cognitive models of sound production and sound perception relevant to the process at hand are explained. Based on both the cognitive and articulatory conceptions of speech production, attention then shifts to the resulting harmonic processes on both the cognitive and articulatory level, by explaining concepts such as assimilation and coarticulation.

Lastly, in Chapter 3, the theory of sound change is introduced, based on the previously elaborated concepts of cognitive and articulatory speech production. Once this theoretical framework has set the stage, the state of the art on this process is elaborated. Both the theories introduced previously and the findings of previous research are then used to illustrate the development of my own hypotheses and research plans.

Chapter 4 introduces in great detail how the theoretical framework and previous research are renegotiated to suit my own research design and procedure.

Chapter 5 presents the findings gathered through these methods. Before regression modeling, trends and patterns in the data are explored visually using tree and forest models to better gauge the direction and future of this change based on additional social factors being modeled in comparison to previous research. Chapter 5 further investigates the patterns present in the meta commentaries of the participants through thematic analysis.

I debate all findings and trends in Chapter 6, with a strong focus on those social patterns of variation that were described differently or narrowly in previous research. Chapter 7 then illustrates the strengths and limitations of this project and sketches pathways both for the ongoing change and future research.

1.1 Sociolinguistics and Language Variation

This chapter provides an introduction to the theoretical foundation of research in sociolinguistics. It is dedicated to the social aspects that directly (factoring into the quantitative analysis) or indirectly (factoring into the qualitative analysis) inform the methodological approach and theoretical understanding of this study. The subchapters are designed specially to introduce the canonical patterns researched in most studies in the variationist framework and to provide a critical depiction of the notions that these factors are based on.

The framework for the present analysis of a possible change in progress in the phonotactics of English rests on the theories of variationist sociolinguistics. There are

several theoretical assumptions necessary to meaningfully research and analyze such variation on the linguistic surface of the communicative code that we call language. The first is the connection between language and the social fabric of life. Sociolinguistics, the resulting field of theoretical inquiry, is relatively young if we consider Currie as one of its "founding fathers", who described the field and its focus in 1952:

> "The present purpose is to suggest, by the citing of selected and salient studies, that social functions and significations of speech factors offer a prolific field for research. It is the intention in this connection to project, partly by means of identification, a field that may well be given the attention of consciously directed research. This field is here designated to *socio-linguistics*. Attention will be called to certain relevant research done or under way. Possibilities for further socio-linguistic research are, in fact, beyond estimation." (Currie 1952: 27)

Although dialectology as a field could generally be considered a predecessor of sociologically informed approaches to language, and sociology in and of itself may present some studies in which language is a factor, the type of data gathered and the interpretation of this data is what distinguishes sociolinguistics as discrete field of research (Chambers 2002: 6). Labov was at the forefront of circumscribing a more nuanced understanding of these "speech factors". When he first introduced his approaches to language to the scientific community, it was not the idea of language factoring greatly in social life that was new. His work was novel because it showed how linguistic variables could be correlated with their social context in a meaningful manner (Labov 1963).

This contrast is exemplified by Chambers' description of the aims of sociolinguistic researchers: "sociolinguists attempt to grasp language as it is used in social situations, which is to say variant, continuous, and quantitative" (Chambers 2002: 8). The quantitative nature of sociolinguistics is best described through a brief introduction of the quantitative paradigm. Bayley names the most influential principles of this approach the "principle of quantitative modeling" and the "principle of multiple causes" (Bayley 2018). The former introduces the general observation of sociolinguistic research that variation in a certain linguistic form under scrutiny may pattern meaningfully in relation to its context. The latter principle takes this variation a step further by arguing that this variation is often caused by more than one context factor.

The relationship between variable and context is important to the understanding of variationist research. Yet, defining the linguistic variable as such may prove difficult and has led to a strong reliance on the principle of accountability. This principle is based on the understanding that research aims "to isolate the largest homogeneous class in which all subclasses vary in the same way" (Labov 1972: 72). The largest homogeneous class in variationist analyses often describes a language internal context in which a pattern can be expected to occur. The subclasses then are the further subgroups within the largest group that was identified. Correctly defining these classes is a precursor for the "principle of accountability" (Labov 1972: 72). A good example for research which fails to meet the principle of accountability in the present case would be an analysis of only highly frequent /str/-items. The largest class is thus not accounted for, because many lexical environments

are left out and findings will be influenced by the focus on only a subclass (frequent words) of the largest class (all words with /str/).

While the discipline's concern with relating language variation with its context and use has remained stable, Eckert perceives changes in the four decades of variationist research in linguistics and names them "the three waves of variation studies" (Eckert 2012). She describes the survey-focused quantitative research introduced above as the first wave of sociolinguistics. She further characterizes: "studies in the first wave interpreted the social significance of variation on the basis of a general understanding of the categories that served to select and classify speakers rather than through direct knowledge of the speakers themselves and their communities." (Eckert 2012: 90). Studies in the second wave change this focus by bringing such concepts as communities and networks into the correlations. This is a shift from perceiving speakers as instances of social category mixes to allowing for agency and locality in the individual. The third wave in sociolinguistic research rests on this foundation while focusing more on the speaker not as a complex pool of network memberships but rather as an active creator of social meaning in language (Eckert 2012).

Eckert's summary of variationist research generally provides a good overview of the various aims that have been included in variationist research. At the same time however, this clear-cut division underestimates the fact that the empirical methods and the social categories employed by Labov are still informing research in sociolinguistics, because there are several ways in which broad social categories still affect speakers lives. Even if we agree that a division amongst gender lines cannot entirely account for the complexity of gendered identities, identifying as female shapes many of the social interactions that define linguistic identity. It is therefore a question of scope rather than a question of evolution to decide why and how to research the connection between a linguistic variable and broad or narrow social factors.

The sociolinguistic scope of research on /str/-retraction is limited compared to other fields, as I will show in Chapter 4. It therefore makes sense to begin by detecting patterns in the broader and canonical social factors prominent in the above described "first wave". Thus, the characteristics of the impact of age, gender, ethnicity, social class and education on language are introduced first. After these inter-speaker categories, I consider the characteristics of the impact of various notions of style and identity on language. This shift changes the focus from per speaker categories to within speaker variation, a shift theoretically informed by concepts in "second wave"-research. Lastly, I introduce meaning in language through the concept of indexicality. While the categories will later be used to statistically establish the social profile of the change, these notions of meaning will aid in making sense of these patterns and connecting them to the metalinguistic knowledge of participants. Before I introduce the social aspects factoring in language variation, I will briefly differentiate between the study of language variation and language change.

1.2 Language Change

Changes in language have generally been investigated from both diachronic and synchronic points of view. These points of view also resulted in two different fields, one that aims to

explain the past and one that wants to predict the future from the present and immediate past. Historical linguistics, on the one hand, has developed various methods to explain language change after the fact, meaning that these linguists observe change from the diachronic perspective. This field developed e.g. the comparative method that is used to analyze various historical stages of the language(s) to describe mechanisms of change and genealogical relations (Rankin 2008). The present study, however, describes a change in the immediate past and thus relies heavily on the synchronic approaches to language change. I will therefore focus on an outline of the development of variationist theories of language change below. Some of the historical notions of change will be referred to later when I introduce the mechanisms of phonological change more specifically.

The variationist understanding of language change is deeply rooted in the theories of Weinreich et al. (Weinreich, Labov & Herzog 1968). In a groundbreaking theoretical paper, they position themselves historically as heirs to theories of language change developed by "moderate" Neogrammarians, Hermann Paul and Ferdinand de Saussure, structuralist Leonard Bloomfield and generativist Noam Chomsky. All of these linguists address the fact that language is changing and that individuals are involved at some level within this process, but they all fail to develop a theory to connect all aspects without negating or excluding certain phenomena. Chomsky is one of the most pronounced opponents to including any type of language other than ideal within linguistic theories:

> "Linguistic theory is concerned with the ideal speaker-listener in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance". (Chomsky 1965: 3)

It is within this theoretical context that Weinreich et al. develop the outline of a theory that can cope with all facets of how and why language changes. In a direct response to Chomsky's understanding of language, they describe it as "an object possessing orderly heterogeneity" (Weinreich, Labov & Herzog 1968: 101). They argue that a perfect structure could and would not change. However, if we accept that the cases of imperfect language we are presented with are not mere chaos, but rather allow for the flexibility necessary to recreate and improve, only then can we apprehend how variable patterns are meaningful. To them, in "a language serving a complex (i.e., real) community, it is absence of structured heterogeneity that would be dysfunctional" (Weinreich, Labov & Herzog 1968: 101). Guy comments that this view of the connection between language variation and language change is one of scope. He states: "variation is the synchronic face of change, and change is nothing more than diachronic variation" (Guy 2011: 179).

Weinreich et al. employ their theoretical understanding of how variation as orderly heterogeneity relates to language change to develop a set of problems that a theory of language change needs to address, if we want to understand how language changes in service to the community. They describe these problems as follows:

1. The constraints problem: "The problem of constraints on immediately succeeding language states […],we too want to inquire into the set of

possible changes and possible conditions for changes which can take place in a structure of a given type".

2. Transition problem: "it remains entirely relevant to ask about intervening stages which can be observed, or which must be posited, between any two forms of a language defined for a language community at different times."
3. Embedding problem: "How are the observed changes embedded in the matrix of linguistic and extralinguistic concomitants of the forms in question? (That is, what other changes are associated with the given changes in a manner that cannot be attributed to chance?)"
4. Evaluation problem: "And how can the observed changes be evaluated—in terms of the effects upon linguistic structure, upon communicative efficiency (as related, e.g., to functional load), and on the wide range of nonrepresentational factors involved in speaking?"
5. Actuation problem: "What factors can account for the actuation of changes? Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times? This actuation problem can be regarded as the very heart of the matter". (Weinreich, Labov & Herzog 1968: 101)

The questions they pose and problems they outline are taken together as a theory of language change able "to deal with nothing less than the manner in which the linguistic structure of a complex community is transformed in the course of time so that, in some sense, both the language and the community remain the same, but the language acquires a different form" (Weinreich, Labov & Herzog 1968: 102).

The presented theory of language change was adapted by many researchers and is arguably the foundation of the variationist linguistic analyses of change. Subsequent research in language change was mainly concerned with further developing the underlying principles of problems 1-3. The constraints principle truly can be considered the basis of the above-mentioned quantitative paradigm. It is the very foundation of considering correlation a meaningful technique in evaluating language in its social context. The evaluation problem is the theoretical backdrop of (metalinguistic) perception and the foundation of the above described second and third wave studies in sociolinguistics. While the actuation problem remains heavily debated, few studies have been able to deliver more than arguments as to why change has a specific timing (Baker, Archangeli & Mielke 2011). In particular, the transition problem leaves room to find more functional and detailed descriptions of patterns differentiating between various types of progression observed in language data. As such, it can be regarded as the foundation of the curvilinear principle as well as theories of lexical diffusion, which will be addressed in greater detail later.

While the five problems outlined in Weinreich et al.'s theory certainly inform the methods and possibilities of describing all factors in a change, they also provide enough room for the meaningful existence of variation. The general notion that all change results from language variation but not all variation causes change in a language was included in the features of change noted by Labov:

1. Linguistic changes show a sporadic character, beginning and ending abruptly at times that are not predicted by any universal principles.
2. Stable, long-term variation that persists over many centuries in much the same form is perhaps even more common than changes which go to completion.
3. It is not uncommon to find retrograde movements, where the direction of changes reverses, or opposing directions of movement in parallel communities. (Labov 2001: 75)

Most notions of change that are of interest in the present work were developed by Labov and closely relate either to the mechanisms of sound changes (mergers, splits, shift) or generalized observation resulting from the plethora of studies carried out by Labov and other researchers. Due to the strong ties with the respective social categories, I will only provide a brief overview here, which will then be complemented by relevant subchapters.

One of the general principles discussed by Labov is "the use of the present to explain the past" (Labov 1994: 1), namely the apparent time construct and the resulting methodological aims (Labov 1972; Labov 1994; Labov 2001). As this rests upon age-related assumptions, it will be introduced as such in the chapter on age. Another general mechanism of language change is described by the notions of change from above and change from below. They deal with the extent to which speakers are aware of certain language features and whether they are used consciously by upper classes (above) or spread through the usage in lower classes without conscious awareness (below) (Labov 1990). Due to their reliance on social stratification, they will be exemplified in the section on social class. These changes are also closely connected to the behavior of females in linguistic change and will therefore reappear in the subchapter on gender.

Language change may also be influenced or motivated by language contact. In the specific North American situation, this contact situation with substrate languages is most meaningful when one considers not necessarily entirely different languages but ethnolinguistic varieties such as African American or Lumbee English (Wolfram 2013). I will briefly describe the theoretical treatment of contact situations in order to better describe the complex concept of ethnicity in the respective subchapter below. Where language contact is at issue, socioeconomic factors often play a role in what the relationship between languages is and how change may come about. Generally, the language with the higher degree of socioeconomic power is referred to as the superstrate while the language of lesser standing is referred to as the substrate language. Language contact can lead to language maintenance, language shift or pidginization (which may later create a creole language) (Thomason & Kaufman 1991).

Power imbalances most often lead to language shift of the substrate language community. Until this assimilation takes place, there is a phase of bilingualism in which speakers may include borrowed terms of the newly learned language into their own language or influence the language they are newly acquiring, which is referred to as substrate influence (Thomason & Kaufman 1991). Thomason & Kaufman mostly describe colonial contact situations in which the incoming language gains power over the local language, leading to language shift. Especially in the recent history of the US however, language contact is taking place through immigration and thus creating a contact situation

where English clearly remains the superstrate language. The English-Spanish contact situation is the most prominent instance of language contact created by immigration, the details of which will be introduced below (Gillian Sankoff 2006).

In contemporary debates on language change, Coupland argues that there is a distinct difference between sociolinguistic and language change that is often not addressed in research. He criticizes: "In the language change paradigm, society is 'where language change happens', but we are often dealing with linguistic changes that have no necessary social significance in themselves" (Coupland 2014). The search for social significance of a change is addressed within the meta commentary section of this study. To what extent the social significance of a sound change can truly be circumscribed, especially when it is as far-reaching as /str/-retraction, is cause for a later discussion.

The present study is a close analysis of a possible language change considering both the problems outlined by Weinreich et al. as well as the Labovian descriptions of language change. Contact is interesting not because we expect contact between languages, but rather because the power imbalances between languages may apply in similar ways to ethnolects such as Chicano and African American English present in Austin. The relevant theories and definitions of phonological analysis are introduced in chapter 2 and detailed theories relating to sound change are elaborated on thereafter.

1.3 Interspeaker Variation and Change

This section continues to set the stage for the social analysis of /str/-retraction by introducing the broader interspeaker and intraspeaker patterns expected in variationist sociolinguistic work. Interspeaker variation concerns the variability of speech when treating each speaker as an individual data point. As such, factors considered within interspeaker variation are often those we employ to evaluate the extralinguistic side of the constraints problem. Whatever their individual patterns are when using the studied feature of speech, they are treated as definable points in the mosaic that is the speech community. Within a community, categorization of social groups is a common process in identity formation. However, there are certain social groups with very generalized membership criteria while others are rather nuanced.

For example, most people identify as a certain gender, belong to a certain age group and can fairly easily be grouped according to their position within the economic and educational system. Similarly, some people may identify as active members of their church community. Yet, while one may generalize findings about the linguistic behavior of women or older speakers in a certain region of the U.S, an active member of a church community in Austin may not show meaningful linguistic patterns when compared to another active member of a different church community in Austin. This is not to say that there may not be community- and network-related language use due to the membership in this group. But it points toward the trends that we can perceive in research on interspeaker variation. The social features that have been involved in many of the statements made about the social structure of language are traditionally age, gender and ethnicity. These are (at least in a western construction of the world) biologically informed and often accepted as given categorizations, because they are visible to some extent.

On the other hand, there are theoretically informed categories, aimed at mapping social positions such as education, social class or dialect region. Although it may intuitively seem as if education or social class are easily distinguishable, categorizations in these realms are often highly artificial or complex. Each of these social factors in language variation are illustrated in the following subchapters. While each of the following sections deal with these categories individually, in reality, they may only become meaningful in combination. Many variationist studies have shown how speakers from the same social class may be different from other social classes but further vary based on their gender and community profile (Milroy & Milroy 1993). All of these social factors will be analyzed in the present study with varying degrees of correlation. Below, I outline these categories and connect them to theories of language change where possible.

1.3.1 Age

Age is central to most sociolinguistic analyses for two reasons. On the one hand, much like gender, age group membership is perceived as a central defining factor in the understanding of one's identity. On the other hand, our theories both of how the perception of one's age proceeds across the lifespan as well as of the aging brain are central to the methods with which we analyze speech patterns and especially language change.

First, we assume that age is of such profound perceptual similarities across western societies that everyone experiences the process similarly and may therefore pattern meaningfully in terms of linguistic variability. In Chambers' understanding, age is "an immutable social fact" (Chambers 1996, 146). The sameness of this experience is the main foundation of a meaningful description of age groups as generations.

Second, age is mostly seen as a process that reaches completion not at the end of life but rather with adulthood. The resulting linguistic stages are closely connected to acquisition. Labov was the first to convert this notion of age-related sociolectal language attainment into six stages. Although these particular stages found little support in later sociolinguistic research (Wolfram 1989; Romaine 1995), Chambers proposes a stepwise model of language acquisition that he names "three formative periods". In it, he argues:

> "that vernacular variables and style-shifting develop along with phonology and syntax from the very beginning of the acquisition. From that point, there appear to be three formative periods in the acquisition of sociolects by normal individuals. First, in childhood the vernacular develops under the influence of family and friends. Second, in adolescence vernacular norms tend to accelerate beyond the norms established by the previous generation, under the influence of dense networking. Third, in young adulthood standardization tends to increase, at least for the sub-set of speakers involved in language-sensitive occupations in the broadest sense of the term. After that, from middle age onward, speakers normally have fixed their sociolects beyond any large-scale or regular changes" (Chambers 1996: 158–59)

There are two important founding principles in this model that inform the notions of change in progress and change in apparent time. On the one hand, adolescence is defined as a time of innovation and nonconformity. It is the time at which the speech of caregivers,

which previously defined the child's life, becomes less important than the patterns and innovations used by peers. On the other hand, there are no stages that follow young adulthood. The adult life is understood as a sociolinguistic plateau with a "fixed" linguistic behavior and a keen understanding of standard language norms.

The innovative language behavior in adolescence is commonly referred to as "vernacular reorganization" (Labov 2001: 415). It is a crucial factor in language change for without this process, adults would linguistically behave entirely like their caretakers, much like young children do. The threshold between caretaker speech and vernacular reorganization appears to be somewhere between the ages of eight and twelve years (Kershwill 1996, 195). Theoretically modeled, the peak of a female-led change is reached at age seventeen (Tagliamonte & D'Arcy 2009). The incrementation model developed by Labov and exemplified by Tagilamonte and D'Arcy not only allows to pinpoint a peak of feature usage but is also able to show that the so-called adolescent peak is indicative of the maximum rate of change (Labov 2001; Tagliamonte & D'Arcy 2009).

The above-mentioned understanding of "language age" as progressing differently in different stages is crucial not only to grasp the importance of age as a social factor but also to understand the methods developed in sociolinguistics. When it comes to adulthood, the stages mentioned above inform the notion that change could be observed through societal analysis. Both the term *change in progress* and *change in apparent time* are inherently connected to this notion. While the ability to even connect change to life stages relies heavily on the definition of sociolinguistic generations and adulthood, we owe the proof that change could be observed on their basis (and thus synchronically) to Labov. Terminologically, *change in apparent time* posits the idea that a speech community undergoing a change will be presenting with various degrees of variable usage spanning generations. If younger generations show a greater feature usage than older ones, the assumption is that this new feature was innovated in their adolescence and what remains is stable in the sense that their usage will remain equally high throughout the adulthood plateau and thus create a change in language usage over generations (Labov 2001).

Labov developed this construct in his study in Martha's Vineyard. He argues that a balanced sample representative of stylistic and social differences in the community would show apparent time differences that closely connect to the changes observed in real time. He supports this argument by comparing his own data with the Linguistic Atlas of New England (Labov 1972: 24). This atlas recorded responses to a questionnaire in phonetic transcription from 1933-39. Four of the 416 participants recorded were from Martha's Vineyard. Labov could thus use these transcriptions as indicators of phonological characteristics 40 years prior to his own research and thus compare the variation in age groups observed in apparent time to a real time point in the past. This belief that the progression of change is observable in real time contradicts the common practice of research in language change at the time, which worked under the impression that "the process of linguistic change has never been directly observed" (Bloomfield 1933: 347).

To this day, research with a focus on change as a result of variation rests on these ideas of progression. The notions of age and vernacular reorganization not only inform the assumptions about how to observe change generationally, but also in which ways this progression is shaped and influenced by so-called incrementation, the process that

implements changing phonology until stabilization. Labov's linear model of incrementation not only describes the specific changes of the individual speaker, but also models how per-speaker incrementation leads to a change within the community (Labov 2001).

However, the method, though highly acclaimed and often employed, is not entirely perfect. Especially in the earlier evidence used to account for the apparent time construct, much like Labov's atlas data, the data to which the respective interviews are compared may not be regarded as similar. Many researchers have therefore started to use panel studies (studies in which participants of earlier studies are followed up and re-interviewed) or trend studies (studies in which the same generations are interviewed at a later point in time) to put their apparent time data into perspective. Sankoff describes this methodological change when she mentions "the past decade […] has seen a great upsurge in longitudinal research as sociolinguists begin to carry out follow-up studies of communities originally studied between 1965 and 1975" (2005, 1003). She compares the outcome of 13 panel- and trend studies, which all show the same change that had been observed in the initial apparent time study. With these results in mind she argues: "Though the field will continue to be surprised by the light that trend and panel studies can shed on the mechanism of language change – especially as it intersects with speaker life spans – our present synchronic methodology is a powerful lens for interpreting the past" (Sankoff 2006, 115).

Not only has the methodology of observing change in progress sustained significant levels of scrutiny over the years but the model has also been tested mathematically in various ways. Croft and Baxter show mathematically how the assumptions made about the progression of change can be proven through modeling. They confirm: "The apparent time construct has been explained in terms of a decline in flexibility in adopting novel variants, or frequencies of novel variants in appropriate linguistic and sociolinguistic contexts, as a speaker ages. We modeled this decline in flexibility with a gradual function that declined rapidly around adolescence or early adulthood. The results straightforwardly confirmed the basic account of the apparent time construct" (Baxter & Croft 2016, 169).

The usefulness of the apparent time construct for the analysis of changes in progress has recently been attested in a corpus study involving several language changes. Fruehwald employed the Philadelphia Neighborhood Corpus for his calculations. This corpus is unique in that it features interviews with differently aged speakers carried out at various points in time. Since the interviewees' age can therefore be treated as a separate variable from their individual birth year, it shares some of the evaluative power of longitudinal studies. Arguing that any and all influence of the time of the interview and the respective age as compared to generational belonging may be modelled in such a corpus, Fruehwald concludes that most changes observed show a much smaller influence of the age or interview time than of the generational identity of the speakers (Fruehwald 2017).

Though the focus of research in age-related language change and variability clearly is on change in apparent time and the methods connected to the construct, some researchers argue that there is a remaining fraction of changes which result from a community moving into a new linguistic direction as a whole. In Labov's initial model, these changes are all changes in which we do not see a change in stability in the individual and change in the

community as illustrated in the following table adapted from Rickford and Price (Rickford & Price 2013: 145).

Table 1.1: *Patterns of language change or stability in the individual and the community.*

Synchronic pattern (age distribution at one point in apparent time)	Diachronic Correlate (between two or more times in apparent time)		Linguists interpretation or characterization
	Individual	Community	
Flat	Stability	Stability	1. Stability
Regular slope w/ age	Change	Stability	2. Age-grading
Regular slope w/ age	Change	Change	2a. Lifespan Change
Regular slope w/ age	Stability	Change	3. Generational Change
Flat	Change	Change	4. Communal Change

The resulting important instances of change that closely connect to the sociolinguistic perception of age are therefore age-grading and lifespan changes. These changes appear contrastive to changes in apparent time. Against the notion of plateaus described earlier, communities in which these changes were observed diverge in two respects. They present with stable patterns of feature usage in the community, while individual speakers present with a changing usage of the variable in question (age-grading). Another diverging possibility is the emergence of changes in both the community and the individual with an increasing proportion of feature use at a point in time across generations (lifespan change). Although all are described as changes, age-grading is not a change in its traditional sense, because the feature usage in the community remains stable over time. Rather than changing the feature usage of a language, it is "cyclic" in that it would progress in the same way across the generational divide in later studies than it does at each other point in time (Rickford & Price 2013: 146). These changes are often said to connect to the pressures of the linguistic marketplace, in which non-standard features result in repercussions for the respective speaker (Bailey 2008). In a lifespan change, speakers of various generations move according to the patterns present in their community over the span of their lifetime. Both types of changes have been presented in a large enough body of research to argue that they may be the explanatory process in some of the instances of language change. They are rather difficult to account for in data that is not of a longitudinal nature and must therefore always be considered a possibility when relying on apparent time data. However, the strong probabilities of the apparent time construct not only strengthened this approach, but were to some extent also able to negate a true influence of factors such as lifespan (Baxter & Croft 2016; Fruehwald 2017) and age-grading (Baxter & Croft 2016).

This treatment of age as a variable of great sociolinguistic importance is criticized on multiple levels. For one, the treatment of adulthood generally is not in line with sociological findings. The assumption that age groups behave similarly is referred to by Riley and Riley as "cohort-centrism", which results in the false preconception that "members of all cohorts age in the same fashion as members of the cohort under scrutiny" (Riley 1971: 81).

Furthermore, adulthood may not feel the same to all ages. In westernized societies, ageism and other ideologies of old age may factor in the observed language behavior in many ways. While some theories of age may argue that old age is characterized by the alienation from mainstream tendencies, others show the effects of retirement and negative stereotypes (gerontophobia) (Coupland 2001). Without exploring these tendencies in any greater depth, they are valid reasons to argue that old age may impact social perceptions to such a great degree that linguistic patterns could reasonably be different. Furthermore, when we look at age as a progression in life rather than a stepwise climb of a linguistic ladder, there are many possibly influential cesuras we might overlook. Eckert states in close relation to this: "progress through the life course involves changes in family status, gender relations, employment status, social networks, place of residence, community participation, engagement in the market-place -- all of which have implications for patterns of [sociolinguistic] variation." (Eckert 1997, 152)

Coupland draws a meaningful conclusion when he states: "The general imperative is for sociolinguistics to take on the task of clarifying the radical uncertainty of what social ageing *means*, and how it is negotiated in language and discourse […] We need to discredit the assumption that ageing is only of interest at the margins." (2001, 202).

While I have introduced both the highly influential constructs of age as well as the criticism of age as a factor in sociolinguistics, my argumentation will later align with the apparent time construct for various reasons. One is the belief that patterns in language create groups rather than are created by groups. This is to say that while cohort-centrism may be problematic, I believe it to be deniable where age is represented through a continuous variable as a trend rather than a delimited categorical group. Furthermore, the phonological nature of the feature under scrutiny is believed to be below the level of consciousness and may therefore be very unlikely to see any age-grading in its current representation. Although I believe that social pressures may conflict with the notion of change in apparent time, I agree with Labov and others in that it may be a functional theory for phonetic features such as the one observed.

1.3.2 Gender

In theory, the difference between male and female may be allotted to the biological make-up (sex) or the socially constructed differences between the two (gender). While both concepts intertwine and are essentially predictive of language patterns only for social reasons, they have a strong influence on identity formation and are thus crucial to the sociolinguistic analysis of language (Eckert & McConnell-Ginet 2003). In the words of Eckert and McConnell, "Gender is the very process of creating a dichotomy by effacing similarity and elaborating on difference, and even where there are biological differences, these differences are exaggerated and extended in the service of constructing gender" (Eckert & McConnell-Ginet 1992: 13).

Although the concept of gender is traditionally binary, research thus far has not proven any entirely binary language patterns in the sense that features are entirely absent from the speech of one gender while occurring heavily in the other (Cheshire 2008). Nevertheless, gender as a concept was often shown to meaningfully describe two subclasses. These are not entirely opposite in linguistic behavior but instead show different

levels of variability. While the most obvious differences may best be observed in the lexical limitations of perceived male vs. female speech such as with adjectives (pretty vs. handsome, diminutives etc.), almost all levels of language production may present with gender-centric differences. It starts out as basic as the pitch level. While the differences in vocal tract size account for some of the difference between male and female speech production, the differences in vocal tract sizes are not proportional to the perceived pitch differences even in young boys and girls (Eckert & McConnell-Ginet 1992: 14).

In terms of phonological variation, gender has been a factor in stratification as much as (or more than) any of the other canonical social patterns. While early research in sociolinguistics was rather focused on social class differences, in studies such as Labov's department store study, gender was soon analyzed as the second most important factor. In fact, Coates reevaluated some class-focused variation studies and concluded that gender would have been the preferable explanatory variable in many of them (Coates 1986). The great similarities in many of the findings from thirty years of gender-related variable evaluation in sociolinguistics finally led Labov to formulate three principles of gender patterns in (phonological) linguistic change. The principles are as follows:

> Principle I: In stable sociolinguistic stratification, men use a higher frequency of nonstandard forms than women
>
> Principle Ia: In change from above, women favor the incoming prestige forms more than men.
>
> Principle II: In change from below, women are most often the innovators (Labov 1990)

These principles may be differentiated and criticized in multiple respects. While all principles relate to gender patterns in language, principle I is descriptive of a stable state of the variability in language while the other principles both describe language change. Furthermore, it should be noted that in principles Ia and II, women essentially show very different characteristics. They exhibit both conservative (prestigious) language behavior while at the same time innovating new forms.

However, defining the orderly heterogeneity in gendered language use cannot provide an insight into why and how these patterns may occur. Principle I generated a great deal of interest in explaining the pattern, ranging from Fasold's proactive understanding of women's conscious non-conformity with local language expectations to index social mobility (Fasold 1990a), to Gordon, who links female withdrawal from local forms to the alleged sexual promiscuity of localized variants (Gordon 1997) and Deuchar, who sees a link to non-face-threatening politeness strategies (Deuchar 1990). Theories of politeness are employed even in recent gender-related studies as successful additions to the evaluation of constraints, as has recently been shown by Holmes-Eliott & Levon (Holmes-Elliott & Levon 2017). Both gender-related politeness approaches are based on the theory of Brown & Levinson that describes politeness in relation to the concept of saving face (Brown & Levinson 1994).

Trudgill's interpretation of female language behavior is amongst the most commonly used perspectives in explanations of gender differentiated speech. He argues that women need to claim social status while men acquire it by occupation. Therefore, women use language to raise their social status through prestigious and/or standard forms

while men show little use of prestigious forms across the board, possibly in congruence with masculine identity expectations. The "covert prestige" of these non-prestigious forms is proven to lead them to overreport the usage of working class forms, even where they may be using the same standard forms as female speakers (Trudgill 1972).

James investigates various explanatory arguments and while some of them appear sounder than others, she concludes that power imbalances are at the core of most of them. While this certainly may aid in a sociologically sound understanding of the observed language patterns, she also warns against the conclusion that power struggles lead all women across various communities to pattern in certain ways (James 1996).

Furthermore, Labov's principle I closely relates to his understanding of speech communities and standard versus vernacular language use. To equate formality with prestige and speech communities with shared social norms creates an undeniable backdrop for gender patterns that may be much shakier than the patterns that are observed because of it. Cheshire criticizes the described lack of empiricism in this approach (Cheshire 2008). While the concepts as indicated above should be evaluated critically, the overwhelming body of research statistically correlating gender with speech clearly points to the value of principle I for analysis.

Principle Ia and II, though they have been validated by Labov's own research, are considered even more problematic than principle I. As was mentioned earlier, they appear contradictory to each other in describing opposing characteristics in females. Labov describes this behavior as the "gender paradox" (Labov 2001). Other researchers have criticized this approach for its unnecessary connectedness with social class (Coates 1986) and argue instead that "females [do not] favor prestige forms, but that they create them; i.e., if females favor certain forms, they become prestige forms" (Milroy & Milroy 1993: 65).

At their core, Labov's principles have further been criticized for underestimating the performance and construction aspect of gender (Eckert & McConnell-Ginet 1992). In Eckert & McConnell-Ginet's definition, "gender is constructed in a complex array of social practices within communities, practices that in many cases connect to personal attributes and to power relations but that do so in varied, subtle, and changing ways" (Eckert & McConnell-Ginet 1992: 484). Based on this community-created differential complexity of gender as a variable, Cheshire stresses the value of ethnographic approaches to gender. These would enable the researcher not only to account for the existent gender differences in a community but also allow for an analysis of what it means to the individual subject to perform their respective gender identity (Cheshire 2008).

While all criticism and complexities of defining gender in sociolinguistic research introduced above may inform the idea that gender should not be dealt with in its binary structure in research, I side with Holmes who argues:

> "We should be concerned with refining generalizations, rigorously confining the area to which they apply, but we should not regard them as useless when exceptions are identified. The fact that there are limits on their applicability should not blind us to the immense value of generalizations and the formulation of sociolinguistic universals in furthering theory development" (Holmes 1998).

Existing preconceptions of gendered language behavior in linguistic change will be a crucial aspect not to identifying a change as such in my data, but to understanding how the patterns observed can be meaningfully put into a research perspective. Identifying a gender pattern will not mean identifying a change but rather identifying a factor in the change that deserves a qualitative explanation through the metalinguistic data I collected. And where sufficient judgment is lacking, this lack may in turn become part of the analysis.

1.3.3 Education and Social Class

Education and social class are discussed in the same sub chapter as they often supplement each other in various studies of language variation (e.g. Labov 1972).

Anecdotal evidence may inform the feeling that positioning in society is highly influential on how a person talks. However, as already mentioned, the manner in which these positions are created are categories shaped and developed in social sciences. One of the most influential thinkers in categorizing the differing social groups in western societies is Karl Marx. Without going into detail on his theories, Marx speaks of social classes that are derivative of the individual's power in respect to others. Power is based on how (much) money can be earned either from labor or through possessions, which is referred to as capital. Marx's theory results in a binary set of conflicting social classes, which can be ranked according to whether they profit from the capitalist system or are exploited by it. Further divisions are based on how much the individual group benefits from or is exploited by the system (with the lower working class and the upper class on opposite ends of the ladder) (Marx 1887).

In contrast to Marx, western sociologist Weber prefers to work with a more complex system, still grouping people into classes but developing more nuanced factors in deciding what these class notions may rely on. Rather than judging the individual's power, Weber bases his approach on the notion of status. Status is the passive position in respect to others in the society rather than the active power that can be put forth in respect to others. This differentiation is important, because it illustrates how the term "class" can not only be interpreted in various ways, but also be used in various ways of social criticism (Edgell 1993).

Bernstein's analyses of working-class children can be considered one of the first approaches to social class notions as predictors of linguistic behavior. He was mainly concerned with the limitations children of working-class speech face in a social environment, where a less "restricted" and "context-specific" register was expected of them (Bernstein 1962).

The first linguist to develop a language specific theoretical concept of social status and power was Bourdieu. He uses both Marxian capital and Weber's status concept and draws a connection to language use. He states: "Linguistic competence (like any other social competence) functions as linguistic capital in relationship with a certain market" (Bourdieu 1977: 651). Bourdieu's market metaphor was taken up by Sankoff and Laberge, who described the so-called "linguistic marketplace". In these ideas, members of societal groups have unequal capital in the linguistic marketplace. While some have the ability linguistically to benefit in multiple negotiations of power through their language abilities, others are rather limited in their linguistic capital (Sankoff & Laberge 1978).

If we consider Labov's research as constituting sociolinguistics methodologically, then social class representations of language have concerned linguists from the very start of the subdiscipline. This concern was often motivated by the link between language and the dire consequences of identifying someone as belonging to a lower class. Labov was one of the first to define social stratification in linguistics, when he argued that "the use of this term does not imply any specific type of class or caste, but simply that the normal workings of society have produced systematic differences between certain institutions or people, and that these differentiated forms have been ranked in status or prestige by general agreement" (Labov 1972: 44). It is important to note, as alluded to above, that this understanding relies on a binary division between a language standard and the vernacular. In most of his early studies, especially his New York City department store study, Labov found significant differences between the social classes he described, and their linguistic behavior. Most of these differences relate to the usage of stigmatized forms by lower social classes and prestigious forms by the higher social classes. Labov differentiates between sharp and fine stratification. In fine stratification, members of the respective social classes linguistically behave different from each other to a similar degree, creating relatively balanced distances for feature usage. Sharp stratification on the other hand shows unequal feature usage, most often resulting in a bigger gap between lower classes and the upper middle and upper class (Labov 1972: 113–14).

Labov's initial department store study, where social class was merely conceptualized through the expected customer and sales representative profile of each store, soon encouraged a new study of New York City English with improved quantifiability. Labov therefore implemented a threefold indicator scale combining education level, family income and occupational rank (Labov 1972: 112). Trudgill later chose a similar approach, reasoning that "even the most affluent manual workers retain the values, ideas, behavior patterns and general culture of the working class, and there has been little embourgoisement of the British working class" (Trudgill 1972: 4). In his Philadelphia study, Labov connects the threefold indicator scale with individual occupation, education and house value. Through this comparison, he provides data showing that occupation is far more predictive of social class identity than education or house value. Yet, the combined factors were still more predictive than each factor taken on its own. The index therefore persisted (Labov 2001). How this relates to women and children further complicates the categorization.

Labov later divided these class-stratified variables into stable variables (those that create variability in the speech community but are currently not undergoing change) and changes. Changes are further subdivided into "change from above" (those variables that are introduced into the speech community by members of the upper classes) as well as "change from below" (where patterns are changing but speakers are entirely unaware of the variation with which they perform). These patterns reliably describe variability in many of his studies. This reliability inspired Labov to take these a step further, namely to aid in the prediction of which social groups lead changes. He finds in his analyses that they "are located in the central sections of socioeconomic hierarchy" (Labov 2001, 190).

However, even in the early days of research in stratification, not all speakers behaved in these highly predictable ways. In some of Labov's studies, he noticed what he

describes as a "crossover" pattern- whereby speakers in the middle class were not behaving within the mid-range on the prestige continuum but rather similar to those speakers that were categorized as speakers of higher social status. Since the usage of prestige forms by middle classes results graphically in non-linear patterns, this behavior is also referred to as the curvilinear hypothesis (Labov 2001).

Education is also a factor that often becomes part of the indicators for social class in variationist research without theoretically going beyond mere categorization. There are several ways to group participants according to their education, such as grouping by years of education or level of education (e.g. Labov 2001; Labov, Rosenfelder & Fruehwald 2013; Burnett, Koopman & Tagliamonte 2018). Since education is often used in conjunction with other social factors, it has never become canonized to the same extent as the above-mentioned social factors.

The issues with quantifying social class have not been solved in present-day studies, although it is often included as a factor. While linguistic heterogeneity seems to be based heavily on education and social class even in the following study, education levels or class differences are not meaningful predictors in the sense that they allow for individual classification. While lack of education anecdotally often results in syntactically and lexically less complex speech patterns (Burnett, Koopman & Tagliamonte 2018), this does not hold true to the same extent for phonetic features. Furthermore, the current status of education in the US would be fully underestimated based on years of schooling or diploma categories. While not completing high school may be meaningful anywhere, completing high school seems to underestimate the vast difference between state and private schools and all other nuances present in the current education system.

1.3.4 Ethnicity

Ethnicity is one of the most complex canonical factors in sociolinguistic research. At its conceptual stage, there are several ways to construct ethnicity based on concepts such as nation, ancestry, religion, tribal membership, race or language. How ethnicity is defined is highly context dependent, and which aspects build the background for the specific community varies greatly around the world. The concept borders on similar constructionism as the concept of gender and it shares the notion that the concept is informed to some extent by differences in phenotype. This is to say that in a society of cultural mixes and complexity, speakers of a certain skin color may include this as a factor within their identity construct or feel as though others use it in their perception of them. The complex entity described as an ethnic group is defined by Zelinsky as follows:

> "The ethnic group is a modern social construct, one undergoing constant change, an imagined community too large for intimate contact among its members, persons who are perceived by themselves and/or others to share a unique set of cultural and historical commonalities, which may be real or imagined. It comes into being by reasons of its relationship with other social entities, usually by experiencing some degree of friction with other groups that adjoin it in physical or social space. Levels of awareness of its existence can vary considerably over time and in accordance with circumstances. Ethnic groups can exist within a hierarchy that ranges from the smallest aggregation

> meeting the stated criteria to a politically sovereign national community or even beyond to entities transcending international boundaries. For individuals, affiliation may be either mandatory or a matter of personal option, one susceptible to change and one contingent upon political and social circumstances." (Zelinsky 2001, 44)

There are a number of aspects in this definition that are especially crucial to the concept of ethnicity in its relation to language variation. First, ethnicity is an identity related concept that is created both through individual identification as well as the perception of outsiders. The friction enforces the notion of ethnic groups, while at the same time creating a very time- and circumstance-related membership identification. The indexicalities of ethnic groups are highly complex and not comparable. In the present case, identifying and being identified as African American is phenotypically and culturally informed, rather than through African heritage or language. Thus, African American culture is perceived as American subculture in many ways and although heritage is of historical importance, African Americans are not generally perceived as immigrants. In contrast, for Hispanic-Latinxs[3], the ethnic identification is closely related to the prevalence of Spanish and immigration. This ethnic identity is therefore often perceived as a transplanted culture rather than an American subculture.

Furthermore, all the sociolinguistic research concerning the concept of ethnicity is both made possible by the construction of "a community too large for intimate contact" while at the same time in need of critical evaluation as each member of this community may make differential use of the positioning in "political and social circumstances". The possibilities of membership identification were explored by Holliday, who was particularly interested in understanding bi-racial identities. She found that bi-racial participants show variability in peak delay intervals when speaking with black interlocuters. (Holliday 2016).

The meaningful representation of ethnic identity in linguistic variation and the perception thereof has been researched especially by Thomas and Purnell, who both conclude that there are multiple cues in being perceived as a speaker of Standard American English (SAE), African American English (AAE) or Chicano English (ChE) (Purnell, Idsardi & Baugh 1999; Purnell 2010; Thomas, Lass & Carpenter 2010). Ethnic identity may create an ethnolect, which is a distinct variety created on the basis of ethnic group membership. Clyne defines ethnolects as: "varieties of a language that mark speakers as members of ethnic groups who originally used another language or distinctive variety" (Clyne 2000: 86). Others have considered ethnically marked speech a dialect or language variety (Labov 1972; Purnell, Idsardi & Baugh 1999; Fought 2007). It is necessary to differentiate here between understanding these as varieties or ethnolects and considering the plurality of variation outside of dialectology. Research in ethnolects may wrongfully create the assumption that regional variation mostly occurs in unmarked White speakers. However, regional features are just as or even more prominent in African American

[3] This term may read unusually, but is a possible gender-neutral category that is currently gaining in usage. Since this is a linguistic dissertation, I have decided to follow this new trend in a movement toward inclusive language use.

speakers and other ethnic groups (Becker 2009). In order to deal with the problem of oversimplifying ethnolects into varieties rather than possibilities for variation, Benor coined the phrase "ethnolinguistic repertoire". She defines this as "a fluid set of linguistic resources that members of an ethnic group may use variably as they index their ethnic identities"(Benor 2010). This solves the issue of wrongfully expecting that features cannot be shared by other ethnic groups, or even that no correlation with social factors other than ethnicity are to be expected. Several interactions between ethnicity and factors such as social class and gender (Fasold 1990b) or age (Fasold 1990a) have been shown in research.

The intricacies of the relationships between minority and majority communities are crucial to understand the functions and limitations of any non-majority ethnic group, whether in ethnolectal or ethnically variable speech. There are a number of ways in which the differences between these communities may be lessened or enforced by linguistic variability. Both convergence and divergence of patterns are shown to result from the intricacies of this relationship (Fasold et al. 1987). Interestingly, in the case of AAE and other American English varieties such as Southern English, differences in convergence or divergence patterns appear not only through the performance of different individuals, but also on the various levels of speech production. Fasold et al. argue that there may be convergence in phonology while grammatical structures diverge (Fasold et al. 1987, 60). Labov finds patterns of divergence and claims that ethnic speakers do not attempt to participate in the majority group's patterns but "are instead oriented to a national pattern of koine formation within the nonwhite groups" (Labov 1994: 157). More recent research has delivered evidence for both the divergence from White speech as well as the convergence with some patterns (Fought 1997; Eckert 2008a; Becker & Coggshall 2009). Eckert even reports "hyperwhite" speech patterns in California as an ethnolect, because they fulfill the same meaning of boundary awareness as ethnolects commonly do (Eckert 2008a).

The identity concept and its relation to ethnicity is described in greater detail below. In its most commonly used understanding however, the positive or negative perception of the majority language community may influence the linguistic performance of any speaker living in the community. This perception of the community may be influenced both by ecological factors (relating to the make-up of the community in ethnic diversity and contact possibilities) as well as ethological factors (relating to the respective attitudes vis-a-vis the other) (Whinnom 1971). Interestingly, some studies have shown that these tendencies are so strong that incoming minority groups may align linguistically with patterns of other minority groups purely to diverge from the White community (Fought 2007).

The two biggest minority groups in Austin are Hispanics (also referred to linguistically as speakers of Mexican English or Chicano) and African Americans. In the following subsection, I present the intricacies of both of these cultural groupings while paying particular attention to the perceived differences.

1.3.4.1 Hispanics/ Chicano English

Hispanic and Latinx migrants comprise the fastest growing minority group in the US, especially in border states such as Texas. More information about the general demographics for Austin are provided in the methods section below. In this subchapter, I focus on general research from the US.

A large body of data has analyzed the speech of Hispanic or Latinx-identifying Americans in various regions of the US (Santa Ana 1993; Bayley 1994; Santa Ana 1996; Callahan 2017). The variety of English spoken by this ethnic group is often referred to has Chicano English (Santa Ana 1993) Mexican English (Thomas & Ericson 2007) and Tejano English (Bayley 1994) among other terms. Fought describes Chicano English as

> "a non-standard (though linguistically rich) variety of English, which is spoken natively primarily by U.S.-born speakers, and which shows the influence of Spanish, especially in the sound system. It emerged historically from a context in which English and Spanish were in constant contact, both across the community and within the competence of individuals". (Fought 2010: 45)

I will therefore continue to refer to these speakers as Chicano, because the term may include both Hispanic and Latinx speakers and is functional without denoting a specific region or heritage country such as Mexico.

The debate on the overall classification of speech patterns in Hispanic and Latinx people was long informed by the notion that the Spanish-English bilingualism of community members created a learner version of English highly influenced by their mother tongue. What was later defined as "interlanguage" may be part of Chicano English in communities with continued influx of migrants even nowadays. However, interlanguage is unable to account for those speakers of Chicano English who present with characteristic features even though they consider English their first language. Precisely at the crossroads, where variability cannot simply be accounted for by deficiency, is where the ethnolectal concept of Chicano English is disconnected from the heritage concept and becomes a factor in variability. Wald argues in favor of a distinct Chicano English that certain features of interlanguage English usage are absent in those speakers who consider themselves first-language-speakers of English rather than Spanish (Wald 1984). One of the examples he introduces is epenthetic vowels, which frequently occur in Spanish monolingual learners of English but can never be found in monolingual Chicano English speakers.

Santa Ana provides a detailed description of the shortcomings of Chicano English research. One of the most prevalent theoretical problems he identifies in describing Chicano English properly is the evaluation of features without differentiating between monolinguals (Spanish or English) and bilinguals (as in Garcìa 1984).

According to Santa Ana, theoretical problems arise where Chicano is not analyzed as a meaningful pattern of variability within the English language, but instead described through a comparison with Spanish. Another source of conflict is basing identification of the membership in this speech community on the knowledge and usage of Spanish (as in Sanchez 1983). However, he recognizes the complex situation of Chicano English by creating a model that differentiates between the different levels of Spanish contact and speech abilities. The aspect that is most important for the following analysis is the fact that Chicano English is a variety of English that may be distinguished through linguistic features and that is based on a choice of identifying with this ethnolect. Whether the speaker is bilingual and uses these features or a monolingual of said variety does not change the variety we consider Chicano English. Nonetheless, the features may be informed to some extent by the characteristic language behavior of non-native speakers. The most important support for this argument can be found in language features that are either not of Spanish

origin or behave similarly on the surface production while being informed by different factors in an analysis (such as syllable sonority) (Fought 2007).

Further issues, which are similarly discussed in the section on African American English below, concern the widespread geographic occurrence of Chicano English. As alluded to in the Zelinsky quote, Chicano speakers are a community "too large for intimate contact". Several researchers have shown aspects of regional difference within Chicano communities in different historical migrant communities (Bayley 1994). Newer migrant communities show very different levels of English proficiency status and new local dialects influencing the development of a Chicano variety in these communities (Kohn & Franz 2009). The resulting understanding of previous research is that there will be a historically developed ethnolect of Latinx and Hispanic speakers in Austin, possibly a version of Tejano English. Since I do not attempt to make a general statement of exactly what Tejano English in Austin is, further distinctions are not a concern here.

1.3.4.2 African Americans

Contrary to Chicano English, African American English[4] currently shows no features identifiable as second language transfer. Rather, it is a variety comprised of unique phonological and morpho-syntactic features that relate to an ethnic identity. Though some influence of AAE on White regional varieties has been investigated (Wolfram 1974), the history of this variety has a complex relationship with the development of White American Englishes in various geographic locations.

Although the aim of the study and the methods certainly is not dependent upon how African American English developed historically, I must briefly explain the different theories concerning the past and current status of this variety, as they underline both the similarities and the differences it shows in comparison with Chicano English.

There are, generally speaking, three different approaches to the theory of how AAE developed into what it is today. One assumption is the so-called English origins hypothesis, in which features prominent in AAE are considered to have developed from historical versions of English. This hypothesis has been put forth mainly by Poplack and Tagliamonte, who focus on the process of divergence as driving factor in creating AAE (Poplack 2006).

The creole origins hypothesis developed as the most extreme contrast to this assumption. It focuses not on the involvement of the English language but rather the contact situation and its unique status for language development. In this hypothesis, AAE is closely connected to African ancestor languages and believed to have been strongly influenced by their features and structure. It was initially and most strongly articulated by Stewart and Dillard (Stewart 1967; Dillard 1972).

However, current research is skeptical of the extreme polarity of both theories and favors a hybrid position that acknowledges both the influence of early creoles in some

[4] I choose to call the language variety spoken by African Americans AAE due to my focus on the English language and my critical view of the term "vernacular".

locations of the historic United States, while leaving room for the significant influence of the superstrate (Van Herk 2015; Mufwene 2015; Rickford 2015; Winford 2015). Both positions are important to keep in mind because they are crucial in understanding, just as outlined above for Chicano English, that an initial contact situation of two different languages such as African languages and English or Spanish and English does not necessarily result in a creole language, and even where some features may be contact-induced, what matters is the variability created and enforced through ethnic identity. The influence of Spanish on Chicano English may be more easily argued for due to the persisting contact situation of new Spanish immigrants with English as a superstrate language. However, the development of AAE shows that an ethnic language variety will eventually develop and change within itself and not necessarily because of contact. The question here is whether the differentiation of English and AAE is facilitated through decreolization, the process through which the reduced language features of a creole language develop into a fully functional language variety.

Historically, it is important to consider the Great Migration between 1916 and 1970 in understanding the relationship between Southern regional varieties and AAE (Wolfram 1974). While AAE may have developed historically (with or without creole stages) in the South, migration patterns have brought and further developed its characteristics in the metropolitan centers and cities in the North (Farrington 2019). It would therefore be wrong to consider AAE in Austin as a purely Southern variety, because this ethnic variability is also influenced by developments in the North. Furthermore, the older AAE speakers in Austin are affected by the impact of these migration-related demographic changes through changes within their community, while younger speakers are impacted by the development of AAE characteristics beyond the South.

Research in AAE is substantial and has dealt at length with racist and other social concerns relating to the linguistic structure of the speech of African Americans over the past 50 years. Among the most prominent such debates is the Ebonics controversy (Williams 1997). Early research was especially focused on those working-class young adults in whom the most extreme vernacular patterns are to be expected (Labov 1972; Fasold 1990b). Wolfram and Schilling criticize: "These accounts emphasized the uniformity of an idealized vernacular variety rather than variation within African American speech in regional and social space" (Wolfram & Schilling 2016: 219). They further argue that an approach to AAE based on the above described ethnolinguistic repertoire is more appropriate because it allows for a much more fluid and flexible creation of desired linguistic meaning. In such an understanding, AAE features are a pool of linguistic opportunities that may aid both in creating ethnic identities and aligning with the social meaning attached to these individual features such as coolness or authority (Wolfram & Schilling 2016).

Another fact about ethnic language variation that is important to the present study is the notion that this variation is non-regional. When we think of local varieties of English, most of the features we think of are features of majority group speech. The connections between different communities and the meaning of the distinction between rural and urban is different for AAE than for other varieties of English. In the aftermath of the Great Migration, current trends show that African Americans are likely to move from

metropolitan center to metropolitan center, and even in formerly isolated African American communities, features are developing within a supra-regional trend (Wolfram & Thomas 2002). And, even those who move into the suburbs "tend to have more extensive contact with African Americans in urban areas than they did a century ago" (Wolfram & Schilling 2016: 233). Furthermore, the sense of African American identity is evolving both due to thriving youth cultures (Lippi-Green 2012) and also because of the "persistent de facto segregation of American society [that] fosters a social environment conductive to maintaining a distinct ethnic variety" (Wolfram & Schilling 2016).

This subchapter has shown how the formation of an ethnic identity in the US may be informed by heritage but has shifted away from the status of non-US and US citizens. While the United States is a historical melting pot, the current situation is arguably a strongly formed categorization of Whites in the US versus all other ethnic communities who speak English (at least to a certain extent in the case of Mexican communities) as their first language. Defining ethnicity purely as a choice would underestimate the social pressures of race, while creating a national heritage informed view of this social factor would underestimate the choices and membership intricacies involved in assuming an ethnic identity. In the following study, the approach is to represent self-identified ethnicity while recognizing that we cannot know to what extent this chosen ethnic identity is a more or less forming and performed characteristic of each individual's life and thus their variable language usage.

This section dealt both with ethnicity as a constructed social category for social actors with racial backgrounds as well as the description of a root to a sociolect (Chicano English or AAE). For the present investigation, the sociolect is the part that does not lend itself to investigation. We cannot know whether participants of a certain ethnic identity are retracting /s/ in /str/ as part of their ethnolect or just using this variable in identity formation practices.

1.3.4.3 Commonalities between African American and Chicano communities in the US

While the feature usage characteristics of AAE or Chicano English may often be described as distinct and divergent from one another, their socioeconomic status as members of minority groups is often similar. This means that both the Chicano and the African American workers in the US often belong to lower socioeconomic classes. The resulting shared feeling of devaluation and inequality can strengthen divergence from the superstratum variety as argued above.

Several studies have found shared features between AAE and Chicano English, especially amongst younger speakers in close contact with African American communities (Wolfram & Schilling 2016). Wolfram has shown that grammatical patterns frequently used in AAE also appear in Puerto Rican teenagers (Wolfram 1974). Furthermore, New York City's r-lessness is a regional feature that is adopted both by Puerto Ricans and African Americans (Becker 2009).

Especially where features are not clearly indexing an ethnic identity, as described for the feature researched in the study below, there is a chance that using a certain linguistic feature may not be directly connected to the respective ethnic identity but rather be used as

a sign of divergence from the white majority group. This is in line with the conceptually open understanding of ethnolinguistic repertoires, in which features are openly available to all speakers inside and outside ethnically formed communities to index social meaning in connection to ethnic identities and/or in divergence from superstrate norms.

1.3.5 Region

Regional identity is a further factor that causes variation in speakers. In contrast to the other social factors mentioned thus far however, this factor has been investigated both by variationist sociolinguists and social dialectologists. While the former focus on variable dispersion, the latter aim at circumscribing those characteristics that classify speech as a distinct dialect.

Region may be taken into consideration for the present study from two different angles. On the one hand, there are certain patterns of the spread of a change that have been proven to occur when feature usage moves in space. Generally speaking, these are the patterns of interest when it comes to evaluating whether and how the variable retraction of /str/ occurs in various dialect regions of the US.

At the same time, region is the foundation of social dialectology. It is because of the respective region a speaker lives in that we can expect them to draw from the pool of regionally charged alternatives in their speech patterns.

The present study does not compare similar data from multiple regions, and therefore, it is not suitable to present any regional effects on the pattern investigated. Nonetheless, I must briefly sketch the respective dialect region and its characteristics as a founding component of the speech community. More specific statements about Austin are made in chapter 4. I then briefly explain how language change is expected to spread based on region.

According to the Atlas of North American English the dialect region that Texas belongs to is generally the South, which encompasses the states in the Southern United States ranging from Texas at the east-most point to the West Coast with Kentucky and Virginia as northernmost points. Many of the features used to establish this dialect area are vowel features, the most prominent change being the Southern vowels shift. Even in those changes in which northern Texas behaves noticeably differently such as the monophthongization before voiceless consonants, Austin remains aligned with the other regions within this dialect. The only consonantal features discussed as Southern features are word-final consonants that are of no concern to this study (Labov, Ash & Boberg 2006).

The spread of dialects is generally discussed in terms of diffusion and transmission. Diffusion is the spread of change across speech communities and region while transmission describes how features spread from speaker to speaker. There are several theories as to whether this type of change is better understood as cascading waves, through a family tree or hierarchical diffusion (Labov 2010). Since the spread of this change cannot be investigated with the necessary detail to model its diffusion pattern, I do not go into any further detail here.

1.4 Intraspeaker Variation and Change

As explained above, interspeaker variation sees the individual as a one-dimensional representation of social categories that inform this individual's language behavior equally. However, speakers do not only vary and change their linguistic behavior based on these categories but are prone to change it in congruence with changes in context. While interspeaker variation was first described mainly as "style", ideas of what "style" is have gradually become more complex and joined with notions of stance and identity performance. The understanding of style has been heavily influenced by developing notions of both behavior in social groups as well as the notion of identity. In the following, I will depict the evolving concepts of intraspeaker variation based on the understanding of what causes them. The categories are style, audience design and identity. The present study tests directly for stylistic differences, but concepts of meaning and identity are nevertheless important for a successful analysis of the data at hand, especially in the thematic analysis of comments.

1.4.1 Style

Early in the discipline and in his own research, Labov noted that there is a problem in the "isolation of contextual styles" (Labov 1972: 70) that had up to that point been ignored in linguistic research: "Since the influence of stylistic conditioning on linguistic behavior is said to be merely statistical, it leads to statements of probability rather than rule and is therefore uninteresting to many linguists" (Labov 1972: 70). To illuminate the order in the heterogeneity of stylistic differences, Labov defines the dichotomy of "careful speech" and "casual speech" (Labov 1972, 79). However, the representation of stylistic differences presents itself in a much more complex manner in respect to how it functions in linguistic variability.

Labov defines five methodological "axioms" in connection to stylistic differences. The first methodological axiom is "style shifting", which describes that the stylistic range speakers utilize varies. Though all speakers shift, the rate and diversity with which they shift is unpredictable. Attention paid to speech is Labov's second and likely most-reviewed axiom. Attention, in Labov's understanding, is linguistic self-monitoring throughout the production process. Said monitoring is hypothesized to have an influence on the degree of carefulness in speech. The more the focus shifts away from this formal awareness, the more likely the speaker is to use their own version of casual speech. Labov's third axiom, the vernacular, is defined as the result of minimal attention paid to speech and thus considered the most desirable data source for sociolinguistic analysis. The other extreme of the speech resulting from attention paid to speech is the notion of formality, Labov's fourth axiom. This is the style of speech very likely to be used in an interview environment as an effect of the monitored situation. A last axiom is good data. Although the ideal data for the researcher is the vernacular, good data cannot be obtained without the interview situation and a recorder present (Labov 1972: 209). Overall, Labov's understanding of style is characterized by the bipolar understanding of language production on the axis from vernacular to formal, and attention paid to speech is used as a singular explanatory mechanism for both hypothesized sides of the spectrum. He acknowledges a similar

constructionism to other social factors in sociolinguistic analyses by addressing the individuality of style shifts.

One of the problems with the axioms used to describe style as a factor lies in distinguishing "'functional varieties' and 'cultural levels' as completely independent dimensions" (Labov 1972: 240). In connection to the study of /ing/ variability he describes, he already mentions the issues of overlap of social and stylistic stratification and stresses the issues with distinguishing "a casual salesman from a careful pipefitter" (Labov 1972: 240). This metaphor points in the direction of problems with understanding style as a one-dimensional behavior under the notion of formality: What if the careful pipefitter chooses to be careful or adjusts his behavior according to context factors other than attention? Other conceptions of style that also involve the (social) environment will be described in the following.

1.4.2 Audience Design

The rather subconscious choice of different variants as a result of varying social environments is described in Bell's "Audience Design Model" (Bell 1984). Bell criticizes Labov's simplistic categorization as a "nonstarter" and argues the "empirical foundation for the attention variable is notably lacking" (Bell 1984: 148). He further criticizes the interview-based focus as incomparable to real speech, such that "Storytelling, reading, and so forth are best seen as interview techniques, which attract attention to or divert from speech and so produce styles analogous to how people talk in different everyday situations. The techniques used to elicit speech at successive points of the style continuum should not be mistaken for factors that actually account for the variation. "Word-list-style" is merely a style produced when someone reads a word list." (Bell 1984: 150).

Two aspects are central to Bell's argument. On the one hand, he sees a close connection between interspeaker and intraspeaker variation. The speakers are not just shifting according to an individual interpretation of formality but they echo the patterns that are meaningful in their immediate environment. Rather than Labov's paralleled understanding of stylistic and social variables, Bell argues that "social variation comes first" (Bell 1984: 151). He then proposes the framework of "Audience Design", which "assumes that persons respond mainly to other persons, that speakers take most account of hearers in designing their talk" (Bell 1984: 159).

The idea that social identities are actively shaped within a certain environment is taken from Beebe & Giles' speech accommodation theory. In a much broader sense than audience design, this theory takes a social psychological point of view. It describes how style shifting is an individual's adjustment to their addressee, either through convergence or divergence. Specifically, they expect convergence with the addressee when speakers "(a) desire their social approval and the perceived costs of so acting are proportionally lower than the rewards anticipated, and/or (b) desire a high level of communication efficiency, and (c) social norms are not perceived to indicate alternative speech strategies" (Beebe & Giles 1984: 8). The conditions for divergence are understood as situations in which speakers "(a) define the encounter in intergroup terms and desire positive ingroup identity, or (b) wish to dissociate personally from another in an interindividual encounter, or (c) wish to bring another's speech behavior to a personally acceptable level." (Beebe & Giles

1984: 8). Studies such as Coupland's assessment of a travel agent's accommodation of certain variables in their clients' speech have underlined this social behavior (Coupland 1980).

However multi-faceted and far reaching Beebe et al.'s speech accommodation framework may be, Bell thought it was necessary to extend the type of "analysis of stylistic shifts or differences [that had been] done perceptually in accommodation theory" (Coupland 2007: 63). He also developed a much more nuanced understanding of all the possible types of addressees in a communicative situation. The first analytic step is taken both in his analysis of Coupland's travel agent study as well as his own research concerning ethnically different interviewees. Instead of describing the audible characteristics in broad strokes, he closely analyzes linguistic variables such as t-glottalization.

In his nuanced view of the addressee, he understands the direct communicative partner as the second person addressee. However, as is true in many daily situations, there may be other listeners present in a communicative situation, who may also influence style shifting. In descending levels of influence, Bell characterizes the possible people on the third person addressee level as auditors, overhearers and eavesdroppers. Auditors are not directly involved in the speech situation but could participate from the speaker's point of view (are ratified). Overhearers would not be able to participate but the speaker is nonetheless aware of their presence. The presence of eavesdroppers however is neither known nor ratified by the speaker (Bell 1984).

Bell's model also provides a first step in differentiating the 'careful pipefitter' more adequately. Firstly, the model allows for style shifts that are responsive (as in the Labovian shift) and style shifts that are initiative. Although Bell initially claims that very few style shifts are initiative, he later extends this part of stylistic shifts (Bell 2001). In responsive shifts, he also allows for a non-audience dimension that is comprised of external factors in the communicative situation such as setting and topic. In initiative style choices, he differentiates between the future referees that may listen to the speaker's communicative production as well as the hyper addressees that are assumed personalities in instances such as public speeches.

Furthermore, Bell discusses the existence of style shifts in relation to social behavior. Instead of treating stylistic variation as an isolated phenomenon that is added to interspeaker variation, he stresses the circular movement in which stylistic variation becomes meaningful only where group identities are shaped and features become salient in individual usage. Many studies have followed Bell's ideas and created experiments to show the influence of the audience on the speaking individual. In his own research, Bell was able to show the differences in interviewee speech when there are perceived ethnic differences (Maori vs. Pakeha (Bell 2001). In Rickford and McNair Knox' interviews with Foxy Boston, they not only showed the linguistic importance of the second person addressee in an interview situation, but also analyzed differences perceived in situations that likely involved the topic of the conversation more so than the audience (Rickford/McNair Knox 1994). Initiative style shifting is naturally difficult to study and has therefore hardly been addressed in sociolinguistic research.

With the research in the complexity of style shifts grew the understanding that although Bell had provided a much more nuanced outline of the possible factors in

communicative accommodation, one particular shaping factor was missing from his analysis. The underlying notion in this shift in perspectives is that speakers not only accommodate to their social environment responsively but may also proactively create their conversational persona through variability (Wolfram & Schilling 2016: 301). Identity and its influence on style theories will be described in the following.

1.4.3 Identity

More recent approaches to style are grounded mainly on an understanding of the concept of identity that is fluctuating and variable. This variable identity concept results in viewing identity in communicative situations as dynamic and variable. The variability exists beyond the negotiation of context awareness and attention paid to speech. As described above for the third wave of sociolinguistics, studies with a "speaker design"-approach are joined by the notion that the speaker does more than just passively monitor and react to context but rather has agency on various levels of consciousness to vary their speech (Wolfram & Schilling 2016). Wolfram and Schilling use the term "speaker design" to explain a number of studies that are based on these ideas, because none of them have claimed a specific model like Bell.

The idea of bridging language and identity as such is described in detail in Le Page & Tabouret-Keller's "Acts of identity". They argue that "linguistic behavior [can be understood] as a series of acts of identity in which people reveal both their personal identity and their search for social roles" (Le Page & Tabouret-Keller 1985: 14). They also include the notion of group membership that had been important for Bell's understanding of intraspeaker variation, but more so in the sense of Coupland's "relational self" (Coupland 2007: 80): "[T]he individual creates for himself the patterns of his linguistic behavior so as to resemble those of the group or groups with which from time to time he wishes to be identified, or so as to be unlike those from whom he wishes to be distinguished" (Le Page & Tabouret-Keller 1985: 181).

The membership in groups allows for certain roles that the individual can take on. These roles are closely associated with the Eckertian understanding of style as "a clustering of linguistic resources and an association of that clustering with social meaning" (Eckert 2001: 123). She criticizes Labov for his predictive and unidirectional style interpretation. Instead of making assumptions about which styles to expect in a given speech situation, Eckert argues that every speaker has the individual ability to choose from a number of socially meaningful speech patterns to relate their own situational identity. The choice relies on the desired identity of the speaker in a given situation, not the speech activity or participant. Style according to Eckert should therefore be researched in relation to the social meaning that may be the reason for perceived stylistic variability in research (Eckert 2001: 126).

A further aspect that should not be considered speaker design, but which is closely connected to the concept of identity is stance. Kiesling sees his approach to this theoretical framework as a response to inductive approaches to style: "The premise that style derives from social group patterning should be reversed: I contend that stance is the main interactional meaning being created, and it is a precursor, or primitive, in sociolinguistic variation: that is, sociolinguistic variants are initially associated with interactional stances

and these stances become in turn associated with a social group meaning in a community over time and repeated use." (Kiesling 2009: 172). Stance is best understood as an individual negotiation of social meaning in a positional manner. Jaffe describes this positioning as being in "respect to the form or the content of one's utterance". Thereby, "speaker stances are […] performances through which speakers may align or disalign themselves with and/or ironize stereotypical associations with particular linguistic forms." (Jaffe 2009: 2). More specifically and in close connection with Eckert's understanding of the formation and usage of style, Du Bois describes stance as "a public act by a social actor, achieved dialogically through overt communicative means, through which social actors simultaneously evaluate objects, position subjects (self and others), and align with other subjects, with respect to any salient dimension of the sociocultural field." (Du Bois 2007: 163).

The stance dimensions Jaffe then describes as possibilities are concerned with "evaluation", "reflecting speaker's positionality" and "attributing position to others" (Jaffe 2009: 5). Other notions often employed to identify the different types of positioning moves are Kiesling's epistemic and attitudinal constructs, with epistemic stance referring to "a person's expression of their relationship to their talk (their epistemic stance – e.g. how certain they are about their assertions)", while attitudinal stance voices "a person's expression of their relation to their interlocutors (their interpersonal stance – e.g. friendly or dominating)" (Kiesling 2009, 172).

Although all the described approaches to style in this section deal with language as a means of identity formation, speaker-design, acts of identity approaches and the Eckertian view of intraspeaker variability are still rather closely connected to identity formation as a whole. The approach to style as stance described lastly functions differently in its inductive reasoning and the ability to develop stance categories from both the speaker's behavior as well as the speech activities being performed.

1.5 Meaning

As the research outlined in the subchapters above has shown, linguistic variability at its core entails that each speaker is presented with a choice of different options. Their individual choice of a linguistic variable, whether conscious or subconscious, may create a difference in how their utterance is being perceived in the context of their specific environment. In this sense, their choice indexes a certain interpretation. In the broadest understanding of the term, the process of choosing a variant and interpreting a variable on both sides of a communicative situation is the very process by which we create meaning in variability. Bakhtin made this connection early on when he stated "All words have the 'taste' of a profession, a genre, a tendency, a party, a particular work, a particular person, a generation, an age group, the day and hour. Each word tastes of the context and contexts in which it has lived its socially charged life" (Bakhtin 1981: 293).

Hay describes this taste as an association and then argues: "rich associations between social meanings and linguistic forms are actively harnessed by speakers to position themselves within their social landscape (Eckert, 2000). These associations are also actively used by listeners, both to interpret the linguistic signal (Drager, 2010) and to

exercise social classification and judgement regarding the speaker (Campbell-Kibler, 2010)" (Hay 2018: 697).

This linguistic creation of meaning is related to a number of concepts that I introduce in the following chapters. These include sociolinguistic salience, indexicality and enregisterment. Although there is some overlap among these concepts, I would argue that the factors involved in creating a linguistic choice on the side of the speaker are best described through the concept of sociolinguistic salience (Rácz 2013). Salience is concerned with the question of how we are able to attach meaning to a variable, and which variants are available to a speaker in this regard. Indexicality on the other hand is used to describe both how the speaker may use a meaningful choice and, on varying levels, how much awareness of this choice's meaning the speaker has (Silverstein 2003). Lastly, the process through which the social meaning of variability becomes attatched to the available variants is referred to as enregisterment (Agha 2005; Johnstone 2016). Similar to the theories on linguistic style, these concepts inform the analysis of metalinguistic comments and the discussions of patterns found in the production of speakers, even though the present study design cannot directly quantify meaning and salience in /str/-retraction.

1.5.1 Indexicality

At its onset, the study of variation was closely connected to the aim of finding meaning in the variable usage of language features. In his Martha's Vineyard study, Labov was concerned with the local construction of meaning in language through the analysis of variability in vowel production. In order to account for the differences between variables that show no potential of inducing a language change and those that promote change or are even consciously recognized as meaningful, Labov divided variables into a threefold categorization. First, *indicators* are those features that are used by all speakers of a certain variety without any conscious awareness. This results in a lack of differentiation between speech activities at varying levels of the formality spectrum. If /str/-retraction were an indicator, many speakers would use it without differentiating between more and less formal activities and even non-users would not notice it.

A *marker* on the other hand shows variable usage when speakers perform different formal language activities. While there is a social meaning attached to variables of this level, speakers within the respective speech community remain unaware of its meaningfulness. If /str/-retraction would be a marker, the variability within the community would be the same as with the indicator, but people outside the community would start to attach meaning such as localness or masculinity to it. The third possible level is described by Labov as a *stereotype*. As the connotation indicates, with stereotypical variables there is a strong connection between the social meaning of these variables and the awareness of said meaning. Speakers who use *stereotype*-variables are knowingly evoking a desired effect in the communicative situation. If /str/-retraction were stereotypical, community members would choose to use it to evoke a meaning knowingly or even mock the social meaning of the variable.

While they are closely connected to Labov's threefold categorizations that were developed through variationist observations, Silverstein's orders of indexicality are more focused on the reference frame of the used features and how it allows for differential

creation of sociolinguistic meaning. He explains his orders of indexicality as deviations from the presupposed state of n-th order. In it, a variable "presupposes that the context in which it is normatively used has a schematization of some particular sort" (Silverstein 2003: 205) resulting in a correlation of the feature with certain ethnographic features of the community in which a feature of this order is used. He refers to this order of indexicality as "scientific". He defines the step of further analysis of this feature (or deepened entrenchment in the respective context) by stating: "N+1st order indexicality is thus always already immanent as a competing structure of values potentially indexed in-and-by a communicative form of the n-th order, depending on the degree of intensity of ideologization" (Silverstein 2003: 194). The third level of indexicality he describes in turn presupposes the first order indexicality. In Silverstein's words: "The values of stereotypes are presupposed in the social-structure-as-indexed according to an ideological model, pure and simple; n+1st-order indexicality has become pre- supposing, in other words, in effect replacing an older n-th-order indexical presupposition." (Silverstein 2003: 220).

Johnstone et al. take these orders into account when they employ them in an analysis of Pittsburghese. To them, the stage of first order indexicality is characterized by unawareness of speakers, while at the same time being identifiable by community outsiders. As such, it can be described as "potential indexicality" (when identified) (Johnstone, Andrus & Danielson 2006: 83). Second order indexicality, in Johnstone et al.'s understanding describes the movement from feature usage to recognition. Like Labov, Johnstone et al. stress "speakers are not necessarily aware of second-order indexicality in such a way as to be able to talk about it" (Johnstone, Andrus & Danielson 2006: 84).

Eckert takes issue with the disconnect of linguistic variables from meaning in variationist literature. She argues in favor of interpreting social meaning through the lens provided by Silverstein and others, but expands the possible meanings of a variable in a more than one-dimensional space. Rather than connecting one variable with one meaning with one of the levels of indexicality, she develops a concept that maps the indexical potential of a variable in its entirety. I describe this concept in its own subchapter below.

1.5.2 Indexical Field

In the framework of the meaning-based studies introduced above, Eckert argues:

> "the meanings of variables are not precise or fixed but rather constitute a field of potential meanings – an indexical field, or constellation of ideologically related meanings, any one of which can be activated in the situated use of the variable. The field is fluid, and each new activation has the potential to change the field by building on ideological connections. Any one of which can be activated in the situated use of the variable. The field is fluid, and each new activation has the potential to change the field by building on ideological connections" (Eckert 2008b: 454).

The resulting concept of this understanding represents the pooled possibilities of all meanings attached to a variable in a specific speech community and environment. It can further portray how this plethora of meanings together influence the perception of a variable. In Labov's Martha's Vineyard analysis, it would allow a connection between the

changing social landscape and various notions of locality, rather than imprinting locality as the only meaning affecting the variability of the vowel phenomena.

1.5.3 Salience

Salience is described by Hickey as "a phenomenon which is generally recognized by linguists but which is notoriously difficult to quantify" (Hickey 2000: 57). Just how difficult this quantification is, depends both on the attempts to describe salience as a characteristic and the temporal logic connected to salience.

According to Rácz, the Labovian distinction between indicators, markers and stereotypes initially created the understanding that variability depends on how noticeable features are in language. However, Rácz cautions that researchers argue inconclusively, because "some imply that salient variables are selected as markers; others suggest that a variable becomes salient when it is selected as a marker" (Rácz 2012: 59)

Early attempts to quantify this variation in awareness of features were undertaken by Trudgill, who developed a set of criteria that describes indicators in opposition to markers. He argues that markers are characterized by:

1. having an overtly stigmatized variant
2. having a high-prestige variant (indicated in the orthography)
3. undergoing change
4. comprising a large phonetic difference
5. being contrastive. (Trudgill 1986; quoted from Rácz 2013: 27)

However, the fact that stigma is overt here while not overt in indicators means that indicators are just markers with decreased salience. Salience is thus an additional and gradual feature but not a distinction.

Auer et. al. take these criteria as a starting point and use them to develop further criteria and distinguish between those that they consider objective (language internal) and subjective (language external) (Auer, Barden & Grosskopf 1998: 167).

Objective criteria	**Subjective criteria**
Articulatory distance	Perceptual distance
Areal distribution	Usage in code-alternation
Phonemicity	Representation in lay dialect writing
Continuous vs. dichotomous lexicalization	Stereotyping/ mimicking
	Comprehensibility

They caution further that while criteria may successfully identify salient features, it is problematic that salience does not predict adoption or divergence of features. Kerswill and Williams further voice concerns with Trudgill's approach both due to the circularity of the stigmatization category and the unclear reasons for acceptance of a variable by speakers (Kerswill & Williams 2011).

MacLeod describes these approaches as "criteria-list approach[es] to salience" and contrasts them with experimental approaches. Experimental approaches use the magnitude of an effect to determine sociolinguistic salience of a feature. An example of such approaches is Fridland's matched-guise evaluation of the vowels used in regional classification of speakers. The accuracy rate of correctly identifying a vowel guise as Southern was used to argue that /e/ was the most salient feature in classifying this regional variety (Fridland, Bartlett & Kreuz 2004). MacLeod concludes that both approaches to

salience show disadvantages and may differ in usefulness based on the data and hypotheses at hand. While the criteria-list approach can successfully identify features relevant to salience, experimental methods can present a gradual notion of salience. The fact that both approaches aim at very different aspects and hypotheses of salience in phonetic accommodation is a first indication that salience is not defined and operationalized in the same way by all phonologists. The matter is complicated further once salience is described in morphology or syntax.

Hickey defines salience as the degree to which speakers are aware of a variant and uses the concept to illustrate pathways of language change. He argues that shared salience in large proportions of a speech community is crucial to language change (Hickey 2000). Maximal salience is thus directly connected to predictability (and thus stereotypical) language use.

A further definition problem with salience is foreshadowed above. If usage has an effect on salience, simple awareness cannot describe all instances of feature usage, because the first contact with a feature will differ from prolonged or reoccurring contact. The theories above however rely on the notion that salience is stable. Jaeger and Weatherholtz propose a clear distinction between "initial salience", which is purely related to the surprisal strength of an item, and "salience at later stages" (Jaeger & Weatherholtz 2016: 1). They define this latter salience as "the cumulative product of an individual's experience related to the lectal variant, including direct experience, as well as discourse about the variant" (Jaeger & Weatherholtz 2016: 1). While initial salience relies purely on awareness as proposed by many others, this distinction clarifies the difference between features such as perceptual distance and comprehensibility and social salience. It may also align with the understanding that salience is crucial to language change without simplifying the reason to surprisal value as Kerswill & Williams have previously criticized.

The definitions put forth by Jaeger and Weatherholtz also connect exemplar models of frequency to salience, by arguing: "while the surprisal of a lectal variant determines how much it "stands out," the frequency with which the lectal variant is observed increases the probability that the variant is perceived and learned" (Jaeger & Weatherholtz 2016: 3). Once the frequency is high, a variable gains sociolinguistic salience in language variation and change when its "(perceived) informativeness about social group membership" (Jaeger & Weatherholtz 2016: 3) is high. In contrast to her own interpretation, this is the salience that MacLeod described as experimental salience, because it relates to the scalar notions of informativeness of a variant. Thus, sociolinguistic salience is later stage salience, while salience as described in the criteria-list approach can show effects both due to initial salience and due to sociolinguistic salience.

This initial salience is similar to Rácz' concept of cognitive salience. He argues for a dual footing of salience in linguistic variation: In a first step, all variables contain cognitive salience in the sense that : "A segment is cognitively salient if it has a large surprisal value when compared to an array of language input (Rácz 2013: 37)". Thus, cognitive salience resembles other notions of salience in that it depends on context features, but is further complemented by the usage-based notion that every instance of variable usage is computed in relation to everything else that had previously been used by a speaker.

The cognitive dimension may be complemented by social salience which is described as the process of being "a marker of social indexation, becoming […] salient for the members of the language community" (Rácz 2013: 37). While all variables contain cognitive salience, he argues that social salience is optional. This social salience then is what creates the crucial difference between variables that are first order indexicals in contrast to variables that carry second or third order indexicality.

1.6 Summary Language Variation and Change

This chapter has provided an overview of the most prominent theories of language variation and change, and introduced the social fabric on which our understanding of these mechanisms is based. To evaluate whether the described retraction of /s/ in /str/-environments can be considered a change in the phonological make-up of English, we must understand both what characterizes a change in apparent time and which social factors have aided in explaining language variation as basis of changes in apparent time.

The most important categories that form the foundation of this understanding are largely influential groupings within society such as age, gender, ethnicity and socioeconomic status. All of these are taken into consideration in this study of /str/-retraction. The methodology section uses these categories to further define the social fabric of Austin and describes in detail how and why notions of social factors were used to develop the specific methods and analyses of this research project.

Before the state of the art and the specific make-up of the speech community under investigation here can be introduced, the phonological background of the variable as such needs to be established, which is the topic of the following chapter.

2. Phonetic and Phonological Background for Sibilant Analysis

The general aim of the study of sounds in language is to identify meaningful patterns with which we are able to produce and perceive speech. This general aim has historically been divided into the study of phonetics and phonology. While the former is usually said to be concerned with the smallest segments of language and the latter with the interplay of these segments, there are many areas with considerable overlap and the present study relies on concepts from both frameworks. This chapter first elaborates on what it means when we consider the initial sibilant of the cluster a phoneme and indicates the historical growth of this distinction. Based on that definition, I then introduce the phonetic concepts that are of importance for the three phonemes under investigation, which are those segmental analyses that concern the specific phonemes present in the cluster /str/. As retraction describes a tongue movement, one of these analyses concerns articulatory phonetics and the motoric formation of the consonants. These classifications are then enhanced by concepts from acoustic phonetics, namely the physical properties of the vocal tract and corresponding acoustic results. They aide in further illuminating the characteristics that the methodological decisions and measurement choices of this study are based on. A brief overview of theoretical developments in phonology ranging from generative phonology to exemplar theory provides the basis of investigations of gradience and lexical effects as properties of the sound change. These are complemented by an introduction to relevant processes of sound change such as assimilation and coarticulation. Lastly, the theoretical introduction to sound change in particular completes the theoretical groundwork on which the methodology and findings of the present study rests.

2.1 The phoneme

Generally speaking, the conceptualization of sibilants in /str/ as more or less retracted versions of a sound is based on the procedural logic of articulation. Once the retracted version is compared to the post-alveolar sibilant however, the distinction between these sounds as separate phonemes becomes questionable, because the concept of phonemes supposes a category of separable entities while the process of retraction expects a gradual spectrum. It lies in the nature of the problematic definition of phonemes that these differences arise. Throughout the establishment of phonetics and the phoneme as a concept, there have been various approaches to deliver an understanding for the distinction.

The problem is well described by Ladd, when he states that:

> "[…] language uses the [acoustic] medium of sound in a very specific way, which involves the human cognitive capacity for creating categories and symbolic systems. This capacity makes it impossible for two physical (acoustic) events that are objectively quite different to count as instances of the same category in the symbolic system, and for to physical events that are objectively very similar to count as instances of two different categories. It also makes it possible for different languages to categorize the physical events of speech in different ways. If we want to understand the medium of spoken language, therefore, it is not enough to consider only the physical aspects of the

> production, transmission and perception of sound; we need to consider the symbolic value of the sounds of speech as well." (Ladd 2014: 348)

This gap between human cognition and the physics of sound is at the heart of the problem with phonemes and phonotactics and the "symbolic value" of the sounds of speech. The problem in the theoretical framework begins with Saussure's differentiation between *langue* and *parole* (Saussure 1978), representative of spoken and cognitive parts of language. The history of phoneme theory continues down two paths. On the one hand, there are linguists who perceive it as a physical reality, arguing that articulation and acoustics are directly related in creating sound: "The speaker has been trained to make sound-producing movements in such a way that the phoneme features will be present in the sound-waves, and he has been trained to respond only to these features and to ignore the rest of the gross acoustic mass that reaches his ears" (Bloomfield 1984: 79).

Bloomfield argues in behavioral terms that normalizing speech sounds is a learned activity performed by speakers of a certain language. He also introduces minimal pair tests as experiments to test these learned phoneme features. The search for the concrete categorization of sounds as described above was at the core of the establishment of the International Phonetic Alphabet. The need for a concept such as the phoneme was therefore described as "There should be a separate sign for each distinctive sound; that is, for each sound which, being used instead of another, in the same language, can change the meaning of a word" (IPA 1999: 153).

The opposite side on the debate of the reality of phonemes are those theorists that think of the phoneme as a psychological concept or fictitious unit. One of the first to voice concerns with limiting properties of speech to "simple, mechanical terms" as "merely mechanical processes, consummated by the organs of speech and by the nerves that control them as a set of shifts in relatively simple sensorimotor habits" (Sapir 1925: 37) was Sapir. Sapir bases his arguments for understanding the phoneme as a "psychological" entity on two problems for phoneme categorization. For one, her argues that individual differences in sound production underline the necessary abstraction and thus normalization of sounds far beyond mechanics. Secondly, quantitative differences arise even where mechanical production may be the same. This is due to learned concepts creating a distinction in language such as for vowel length. These concepts are not objective and yet can be perceived in language relative to context, not based on measurable physical differences (Sapir 1925).

Twaddell evaluates both theoretical positions in defining phonemes: the psychological reality proposed by Sapir and the physical quantification based on shared features proposed by many others. He shows that neither theory can successfully describe the phoneme as a fundamental unit, one because variability in physical production is not entirely accounted for and the other because psychological abstractions cannot be tested and therefore not be used as classification techniques. To solve these issues, he proposes to think of the phoneme as an "abstractional fiction" (Twaddell 1935: 36).

Current debates on whether the phoneme is an artificial category or a meaningful distinction have focused more on the distinguishing features of these categories (distinctive features, feature specification and underspecification, markedness theory, contrast) and the

structure of mental lexica where the phoneme input is not immediately normalized in mental storage (exemplar theory) (Dresher 2011).

The state of the art will show that the focus in understanding /str/-retraction varies along with the (historical) conceptualizations of the phoneme. While Shapiro makes an argument based on feature specification, newer studies employ acoustic measurements to represent a continuous understanding of the variable, so that phoneme categorization is no longer an issue. This continuous and acoustic focus is further underlined by the reliance on exemplar theory to explain language change.

In the history of the term phoneme, there has been a gradual movement from treating language production and language storage as separate entities to defining perceivable features of phonemes based on the motor activity and creating the phoneme as a resulting prototype of these features. In the following, I define both the articulatory and acoustic characteristics of the phonemes in question here, before expanding the view to phonological aspects of cognitive language access and suprasegmental processes in language production.

2.2 Articulatory Phonetics

Articulatory phonetics is concerned with the distinction of sound units based on how we use our articulators (active and passive) to produce them. It generalizes tongue position and air movement for categorization and is the basis for the IPA as described above. When the focus is on the gestures and gestural timing of speech, Bowman and Goldstein differentiate between segmentation, a process based on the acoustic output, and gestures, which they understand as those movements of the articulators that produce sounds without any connection to output (Browman & Goldstein 1992). The reality of sound production is certainly one in which continuity and variability are at play. If we were to measure distances between lips and stricture, we would see a plethora of different values for each type of sound created by any speaker. However, as exemplified by feature theory (Mielke 2008), articulation can be divided into certain zones, which are then described as places of articulation (Ladefoged & Maddieson 1996).

2.2.1 Articulation

Consonants in articulatory phonetics are generally described as those entities that combine active and passive articulators to obstruct the airflow. The airflow can be obstructed in multiple ways: For the production of stop consonants, the airflow is blocked entirely. For approximants and fricatives, the airflow is impeded with varying degrees of air channel size. Nasals are those consonants for which the airflow is diverted through the nasal tract. Consonants are further characterized by the presence or absence of vocal fold vibration (Ladefoged & Maddieson 1996). The acoustic results of these characteristics and the way in which they are relevant to the acoustic measurements in the present study are discussed below. There are two different pathways to describe articulation, either through binary features/ natural classes (Mielke & Hume 2006) or on the basis of place and manner of articulation, as is the basis for the international phonetic alphabet. In line with newer approaches to describing the articulatory differences in /str/-retraction, I focus on a description of the relevant sounds in the latter framework below.

2.2.2 Articulation of /s/ and /ʃ/

Both sibilant alternants described in this study are fricatives, meaning they are created by moving the tongue into close proximity against a passive articulator, in this case the alveolar ridge or palate. The closeness of the articulators creates the friction that is characteristic for the manner of articulation described as fricative. Ladefoged and Maddieson mention that this specific constriction "has a greater degree of articulatory precision than that required in stops and nasals" (Ladefoged & Maddieson 1996: 137). This degree of precision over a perceivable time frame needed to perform a successfully perceivable fricative results in a "greater constancy of shape in varying phonetic contexts in comparison with the corresponding stops t,d, and n" (Ladefoged & Maddieson 1996: 137).

In English, there are two common articulatory patterns for /s/. The constriction is usually located on the fronter part of the alveolar region, which is the boney extension behind the front teeth. American English speakers have been found to produce the constriction with either the tip of the tongue (apical) or the blade of the tongue (laminal) at this place of articulation. These differences correlate with constriction length and place of articulation in that apical realizations of /s/ in Dart's study of 20 English speakers were both fronter and had longer constrictions (Dart 1991: 29). Dispersion-wise, laminal articulations were slightly more common at 52.5 percent. However, Dart argues that the variability in the production may not be entirely described by the laminal/apical distinction. She presents evidence that laminal productions of consonants may also include characteristics such as a high tongue body, a wider pharynx and a more closed jaw position than apical pronunciations and therefore argues that a distinction between apical and laminal should rather consider the tongue body (Dart 1991: 62).

Her study provides evidence that English /s/ may present with interspeaker variability in place of articulation, constriction length and apicality (Dart 1991: 150). Other individualized factors in /s/ production are "amount of protuberance of the alveolar ridge, and the relation between the lower jaw and the upper teeth" (Ladefoged & Maddieson 1996: 146).

Beyond these individual differences, there are general characteristics in the articulatory make-up of /s/ identifiable for all speakers. First of all, stridence, the close positioning of the teeth, is present in all productions of /s/. Secondly, Ladefoged and Maddieson describe a narrow groove in the posterior-anterior direction of the tongue as characteristic of /s/. In a comparison of two speakers they exemplify a groove in both raised and lowered tongue body realizations of /s/ (Ladefoged & Maddieson 1996: 147–49).

The place of articulation for /ʃ/ is described varyingly by phoneticians. The current version of the IPA (2018) categorizes it as a post-alveolar sound, which is located centrally in the alveolar region, involving the alveolar protuberance. It compares to the articulation of /s/ in being strident (close approximation of the teeth). However, the tongue shape and position characterize the differences between the two sounds. The constriction itself is located at the backer part of the alveolar region and is generally wider (which is relative considering the effect of laminal and apical production differences described above, with laminal productions yielding wider constrictions). The post-alveolar sibilant is further

differentiated by the rounding of the lips, which lengthens the resonator (results of protrusion will be described in the acoustics section).

The articulatory difference that Ladefoged and Maddieson stress the most is the shape of the tongue behind the constriction: the part of the tongue immediately behind the constriction is raised instead of the groove created for /s/, resulting in a dome-like structure with the sides of the tongue pointed downward from the center. This domed tongue shape is created regardless of apical or laminal articulation processes. Ladefoged and Maddieson describe the sibilant as palato-alveolar since the tongue shape creates a degree of palatalization. The heterogeneity in the specific part of the tongue (laminal or apical) actively creating the constriction is similar to the production of /s/. It can therefore be concluded that laminal and apical articulations of the sibilants are equally possible in English and are not meaning-distinguishing. A palatographic survey in 1957 reported that all participants differentiated between the articulation of the sibilants in the words *ship* and *sip* by creating a narrower channel in a (usually) fronter position for *sip*. A wider channel physically results in a lowered velocity with which the airstream hits the down-stream obstacle, an acoustic phenomenon that will be elaborated in the subchapter on acoustics below (Ladefoged & Maddieson 1996).

2.2.3 Articulation of /t/,/p/, /k/

Stop consonants are only considered here because they may possibly show coarticulation effects with the respective production of /s/ as will be described below. Typically, stop consonants are distinguished from an articulatory point of view by a complete blockage of the airflow. In contrast to the fricatives described above, the sound is much shorter and quick changing (Ladefoged 2006: 49). It is also audible in its articulatory movement in conjunction with a vowel rather than a steady state. The tongue moves into the blockage position for the articulation, blocks the airflow and then quickly releases the air-buildup behind the blockage.

The places of articulation for sounds relevant to the second position in the consonant cluster under investigation in the present work cover the entire range of possibilities for oral stop consonants in English. Phonotactically possible from front to back are the bilabial, the alveolar and the velar voiceless stop consonants. The bilabial blockage is created with both lips, the alveolar blockage with the tongue pressed against the alveolar ridge and the velar blockage with the back of the tongue against the velum. The same variability in part of the tongue used for production that was described for both sibilants above exists to produce the alveolar stop closure (Reetz & Jongman 2018). However, the percentages of apical production of the voiceless alveolar stop by Californian English speakers is higher for stops in Darts' study (66,67%) than for alveolar voiceless fricatives (42,5%) (Dart 1991). The research literature on Hungarian has shown that palatalized alveolar stops are created with a laminal gesture for those speakers while the plain stops are created with apical tongue body (Ladefoged & Maddieson 1996: 23).

2.2.4. Articulation of /r/

The class of rhotics can generally neither be defined by place nor by manner features. They are however phonologically interesting sounds, because they

> "often occupy privileged places in the syllable structure of different languages. They are usually the only consonants allowed as second members of clusters in the syllable onset, or as first members of clusters in coda position. More generally, we may say that in languages with consonant clusters, rhotics tend to occur close to the syllable nucleus" (Ladefoged & Maddieson 1996: 216).

Interestingly, this is exactly the case in the phonotactics of English, where /r/ is the only consonant that appears in triple consonant clusters with /s/ in syllable onsets. Importantly, /s/ is the only sound even able to be used in triple consonant clusters and is only connected to oral stops and rhotics.

For the pre-vocalic position in American English that matters in the present context, an approximant articulation of the rhotic is in either alveolar or post-alveolar position. Some speakers also produce a bunched /r/, which is produced with narrowed passages both in the lower pharynx and at the center of the palate without involving the tongue tip or blade. Some speakers narrow their lip opening slightly. Acoustically, both the pharynx constriction and the lip rounding result in a lowered third formant. A more or less retroflex pronunciation of this approximant has also been found for American English speakers, which shares the constriction at the pharynx and lips, thus also resulting in lowering of the third formant. Ladefoged and Maddieson conclude "that speakers of American English combine several articulatory mechanisms to produce a low third formant for whichever variety of this segment they employ" (Ladefoged & Maddieson 234).

2.3 Acoustic Phonetics

Acoustic Phonetics is the study of physical sound structure and the way in which in relates to speech production. Acoustics or more specifically sound waves are the product of the movement of the articulators and the vocal folds by each individual speaker. The findings in speech acoustics are grounded on a threefold theoretical foundation according to Harrington (Harrington 2010). Research in speech production conceptualizes the configuration of an idealized speech apparatus to deduce theories of the physical aspects of language sound creation, such as the theories developed by Fant and Stevens (Fant 1971; Stevens 1998). Other foundational work, described by Harrington as "Linguistic Phonetics", developed models to explain the connection between the acoustics/articulation interface, such as quantal theory (Stevens 1998). The third foundational effort lies in the study of language variation, which took concepts from both linguistic and acoustic phonetics to establish patterns of linguistic and social factors involved in speech production differences (Harrington 2010). I briefly describe the findings alluded to in the first two dimensions below. Social factors in language variation are outlined in the previous chapter and the relevant linguistic factors are outlined in the following section on sound change.

2.3.1 Linguistic Phonetics

In an attempt to combine the articulatory notions of segment production or gestural theories with the acoustic realities of sounds, a number of researchers have proposed theories of how the continuous realities of language production are successfully turned into perceivable segments (Fant 1971; Stevens 1998). This problem is also referred to as the "invariance problem" (Blumstein 1989), whereby the variance in acoustic measurements

of phonemes is surprising since the perception and segmentation of speech sounds is relatively invariant and stable. Attempts to cope with variability in production have shown that speech production is not only affected by the production mechanism itself, but also by context factors (Luce & Pisoni 1998).

The most physically oriented theory in this realm is Stevens' quantal theory. It combines the articulatory accounts that Chomsky and Halle base their research on with the phonetic approach by exemplifying how these features are representative of certain stable states in speech production (Chomsky & Halle 1968). He argues:

> "Since the articulators move within time, the sound sources and the filtering also change with time. Examination of the time-varying sources and filtering leads to the observation that some aspects of the transformation from articulation to sound are categorical. That is, the types of sound sources and the filtering of these sources can be organized into classes. These classes are closely related to the discrete linguistic categories or features that describe how words appear to be stored in the memory of a speaker or listener[…] An attempt is made to show that, when reasonable assumptions are made about the physiological parameters involved in producing a sound sequence, acoustic theory can make predictions about the sound pattern, and these predictions agree well with the measured pattern" (Stevens 1998).

Supposedly, these stable states are favored by speakers, because they offer a significantly decreased confusion rate as compared to other places in the oral tract. They can be understood as limited thresholds for optimal production and discrimination of sounds based on the physical reality of the (constricted) airflow in the tube that is our vocal tract. The remaining invariance problem can be interpreted in two ways: either as proof that no such categorization as underlying features exist in English or as evidence of predictable modification of features that are representative of the process on some level. Stevens and Hanson argue: "Because the modifications are systematic, the underlying discrete representation should be recoverable from the acoustic signal" (Stevens 1998: 420).

In greater detail, the argument is that articulator-bound gestures such as the place of articulation relevant to the phoneme difference in the current study have non-continuous magnitude of impact on the acoustic signal. This means that the shape of the vocal tract is such that a slight movement on the alveolar ridge has less impact on the acoustic signal than the change from the flat and boney surface of the alveolar ridge to the grooved and hollow space after it. In other words, a millimeter on the alveolar ridge does not have the same impact as the millimeter between alveolar ridge and palate. The difference between sibilants is therefore created at precisely the space in the vocal tract where "there is a rapid change in an acoustic parameter for a relatively small change in the articulatory parameter" (Stevens & Hanson 2010: 429).

In the present case, quantal theory would indicate that there is a certain production threshold within which speakers normally produce /s/ and /ʃ/, which would explain why there is no further subdivision between alveolar and post-alveolar production, or anterior and posterior. This is a necessary condition if we want to consider a change from one state to another. However, while this threshold is certainly at play, it negates many of the context factors that have also been proven to impact sounds, even in the case of /str/-retraction

where medial position has been shown to pattern differently than initial position (Wilbanks 2017). Another problem that quantal theory cannot provide satisfying answers to is individual differences and thus our ability to perceive one speaker's alveolar sibilant as /s/ while a similar sound wave is interpreted as /ʃ/ for another speaker. Stevens and Hanson describe these individual differences in production as "enhancement" and argue that these modifications do not deny the underlying schematicity of the process (Stevens & Hanson 2010: 447).

2.3.2 Acoustic Theory of Speech Production

Above, I introduced both the articulatory dimension of speech production and gave an overview of how these articulation based theories of place and manner of articulation (IPA 2018) and distinctive binary features (Mielke 2008) have been connected to acoustic analyses of speech (Stevens 1998; Stevens & Hanson 2010). In the following sections, I focus on the relevant concepts of acoustic theory and then apply them to describe the acoustic structure of fricatives and their technological quantification. Since the measurements in the study do not measure the acoustic patterns of the other sounds involved in the consonant clusters, I did not include a section on the acoustics of these sounds. The relevant aspects for a phonological analysis have already been delivered in the articulatory section.

The aim of acoustic research is to appropriately describe sound waves. The general term "sound" describes two different types of signal that are both relevant to speech production: periodic and aperiodic sounds. Simple periodic signals are also called sine waves because of their even shape. The motion that propagates periodic waves is entirely harmonic in the sense that it disperses air in an even, wave-like movement. The defining characteristics of any simple periodic wave are its frequency, its amplitude and its phase. Frequency is defined as the number of times an entire cycle of the wave if performed at a given time from the point of zero to its highest, lowest and back to its zero point. This zero-point is represented by the x-axis on a graph. The number of cycles that are completed per second are expressed in Hertz. It has to be noted however that frequency is a linear representation of sound that reflects dispersion correctly but is not representative of the way in which we perceive sounds (Johnson 2003: 7).

Each peak or maximum distance from the zero-point is representative of the amplitude of a sine wave. The third characteristic, phase, describes the progression of a sound wave relative to a certain point in time. Phase is of particular importance when a complex wave is being described, as some waves may differ just in that respect (Johnson 2003, 8).

The difference between a simple periodic wave and complex periodic waves is their general make up. A complex periodic wave is made up of two or more simple periodic waves. If it becomes necessary to calculate all simple waves combined to create a complex wave, the transformation developed by Fourier in the 17th century can be used to derive its frequencies, amplitudes and phases (Johnson 2003, 12).

The other type of sounds, aperiodic sounds, differ from periodic sounds by lacking a regular repeating pattern. There are two types of aperiodic sounds: continuous aperiodic sounds and transients. When performing a Fourier analysis on a continuous aperiodic

sound, there will not be any sharp peaks but rather equal amplitude for the possible frequency components. Speech sounds that would acoustically be characterized as continuous aperiodic sounds are for example fricatives, the type of sound relevant for the present study. The other type of aperiodic sounds are transients. In contrast to continuous signals, these are sudden differences in air pressure without sustained length such as sharp bursts or sounds created by objects falling or breaking.

The phonemes relevant in the present study are all consonants and can very generally be classified as aperiodic sounds. Sibilants are continuous aperiodic sounds, the stop is a transient and the /r/ is a continuous aperiodic sound that includes some characteristics of a complex periodic sound wave and is thus a sonorous consonant.

While the explanations above provide a good overview about the types of signals we deal with in this study, they do not yet account for differences in articulator movements as they were categorized in the section on articulatory phonetics. This will be accounted for by source filter theory section below.

2.3.2.1 Source filter theory

Resonance in speech production is the way in which the oral cavity functions as a filter of certain wave-lengths. It is theoretically described by the source-filter theory that was developed by Fant (Fant 1971). This theory explains how the complex wave perceived as speech is initiated at the vocal folds, but crucially shaped by the respective shape of the vocal tract and further influenced by the radiation properties of the environment once it exits the oral tract at the lips. It thus connects articulation with the physics of sound propagation, by providing an explanation for how a single origin of an acoustic signal can result in such a plethora of different sounds. Resonance, more generally, is the preferred oscillation frequency of any given object. This functions as a filter of the complex signal we create when producing speech. The types of waves filtered by the vocal tract are based on its properties as a tube. For speech production, the tube as such can either be open at one side or open at both sides. In a tube open at one side, the air pressure at the opening will be the same as the air pressure outside of the mouth. This is referred to as the node. At the opposite side, the air pressure is either at its highest point (when closed) or at its lowest point again (when open). The point of maximal change in pressure is described as an anti-node, which relates to the point of maximal change present in Stevens' theory (Reetz & Jongman 2009).

Depending on these nodes, different types of waves are amplified. For a tube open at one end, 1/4 of the wavelength may fit into the tube and thus exactly a quarter of the wavelength may fit into this type of resonator. However, it also allows for 3/4 or 5/4 or the wavelength to be amplified. This is the case because the odd multiples of the respective wavelengths also fit the tube. The second resonance (half of the first resonance) thus has two nodes and two anti-nodes that will affect that wave in different ways. Fricatives, much like vowels, are the result of a resonator opened at one side, a so-called quarter wavelength resonator. Constrictions within the tube can slow down or speed up the movement of the molecules. If there is a constriction at a node, the molecules move at decreased speed, thus resulting in lower frequencies. A widening at a node will increase the speed and result in higher frequencies. Widening or constriction at anti-node result in a reversal of these

frequency characteristics, meaning a constriction results in higher frequencies and a widening results in lower frequencies (Reetz & Jongman 2009).

2.3.2.2 The Acoustics of Fricatives

Fundamental to the creation of fricative sounds is the manner in which the air flows through the oral cavity as a resonator. It is the acoustic consequence of the speed and volume of air that is being moved. The chaotic aperiodic sound acoustically created by this practice is a result of the fast-moving air exciting the oral cavity and hitting inert outside air. The amplitude or relative acoustic energy of fricatives is a product of velocity and turbulence. Velocity is the speed with which, influenced by the size of the resonator, particles may pass. If there is a great volume of air forced through the oral tract as a channel, it increases the speed of air molecules due to the increase in pressure. The smaller the constriction in the mouth the more amplified the sound will be. The energy of the fricative is further enhanced by the turbulence of air created when hitting a down-stream obstacle. For the fricatives of interest here that down-stream obstacle are the teeth. This difference warrants the distinction between sibilant and non-sibilant fricatives (Fant 1971; Stevens 1998; Johnson 2003).

The frequency of individual fricatives is dependent upon the length of the cavity after the constriction. This is a result of resonance as described above. The voiceless fricatives /s/ and /ʃ/ are created by the velocity of the airstream forcing friction at the tip of the tongue. This frication noise is the main source of acoustic energy of these sounds and results in an aperiodic signal. The second obstacle created by the teeth results in "clear, distinct spectral shapes" (Reetz & Jongman 2009: 191) for both the alveolar and the post-alveolar fricatives.

The difference between /s/ and /ʃ/ is acoustically represented by an energy concentration at lower frequencies for the latter. The lower frequency is a result of the longer anterior cavity. Due to coupling, the posterior cavity has no significant influence on the acoustic profile of these sibilants (Stevens 1998). But /s/ and /ʃ/ differ in more respects than just the retraction of the tongue. Research has shown that speakers of American English also differentiate between the two sounds by creating a sub-lingual cavity for /ʃ/ which further lowers the frequency (Shadle, Proctor & Iskarous 2009). The acoustic differences are further enhanced by lip rounding or protrusion (Shadle & Scully 1995). According to Ladefoged, this lowered frequency is also enhanced by a wider channel for the production of the post-alveolar sibilant that decreases the velocity of the molecules (Ladefoged 1993). The fact that there are large cavities involved for both sounds not only influences the frequency range of the energy concentration but the concentration of energy at a more defined frequency than for fricatives with smaller cavities such as /f/.

The differences between laminal and apical production of the alveolar sibilant as they were investigated by Dart show differences in the spectral shape of the sound. Acoustically, the spectral shape of alveolar productions of /s/ shows a peak around 500Hz while the spectral shape of laminal /s/ is relatively flat. Postalveolar laminal realizations of /s/ show a lower frequency peak than alveolar realizations. This same pattern holds true for apical realizations in French, while there are no apical post-alveolar realizations in her

English sample. Although both alveolar realizations show an increase in energy at higher frequencies, the increase is earlier and steeper in apical realizations.

The different aspects that create the frequency patterns perceived as alveolar or post-alveolar fricatives are relatively speaker-dependent (Shadle & Scully 1995; Shadle, Proctor & Iskarous 2009). However, these differences are consistent within the speaker. Thus, irrespective of the degree of lip-rounding, exact tongue positioning or sub-lingual cavity creation, speakers are consistent in the pattern they choose to produce the acoustic difference between the two voiceless sibilants.

Ladefoged and Maddieson describe the highest energy concentration for the alveolar fricative of an English-speaking male to be located at approximately 5 kHz, while the highest energy concentration for the post-alveolar lies at 3,5 kHz (Ladefoged & Maddieson 1996). Averaged across both female and male speakers in their study, Jongman et al. report a spectral mean of 6,1 kHz for /s/ and 4,2 for /ʃ/ (Jongman, Wayland & Wong 2000).

2.3.3 Acoustic Quantification of Fricatives

In attempts to establish a correlation between acoustic patterns and frequency distributions of fricatives, several methods have been established that may define classification of place within the acoustic signal. Forrest et. al. were the first to show that spectral moments could be used to classify place of articulation differences in a study with 5 male and 5 female speakers. These findings were tested by Jongman et al., who focused on several measurements previously used in attempts for acoustic description of place in consonants. Since the sibilants in focus of the present study are not followed by vowels, some of his quantification techniques such as relative amplitude and locus equations are not of interest and will not be described here. In Jongman et al.'s research, both spectral peak location and all spectral moments distinguish /s/ and /ʃ/ (Jongman, Wayland & Wong 2000). The measurements of spectral moments are reported in table 2.1 below.[5]

Table 2.1: Mean spectral moment values for each place of articulation, averaged across speakers (20 students, balanced for gender), window location, voiced and voiceless tokens, and vowel context.

Place of articulation	Spectral mean (Hz)	Variance (MHz)	Skewness	Kurtosis
/f,v/	5108	6.37	0.077	2.11
/T,D/	5137	6.19	-0.083	1.27
/s,z/	6133	2.92	-0.229	2.36
/S,Z/	4229	3.38	0.693	0.42

All spectral moments distinguish the differences well, although variance delivers the least promising results judging from the table. Jongman also reports that kurtosis and variance do not establish a difference between sibilants in all windows, which cannot be seen in table 2.1.

Tabhain's study delivers similar results for the spectral mean (which she refers to as the center of gravity). She further shows that the spectral center of gravity is correlated with the center of gravity of the electropalatographic measurements, a further case in point

[5]Modeled after Jongman et al. (Jongman, Wayland & Wong 2000: 1257)

that acoustic invariance does prohibit acoustic quantification of sound segments (Tabain 2001).

In the present study, the acoustic profile of /s/ may be influenced by the fact that participants were not recorded in a soundproof booth and recordings are therefore expected to include non-speech sounds. In spectral analysis, we usually focus on certain ranges in the frequency of sounds as they appear most characteristic. For sibilants, we focus on higher frequency ranges as these types of aperiodic signals show the highest energy at these levels. One way to deal with this is to apply a regular filter to parts of the signal so that higher or lower frequencies than the ones of interest are cut out. Another means of dealing with this without direct filtering is to emphasize relevant parts of the signal, which is achieved by pre-emphasis. Rather than filtering the signal, this process amplifies certain frequencies (Johnson 2003, 41). In the present study, pre-emphasis will be applied to the tokens prior to the acoustic measurement of spectral means.

2.4 Suprasegmental Theories and Fundamentals

I have thus far focused on the phonetic approach to both the articulation and the acoustic characteristics of the speech sounds relevant to this dissertation. However, the fact that the focus lies on the influence of other consonants on the acoustic characteristics of /s/ shows that a purely segmental approach cannot entirely account for the variability observed. The missing theoretical puzzle pieces concern suprasegmental questions. I first describe exemplar theory as a model of language sound cognition. Based on this model, I then explain linguistic processes of sound change that have been used for explanations of the phenomenon in previous research. These include the phonotactics of triple consonant clusters, assimilation and coarticulation.

2.4.2 Exemplar Theory

Exemplar theory is a general attempt to account for "phonological generalizations and the coexistence of gradience and categoriality in phonological knowledge" (Johnson 2005: 40).

Exemplar concepts of memory storage are no invention of linguistics, but are instead based on memory studies such as Semon's 1923 account on sensory experience, in which he states: "For the sensation engendered by an original stimulation, say a peripheric stimulation, one coming from the outer world, finds in the mental organism not only a receptive response but, what is much more, a nourishing matrix of images, of *engrams*, left by previous sensations." (Semon 1923: 11). This engram as image of a outer world sensation is grouped within this matrix, so that "Every moment of individual life adds something to the already existing sum of simultaneous engram-complexes" (Semon 1923: 171).

How these "engram-complexes" apply to language was first investigated by Goldinger. He was one of the first to experiment with speech under the hypothesis that the mental lexicon is not an abstract storage of "ideal, modality-free units" (Goldinger 1996: 1166) and thus developed a model of language memory that described Semon's "engrams" as "traces". He concludes from his voice experiments, that "words are recognized against a background of countless, detailed traces. Speech is not a noisy vehicle of linguistic content; the medium may be an integral dimension of later representation." (Goldinger

1996: 1180). The notion of countlessness was later addressed by theorists such as Johnson by providing weighing mechanisms for attention and association in speech processing (Johnson 1997).

The idea about the unfiltered storage capacities of the mental lexicon was broadened by other theorists and creates the juxtaposition between abstractions at the interface of phonetics and phonology and episodic memory. In an attempt to further specify these differences and provide a concrete example of how "traces" become useful, Pierrehumbert refers to all traces as "exemplars". She states that "people have detailed long-term memories of particular percepts, and these are stored as locations on the map" (Pierrehumbert 2002: 113). These exemplars are for examples all memories of the word STREET that a listener encounters.

Exemplars define the individual regions of each and every segment. Whenever a new stimulus is decoded, it is classified based on the densest exemplar region on the map. "The basic insight, which appears to originate with work on motor control by Rosenbaum et al. (1993), is that activating the group of exemplars in a subregion of the perceptual map can specify a production goal which corresponds to the aggregate or average properties of the members of the group" (Pierrehumbert 2002: 114). As a result, speakers learn implicit frequency distributions over phonetic outcomes, these distributions are stored in long-term memory, and they are subject to incremental updating. These phonetic outcomes and thus the representational level of speech within the mental lexicon are argued to be auditory and not articulatory, which makes acoustic quantification of sounds all the more appropriate in the present case (Coleman 1998). The exemplars of the word STREET are thus constantly added to and frequency influences probabilistically which outcomes are possible and how likely they are to occur.

A good example of the categorization and storage process in exemplar theory is provided in visual terms by Lavie. In his example, the decision on whether an input variable is an /ɛ/ or an /i/ is based on the strength and number of exemplars in the periphery of the acoustic input, where the spread of exemplar realizations is visible on the x-axis and the exemplar strength on the y-axis (Lavie 2007). Exemplar strength directly relates to the frequency of exemplars, whereby those exemplars that are frequently activated involve less computation and remain in storage, while infrequent exemplar eventually fade (Bybee 2002). Depending on where on the x-axis the concrete realization of the phoneme is located, the neighboring strong exemplars will influence how it is being categorized.

The structure of exemplars is described in multiple ways in the literature. Some theorists argue for the storage of items of speech as exemplars (Johnson 1997), while others believe that exemplars are contextually rich allophonic representations that build the basis for further phonological generalizations performed by the speaker (Pierrehumbert 2002). It is important to stress that all models "assume that we store phonetically and contextually rich memories of our past linguistic experiences, which we then access in speech production and speech perception" (Hay 2018: 697).

Hay extends the concepts of exemplar theory to other contexts and thus connects it to notions of indexicality, when she states:

> "If detailed phonetic information accumulates at the level of the individual word, and if phonetic implementation varies across context, then we might

> predict that a word's production and perception will be influenced by the full range of contexts (linguistic, social, environmental) in which we have previously encountered it. There is increasing evidence in support of this hypothesis." (Hay 2018: 702)

She supposes that the incremental updates and stored information per exemplar not only include the linguistic shape but also other environmental information. This would mean that the exemplar cloud not only includes the many acoustically possible productions of STREET and their probabilities of appearance but also connects these varied productions to context factors. These factors include i.e. knowledge like retracted versions of the word are often occurring in fast speech or young women are more likely to produce one version than another.

2.4.3 Phonotactics of English Consonant Clusters

To better understand both status and particularities of the specific consonant cluster researched in the present study and other clusters mentioned in the literature on retraction, it is necessary to briefly outline the structure of phoneme sequences in English. In line with the usage-based approach of exemplar theory, these structures should be understood as probabilistic rather than rule-guided (Välimaa-Blum 2009).

First, there are diverging views of what a consonant cluster is and how it should be understood linguistically. While the general view may be that consonant clusters are any occurrence of multiple consonants in one word, Bauer differentiates between two types of multiple consonant occurrence based on suprasegmental considerations. In her definition, consonant clusters are those appearances of multiple consonants that occur within the syllable and are thus more phonotactically regulated. Consonant sequences on the other hand are those appearances of multiple consonants that occur across syllable and word boundaries (Bauer 2014).

Double consonant clusters in English are usually a combination of stop or fricative consonant and a more sonorous approximant or nasal. Although it does occur in some words (e.g. Sri Lanka), the combination of /s/ and /r/ rarely exists in English and is often pronounced with an initial post-alveolar sibilant despite the spelling (Rutter 2011; Bauer 2014: 450).

Most triple consonant clusters in English begin with either /sp/, /st/ or /sk/. This has started a debate on the singular nature of these sequences. The quality of the second consonant in clusters /sp/, /st/ and /sk/ is indistinguishable between voiced and voiceless stop, which would allow a reanalysis as monosegmental, similar to an affricate. Davidsen-Nielsen argues against such reanalyses because the segments function differently in syllable-final position (Davidsen-Nielsen 1975). A further point of discussion is the nature of the stops in these clusters. Davidsen-Nielsen reports that the underspecification of the voicing of the stop in these clusters has fueled the debate as to whether these should be considered specific phonemes or archiphonemes. In the case of /s/-retraction, there has been a similar debate concerning the acoustic perception of /s/. Since there are no naturally occurring triple consonant sequences with the post-alveolar fricative at the phonological level (c.f. Bauer 2014: 451), there have been some proposals that the sibilant in these cases

may be underspecified for place, and the mere presence of turbulent air is perceived as /s/ (Janda & Joseph 2003b; Rutter 2011).

The investigation of The Cambridge English pronouncing dictionary 2003 shows that onset triple consonant clusters only exist in a varied combination of /s/ + oral or nasal stop + approximant. The impossible combinations are limited to /spw/ and /stl/. The clusters including /w/, though theoretically possible, are not apparent in any word examples from the dictionary. In terms of frequency of syllable initial clusters for the dictionary, Bauer reports that /st/ is in the largest group of clusters with more than 500 occurrences. In contrast, /sk/ appears 21-100 times and /sp/ occurs between 11-21 times. Without further details on the triple consonant clusters, this can serve as a first indication of the varied dispersion of these clusters that will warrant frequency analyses in the present study.

2.5 Phonetic and Phonological Processes involved in /s/- retraction

The phonological and phonetic descriptions of the individual phonemes involved in the cluster under scrutiny here as well as the underlying phonotactic concepts allow for certain patterns of change as described below. I first differentiate the process of coarticulation from assimilation to then introduce other influential concepts such as consonant harmony.

2.5.1 Coarticulation

Coarticulation is understood as the result of timing differences in the motor processes involved in producing speech. In spoken language, the underlying segments the articulators aim for are not created in perfect succession but rather overlap and influence one another in all factors involved in speech production, namely the muscular movements of lips, tongue, velum and larynx (Farnetani & Recasens 2010). The existence of coarticulation has been proven both in articulatory terms with EPG measurements (Hardcastle 1985) as well as in acoustic terms (Farnetani & Recasens 2010).

The reasons for the process of coarticulation however are highly dependent upon theories of speech production, speech perception and the neurolinguistic storage and planning of speech segments. Theories on coarticulation lie on somewhat polar ends of a spectrum between generative and articulatory phonology. The generative phonology approach in the definitions of Chomsky and Halle describes phonetic features as those "properties of the signal that are supplied by universal rules […] [including] the different articulatory gestures and various coarticulation effects- the transition between a vowel and an adjacent consonant, the adjustments in the vocal tract shape made in anticipation of subsequent motions, etc." (Chomsky & Halle 1968, 295).

The term universal rules here refers to the phonetic component of Chomsky and Halle's overall understanding of the mental representation of language. It further connects to the notion of language universalism creating articulatory systems inherent in human language production in general with no differences between different languages. Importantly, this notion of coarticulation assigns speech production as a factor in this process while denying any influence of production planning and bottom-up knowledge transfer.

In later interpretations of coarticulation as a process, it is either used synonymously to assimilation (Ohala 1993a) or differentiated in terms of physical properties as opposed

to phonological processes (Wood 1996). Exemplary of this type of differentiation is Wood's definition of coarticulation as "local articulatory adjustment of all phoneme instantiations to their current neighbors" (Wood 1996: 139).

Solé further extends the concept that there are different processes at work creating harmonic pronunciation: He shows that nasalization can be both language specific and universal. The universal nasalization was affected by changes in speech rate and thus arguably affected at the production level, while other instances of nasalization showed no such effect and were thus argued to be impacted at the phonological (planning) level. The latter type was further shown to be language specific (Solé 1992).

2.5.2 Assimilation

Assimilation generally describes the process by which a phonetic segment becomes more like one or more of the segments surrounding it. Like coarticulation, it is thus context dependent. Whether it is considered synonymous or not has been heavily debated as indicated above. Generative phonology clearly differentiates between the two based on competencies involved. To Chomsky and Halle, assimilation is rule-guided in the sense that the categorizations in a speaker's linguistic competence are necessary for these processes to occur (Chomsky & Halle 1968). It is therefore based on individual phonological rules and language- specific grammar (Farnetani & Recasens 2010). In contrast to this rule-based conception, coarticulation in these terms would be a result of linguistic performance because it "results from the physical properties of the speech mechanism and is governed by universal rules" (Farnetani & Recasens 2010: 320). As an example of coarticulation, Chomsky and Halle consider "the transition between a vowel and an adjacent consonant, the adjustments in vocal tract shape made in anticipation of subsequent motions, etc." (Chomsky & Halle 1968: 295).

Ohala and Solé argue in contrast that "common cross-language sound patterns arise from the universality of physical phonetic constraints" (Ohala & Solé 2008: 340), not from phonological universals.

When assimilation is assigned a functionality beyond coarticulation effects, there needs to be a language-inherent need for these processes. Attempts to explain these needs are based on two concepts. On the one hand, assimilation has been described in phonological theory by optimality theory. Otimality theory has been developed to connect generative notions of phonological universals with the variability in output. In terms of assimilation, it may explain how universal language segments end up in differing phonetic manifestations depending on the respective language. This factor has been described in terms of assimilation by McCarthy as "homogeneity of target/heterogeneity of process" (McCarthy 2007).

In terms of language change, co-articulation arguably plays a role in both production and perception of reduced speech. This connects the levels of production and planning by arguing that the accumulation of changed exemplars due to coarticulation results in a recalibration of characteristics and thus a phonologization of the observed patterns. Mielke and Hume further posit: "If assimilation is the result of phonologized co-articulation (Baudouin de Courtenay 1972 [1895]; Ohala 1993; Blevins 2004), then an account of co-articulation and its phonologization could possibly account more directly for

sound patterns interpreted as feature spreading" (Mielke & Hume 2006). It is therefore useful to distinguish the physiological process from any other impactful factors.

However, exemplar theory negates the existence of a rule-governed universal grammar. A strict description of assimilation as a purely cognitive process in opposition to coarticulation oversimplifies the ties between perception, planning and production of speech. The most useful concept, therefore, appears to be one that conceives of feature spreading as a process that begins with co-articulation and ends in assimilation, without ascribing a stage in the pattern to all speakers. The observation of non-categorical choices within an ongoing change thus becomes no more than varying levels of exemplar remodeling. It offers the ability to consider the gradual impact of coarticulation as one with cognitive impact from the beginning without necessitating a categorical switch from motor-activity to cognitive concept.

2.5.2.1 Long-distance assimilation

The problem with the present analysis of assimilation is the assumption that it may be "long-distance" assimilation to /r/, which conflicts with any linear representation of coarticulation or assimilation. Several phonologists have analyzed processes of assimilation that extend the understanding of which aspects of phonological context may lead to alternations in speech production (Shaw 1991; Hansson 2001; Rose 2011; Rose & Walker 2011).

More specifically, they analyze processes of production alternations where those context items that have an apparent influence on the respective item have an effect across other segments.

The process of long-distance assimilation rests on the concept of phonological harmony, which refers to the process by which featural similarities among phonological segments are increased. In contrast to vowel harmony and vowel-consonant harmony, consonant harmony is not limited to immediate context but has been found to spread across several unaffected units. It is therefore referred to as long-distance assimilation (Hansson 2001; Rose & Walker 2011). If we assume that /s/ shares more similarities with /r/ than with /t/, it would warrant the conclusion that /t/ is unaffected by the process.

There are several theoretical assumptions both about possible unaffected units and default directions of assimilation. Hansson has argued for regressive assimilation (leftward) to be the default position and for consonant harmony to only spread across unaffected vowels (Hansson 2001). In line with the trends we observe in /str/-retraction, where consonant harmony is expected to leave the stop unaffected, Rose & Walker provide evidence conflicting with these claims (Rose & Walker 2011).

Rose and Walker discuss the case of sibilant assimilation in Ineseño Chumash. In one of these examples, the final sibilant affects the word initial sibilant, such that /s-ixut/ (it burns) remains alveolar [s-ixut], but /s-ilakʃ/ (it is soft) assimilates so that the initial sibilant is produced post-alveolar [ʃilakʃ]. (Nolan, Holst & Kühnert 1996; McCarthy 2007; Rose & Walker 2011: 242). In these examples, the harmonic process spans the entire word, including both vowels and consonants, so that initial or sibilants in initial syllables match the last sibilant.

How these features are shared is based on phonological theories and the related understanding of articulation. It was first described as autosegmentation based on featural geometry that orders the above-described features hierarchically and identifies characteristics for which phonemes are underspecified. Where this underspecification exists, phonemes can become transparent in assimilation so that they are not affected by changes occurring in other phonemes (Clements 1975; Halle, Vaux & Wolf 2000).

In contrast, Rose & Walker propose a disconnect from the cognitive load necessary to argue for feature-based analyses of assimilation and rather center on speech production. They report establishing successful models of assimilation by employing similarity matrixes based on natural classes (Rose & Walker 2011).

2.6 Summary: Phonetic and Phonological Theories for /str/-retraction

This chapter has outlined defining features of the sounds that will be investigated in the present sociophonetic analysis. I started by distinguishing those concepts that are phonetic in nature from suprasegmental accounts of coarticulation or assimilation. The consonants involved in this investigation were first introduced in articulatory terms. I then described the mechanics of speech production and the specific settings that result in the production of sibilants. I also briefly illustrated how the acoustic profile of the sound that this setting creates can then be quantified. Next, I dealt with cognitive models of speech storage and production. Within the framework provided by this cognitive understanding, I then introduced the specific theories of coarticulation, assimilation and consonant harmony that have been employed in analyses of /str/-retraction in previous studies. This chapter is knowingly lacking further theoretical explorations such as feature theory, feature geometry, natural classes and markedness theories. The reason for these omissions is the fact that using exemplar theory as foundation of the current approach and also more recent studies published on /str/-retraction renders many of these theories only marginally relevant. The third theoretical component that a sociophonetic approach to /str/-retraction needs is the theory of sound change, which will be provided in the following chapters.

3. Sound Change

I addressed the general mechanisms of language change and how they relate to social factors above. In this chapter, the specifics of sound change are organized from origin to initiation and beyond. I begin with theories of why sound change exists and discuss the teleological and non-teleological approaches. In the second section, models of how sound changes progress are connected to these varied initiation stages. From a micro view of how change happens in the individual speaker or listener, I further elaborate the overarching question of how sound change leads to phonologization as its final stage.

The different roles of speakers and listeners in acoustic speech production are crucial to the theories and notions that are outlined here. While variability in production and individual differences in speech unite all assumptions of how language change progresses, acoustic speech production can be studied and measured continuously in contrast to for example syntactic change. Sound change further lends itself to a more mechanical interpretation of cause than other types of change. The resulting notion is that a reinterpretation of patterns (change) could be a part of the communicative situation rather than an invention or a maximum performance of individuality on the side of the speaker.

3.1 Initiation of Sound Change

Discussions on causes for sound change often treat the process as a purposeful mechanism, which is problematic because of its teleological approach. Ease of articulation is one of the strongest arguments when it comes to teleological explanations of sound change. It assumes that sound change is a result of the aim to reduce articulatory effort while preserving the least necessary amount of communicative clarity.

Lindblom developed his hyper- & hypospeech model (H&H) following the intuition that speech production functions similarly to all other motor processes of the body, which means that both economy and plasticity influence speech production. He defines economy as "low-cost form of behavior" (Lindblom 1990: 412) in any motor system. It is this economy that directly relates to the idea of ease of articulation. In his model, economy creates hypospeech in instances where the context aids in contrast and creates a situation in which articulatory effort is directly connected to the notion of "sufficient contrast" (Lindblom 1990: 405) as evaluated in each situational context by the speaker. Hyperspeech on the other hand is realized in instances, where output-control is perceived as necessary due to contextual factors. This ability is a result of plasticity, namely the ability to adapt as desired to a given situation, be that a physical obstruction such as a bite block or social context. He defines these different hyperspeech factors as "production constraints" and "reception constraints" (Lindblom 1990: 419). The theory thus accounts for the invariance problem by explaining how and why speakers can vary their production. He stresses that "speech perception is the product of both signal-driven and signal-independent information" (Lindblom 1990: 431), which unites production and perception theoretically. This is famously described as a tug of war (Pouplier 2012).

While ease of articulation in sound change may easily explain processes such as assimilation, it is burdened with theoretical shortcomings. Not only is it circular in naming articulatory effort both as the source of a change and its outcome but it also fails to explain

many other aspects of historical sound changes such as fortition and dissimilation (Pouplier 2012; Thomas 2016).

This theoretical shortcoming is further supported by findings on "metabolic cost" in both speech production and biological effort. Metabolic cost refers to the notion of efficiency in energy usage such as oxygen consumption and hypothesizes that more effortful activities would result in less efficient energy usage. However, Pouplier shows that both for general movement and for movement of articulators in particular, activities that we are skilled to perform will be carried out in an energetically low-cost manner, irrespective of the type of movement. Pouplier concludes from her study that speech should be viewed as "an optimized motor system *because* it allows for context-specific articulatory reorganization; a characteristic that is the hallmark of skill rather than laziness" (Pouplier 2012: 160). This further underlines the positive notion of economy as a skill rather than lack of effort similarly to Lindblom's cost-based distinction.

In an attempt to deal with the other extreme of ease of articulation and thus to explain how contrast in language is retained in most cases, the principle of maximal dispersion was developed. According to this principle, contrast in language is realized by preserving the greatest possible difference in the perceptual space. It was explored more specifically for vowels by Lindblom and Liljencrants & Lindblom (Lindblom 1986; Liljencrants & Lindblom 2011). However, their approach follows the inherent assumption of language optimization and is thus a teleological approach.

Boersma and Haman investigate this behavior for the vowel spaces in various languages. They describe the resulting effect as two auditory dispersion effects. The first effect predicts that all sounds in a category will be spaced at equal distances from one another and secondly that larger inventories will increase the auditory space, while decreasing the acoustic difference between the sounds. They use this notion of dispersion to explain the acoustic difference between /s/ and /ʃ/ in English and prove that learning this dispersion can be acquired by computing phonetic constraints. If appropriate spectra for sibilants are learned from production, where every possible frequency profile represents a constraint, the listener and the speaker can be expected to undergo the same rankings (bidirectionality) to arrive at an appropriate computing of the targeted production/perception. (Boersma & Hamann 2008).

Boersma and Hamann's theories aim to unite the teleological understanding one derives from the unity of changes across similar languages without burdening the language user with a teleological approach to their individual production and perception of speech. They argue "sound change is teleological at the abstract level of the observed language, but non-teleological at the concrete level of the language user; this situation is analogous to that in evolutionary biology (Darwin 1859), where adaptions to the environment are observationally optimizing but underlyingly non-teleological." (Boersma & Hamann 2008).

Including theories of evolution in explaining the existence and initiation of sound change relies on the theoretical findings of Blevins (Blevins 2006a). Her concept of evolutionary phonology adds to these physiological patterns, but her reasoning is more phonological in nature. She defines "natural or system-internal sources of similarity" and "unnatural or external influence". Natural sources of sound change to her are "innate

aspects of speech perception & production" (Blevins 2006b: 122). Thus, changes come about not due to actively pursued articulatory effort or contrast in production, but because some patterns are marked for naturalness. It is not her attempt to account for intraspeaker variability, as seems to be the grounding for Lindblom, but rather to explain why certain sound changes appear more often cross-linguistically and diachronically than others. Her focus is fairly apparent when she argues:

> "Multiple factors contribute to instances of recurrent sound change, the most important being variable articulation, universal biases in perception, and language-specific perceptual biases. These multiple factors act as filters, determining differential rates of successful sound pattern 'replication' or transmission across successive generations" (Blevins 2006b: 122).

This understanding that the diachronic patterns of sound change may deliver insights into explanations of sound change initiation and probabilities attached to certain outcomes is further elaborated by Garrett and Johnson (Garrett 2015). They argue for two groups of bias factors, where one group is concerned with the "production and perception bias" and the other group consists of the "systemic bias". They also argue for a phonetic basis of sound change but not from a purely speaker-focused point of view but rather including speech perception and production.

On the other hand, Blevins has argued that hyperarticulation may lead to sound change as well. Lindblom's concept of maximal dispersion may aid in explaining vowel-related changes. However, Blevins describes the problem with teleological understandings of sound change at both sides of the spectrum with maximal contrast and minimal effort as follows:

> "The principle of maximal perceptual contrast suffers the same general weakness as that of articulatory ease: it is too strong, since complete mergers and contextual neutralizations are quite common and it is too weak, since it does not account for the fact that sound change results in maximizing perception only where it has a clear source."(Blevins 2006a: 293).

Thus, neither contrast nor ease can account for all the mechanisms observed historically in sound change. Some sound changes are well explained by contrastive needs, others lend themselves to economic explanations but no mechanism fits all changes.

Contrary to Blevins, Ohala argues that sound change is a probabilistic phenomenon, so that the onset of change cannot be predicted by theory, but rather a number of factors can greatly increase the likelihood of change. He decidedly argues against teleology of sound change, even though he defines sound changes as those cross-linguistic phenomena that "are most likely to arise from language universal factors" (Ohala 1993b: 238). Ohala argues that the origin of sound change is a change in phonetic forms that spread through a speech community, which results in the understanding that phonetic factors alone account for the initiation of sound change and social factors are only influential when it comes to the spread of change (Ohala 1993b).

A third idea explored by Bloomfield, Janda and Joseph and most recently by Fruehwald is that phonological factors may be at play in sound change (Bloomfield 1984; Janda & Joseph 2003b; Fruehwald 2016). In his structuralist account of phonetic change, Bloomfield theorizes: "Phonetic change acquires significance only if it results in a change

of the phonemic pattern" (Bloomfield 1933: 369). Janda and Joseph agree with Ohala that sound change is abrupt and phonetic at its origin. They start out defining sound changes as those that truly result from innovation and not historical correspondences. They define an innovation as "the dynamic alteration that occurs at the point of origin" (Janda & Joseph 2003a: 4). As such, they identify /s/-retraction as an example for innovation. They argue that neither the claims of distant assimilation to /r/ nor direct assimilation to affricated /t/ can reasonably account for the pattern, which would in turn be phonological explanations. It is thus a dynamic alternation- an innovation. They conclude in their "big bang theory":

> "in its purely phonetic manifestations, sound-change is ephemeral, though fully regular within narrow bounds: in fact, the narrower and more circumscribed the original context is, the better we can define and determine the likely associated regularity. Moreover, sound-change rapidly yields to generalization along non-phonetic (phonological or morphological) and social lines that may contribute further regularity via extension to broader contexts" (Janda & Joseph 2003b: 13).

Fruehwald, in contrast, shows that there is no phonetic involvement in the initiation of change but rather that, at least for the change in /ay/ he discusses in his paper, phonological factors initiate the change. The two phonetic precursors discussed in the literature are pre-voiceless shortening and offglide peripheralization. By showing that the sound change results purely from the phonological voicing of the following segment, he stresses that the onset of a change may also be phonological in nature. He is thus providing a view on sound change that is entirely removed from language production or auditory processing and placing it a level up in the cognitive or representational sphere of speaking and hearing (Fruehwald 2016).

A further pathway in explaining the initiation of sound change at the phonetics-phonology interface includes both hypercorrection and phonological factors as was proposed by Baker et al. While they consider the resulting change a phonetic change in the hypocorrection-sense of Ohala, they consider the involvement of phonological factors at the initiation or actuation stage (Baker, Archangeli & Mielke 2011).

They base this assumption on the fact that isolated phonetic factors cannot explain the actuation of sound change:

> "We do not believe that isolated phonetic factors have any bearing on sound change. Indeed, every word uttered by a speaker is influenced by coarticulation, and every word perceived is subject to coarticulatory effects; these phonetic factors overwhelmingly do not give rise to sound change. Rather, it is when phonetic factors are arranged in a way that facilitates a wide range of speaker variability that sound change is more likely to be initiated"(Baker, Archangeli & Mielke 2011: 366)

In the case of /s/-retraction, they thus argue that accumulation of error cannot explain the initiation of the sound change, because it would have occurred much earlier if this were the case. Instead, they argue that "only speaker-specific patterns that vary noticeably from one another are candidates for sound change" (Baker, Archangeli & Mielke 2011: 367). Great phonetic variability facilitates the initial salience and thus the potential to exaggerate and then use phonetic tendencies for indexical purposes. But

without this noticeable difference, change is unlikely. They thus predict that phonetic items with a high degree of coarticulation such as /s/-retraction are first sometimes perceived as exaggerated, then truly exaggerated and finally spread through social networks. At all of these stages, sound change is based on likelihoods and finally network strength.

3.2 Models of Sound Change

The study of sound change initiation provides a couple of directions for how sound change can be modeled after the initiation stage.

The articulation-based account of Ohala views every instance of language change as the mis-parsing of the acoustic signal leading to a re-interpretation of the prototype in the mental representation. Once this changes the production of the speaker, phonologization takes place. This is a result of the ever-present coarticulation in speech production. He further states that this is a two-way street and can account for both assimilation and dissimilation as both are possible and functional in the communicative situation.

He argues: "If the listener fails to correct the perturbations in the speech signal, then they will be taken at face value and will form part of his conception of its pronunciation" (Ohala 1993b: 246). This in turn results in a changed representation within the individual's mental lexicon and thus phonologization of a new phonotactic pattern.

He describes these changes in the mental lexicon as "mini-sound change". Thus, circumstances that would elsewhere be described merely as instances of variation within an individual are treated the same way as interspeaker patterns and phonologization at a larger scale. In Ohala's model, all change starts with the individual. "In hypo-correction the listener fails to parse or associate a given perturbation or variation in the speech signal with the conditioning environment. In hyper-correction the listener erroneously parses a given event with another event." (Ohala 1993b: 265).

Sound changes to Ohala are thus those instances of petrified reanalysis in a speaker's system that spread as a new norm across speakers. It is thus a very unlikely event, which he takes to be a good explanation why the variability in speech production and perception often does not result in change (Ohala 1993b).

In some ways, Blevin's Change-Chance-Choice model could be considered an extended version of Ohala's understanding of how sound change occurs. She labels different types of sound change as either one of these three types. The fundamental understanding is that "In each phonetic source, there is inherent ambiguity which can give rise to reanalysis on the part of the language learner." (Blevins 2006a: 35). For CHANGE, she hypothesizes that the listener has a greater-than-chance likelihood of mis-parsing a specific sound pattern. In CHANCE, the ambiguity is not created by the listener but rather a possible analysis of the specific sound pattern. The interpretation of the listener is random and could result in any of the ambiguous patterns. In CHOICE, the variability in the signal permits multiple analyses and the listener chooses which of the presented patterns is the underlying form. Blevins argues that this "model of sound change covers both subphonemic changes in pronunciation, changes in phonological representations, and shifts from phonetic processes to phonological ones." (Blevins 2006a: 262).

Harrington et al. look at the connection between production and perception in greater detail and find that variability is discernable from sound change when one looks at the discrepancies between perception and production (Harrington, Kleber & Reubold 2008). For /u/-fronting in Standard English, they find that a misalignment between production and perception can be found in younger speakers who have completed the change. They argue, in line with Ohala's assumption, that sound change is a failure to compensate correctly for coarticulation but also state that an exemplar understanding of the resulting sound change better predicts the ongoing change than a clear distinction between phonetic and phonemic change. They further find that not only the output may change over time but also the perception. While older speakers seem to classify the vowel based on the surrounding environment and are thus highly affected by coarticulation, younger speakers tend further toward a fronted categorization and are less affected by the phonetic environment. Harrington argues that the connection between perception and production is therefore correctly assessed by Ohala, but how the initiation of sound change leads to incrementation lacks an understanding of changing exemplar representation rather than phonologization. He refers to this as an "episodic model of perception and production in which the frequency and probability with which /u/ occurs in a certain context is the trigger for sound change" (Harrington, Kleber & Reubold 2008: 2833–34).

Models of sound change were then further elaborated by Yu and Baker et al. who further distinguished not only that the individual varies greatly within a sound change, but that the social settings importantly contribute to questions of actuation and incrementation of a change (Baker, Archangeli & Mielke 2011; Yu 2013; Fruehwald 2016). Baker et al. state: "sound change can be initiated when phonetically-determined inter-speaker variation happens to result in variants being distributed in a community in a way that favors the interpretation of extreme phonetically-motivated coarticulation as an exaggeration beyond the phonetic motivation" (Baker, Archangeli & Mielke 2011: 369).

Fruehwald further shows that not all sound changes can be modeled through any of the above-mentioned models of hypercorrection, because not all early stages of change involve coarticulation. He uses the example of /ay/-raising in Philadelphia to show that phonetic expectations predict an interaction with flapping that is not present in his analysis. The condition for change seems to be voicing of the following segment and thus phonological in nature (Fruehwald 2016).

3.3 The Propagation of Sound Change

On a larger scale, whether sound change involves mis-parsing, imitation or other factors, the debate has been whether or not sound change progresses in a regular manner. Regular sound change, also called Neogrammarian Sound Change, understands sound change as a regular process whereby the change affects all instances of a phoneme in a particular phonetic context in a gradual manner. On the other hand, exemplar-based models of sound change have found lexical effects indicating abrupt and categorical changes.

Neogrammarian sound change and its theoretical underpinnings are closely connected to the diachronic perspective on sound change. The Neogrammarians deduced patterns of sound change from earlier stages of change among related languages

(comparative method) and different stages of a language (internal reconstruction). The comparative method is crucial to their understanding of how various stages within and across languages can be detected and evaluated. I have introduced these concepts briefly in chapter one. Phonological reconstruction in particular compares corresponding sets of cognates to deduce proto-phonemes and change patterns (Rankin 2008). Only those cases in which change is evidently the direct result of phonological context factors are considered a sound change. The involvement of grammatical or semantic factors would in turn directly negate the interpretation of a change as sound change. To accept only those changes as sound changes that are truly phonetic and in their completed form, emanates from the comparative method. Changes behaving differently than these would often be described as sporadic changes, providing a term for all those instances in which the outcome could not be predicted from the input (Hale 2008).

Structuralists such as Bloomfield followed the Neogrammarian paradigm in their understanding of how a language changes its sound. In many ways, Bloomfield's description is even more reduced and very much based on a certain conception of mental representation and phonemes. He negates many of the factors that will later be shown to affect change (even if they may not initiate change). He describes exactly this notion of sound change when he concludes:

> "Sound-change is merely a change in the speakers' manner of producing phonemes and accordingly, affects a phoneme at every occurrence, regardless of the nature of any particular linguistic form in which the phoneme happens to occur […] Phonetic change is independent of non-phonetic factors, such as the meaning, frequency, homonymy, or what not, of any particular linguistic form. […] The whole assumption can be briefly put into the words: phonemes change." (Bloomfield 1933: 353–54)

The earliest, relatively uncontested pieces of evidence against this articulation and phonetics-informed approach to sound change was found in the development of tone languages. Chen and Wang analyzed the case of tonal changes from Middle Chinese to present articulation in the Chinese dialect Chaozhou (Chen & Wang 1975). A phonetically conditioned split, which should have resulted in a voicing conditioned split of the entire lexicon, shows a further split in tones in the conditioning environment of the voiced initial for Modern Chaozhou. No phonetic factors can account for this split, yet it appears highly affected by lexical items. In their analysis of homonyms, Chen and Wang find that 58% of former homonyms are now produced with different tones. Based on this analysis, they define the concept of lexical diffusion. A case of lexical diffusion is classified as a sound change in which "a phonological rule gradually extends its scope of operation to a larger and larger portion of the lexicon, until all relevant items have been transformed by the process" (Chen & Wang 1975: 256).

This definition already addresses the level of gradualness and transformation. However, it is more a statement about the type of progression than it is about the individual words affected by the sound change. In a later analysis, their understanding of gradualness is more direct:

> "A closer look at changes in progress and a more careful examination of large quantities of residual forms lead us to conclude that most (not necessarily all)

> types of phonological change are phonetically abrupt but lexically gradual. As contrasted with Bloomfield's conception, we hold that words change their pronunciations by discrete, perceptible increments (i.e. phonetically abrupt), but severally at a time (i.e. lexically gradual) rather than always in a homogeneous block. This latter conception of phonological change may be called LEXICAL DIFFUSION" (Wang & Cheng 1977: 150).

This later definition presents sound change as a change that is phonetically abrupt for each word it affects but gradual in its spread through the lexicon. They considered this change to be lexical in nature, because it speaks for the fact that the changes happen to the words stored in exemplar clouds, not to the segmentation of sounds.

Labov, in an attempt to provide room for both the neogrammarian hypothesis of sound change and exemplar approaches, investigates a number of phenomena to draw the conclusion that both types exist and pattern meaningfully in certain environments. While he expects reductive processes of change to be affected by lexical items, he argues that changes in place of articulation are likely Neogrammarian types of change that, due to the step by step incrementation of phonetic realizations, progress in a phonetically gradual manner (Labov 1981a). Crucial to his understanding of lexical diffusion is the idea that some words behave differently than others with no phonetic explanation. One such word would be *sad* in the short ae shift in Philadelphia which behaves differently than comparable adjectives such as *mad*. According to his interpretation, lexical diffusion negates the involvement of phonetic factors in sound change:

> "Lexical diffusion implies a rejection of the idea that phonetic conditioning fully accounts for sound change: that there are at least some words whose behavior is not predicted by their phonetic composition. If the word is a fundamental unit of change, it is because some words undergo change for reasons that are not phonetic" (Labov 1981b: 279).

Pierrehumbert understands the process of speech production as an interplay at the phonetics-phonology interface. She argues: "In a consensus view of phonetic implementation, lexemes (the phonological representations of words) are abstract structures made up of categorical, contrastive elements. The phonetic implementation system relates this abstract, long-term, categorical knowledge to the time course of phonetic parameters in particular acts of speech." (Pierrehumbert 2002: 101–2) While she argues in favor of the central function of exemplar mental representations in both language production and perception, she categorizes lexical effects as "second order":

> "Phonetic implementation rules are modelled through a correspondence between phonological labels and frequency distributions over the phonetic space. Individual words can bias the set of exemplars which serve as production goals. By assuming that words bias productions - rather than providing holistic production goals - the approach captures the fact that word-specific phonetic effects are second-order effects" (Pierrehumbert 2002: 134)

New to her approach to lexical diffusion is both the application of exemplar theory as well as a focus on the effect of frequency. These two are equally addressed by Bybee (Bybee 2002).

However, Bybee stresses two further characteristics of lexical diffusion. Her approach, in contrast to Wang & Chen's initial characterization, does not include categorical phonetic change in the lexemes, arguing that "phonetically conditioned changes that affect high-frequency words before low-frequency words are best accounted for in an exemplar model of phonological representation that allows for change to be both phonetically and lexically gradual" (Bybee 2002: 261). She also remains focused on the articulation of sound and describes sound change as fundamentally reductive, not only in those cases mentioned by Labov (Labov 1994). Like Pierrehumbert, she specifies frequency as a language external effect through which words get chosen first, providing an additional component to the progression of lexical diffusion that had not been stressed by Chen and Wang. This focus on frequency derives from an exemplar theory-informed view of the mental representation of the sounds at hand. She exemplifies this notion when she connects reduction and word-based phonetics:

> "Given a tendency for reduction during production, the phonetic representation of a word will gradually accrue more exemplars that are reduced, and these exemplars will become more likely to be chosen for production, where they may undergo further reduction, gradually moving the words of the language in a consistent direction. The more frequent words will have more chances to undergo online reduction and thus will change more rapidly." (Bybee 2002: 271)

In a more recent article she presents a more complex view of the involvement of phonetics in the change. By showing that word-internal sounds change at a different rate than those which are located at the word boundary, she redefines the understanding of phonetic context. She argues that a frequent context for a word in a sound change may have an effect, not just the phonetic context within the word. She thus presents data that allows for an interpretation that connects phonetic factors with frequency rather than just based on how often a word is used (Bybee 2017).

Tamminga stresses the importance of exemplar theory for establishing a connection between phonological representation and online processing mechanisms in connection to Bybee's and Pierrehumbert's claims. Where the mental lexicon and processing are not considered and accounts are based on production, a frequency effect could otherwise be explained by the mere process of articulation without having any lasting effect on word production and thus change. Exemplar theory is therefore most important to understand how usage shapes the processing and storage of exemplars, which articulation-based accounts of gradualness fall short of (Tamminga 2014).

In a review of studies on sound change and variability, Hay concludes that examplars are stored with more than just phonetic and word frequency detail. Beyond these factors, she defines the "full range of contexts"(Hay 2018: 701) involved in production and perception of speech to also include social and environmental information for each exemplar. Evidence for this type of detailed storage is for example that listeners have been shown to change their interpretation of the same acoustic cue as /s/ or /ʃ/ based on gender information on the speaker (Strand 1999). Based on exemplar models of language production, she argues that "speakers attend to statistical associations between linguistic variants and social and environmental context" (Hay 2018).

Labov remains critical of accepting lexical diffusion as a significant factor in sound change. He uses a close statistical analysis of the ANAE corpus data to evaluate whether any of the three vowel changes he looks at show a significant influence of lexical items. In regression models of all three changes, lexical factors cannot account for any of the variability in the data set (Labov 2010).

Tamminga uses a study of various homonyms of the different grammatical LIKE-productions to show that exemplar models cannot entirely account for sound change. These exemplar models of sound change would predict that homonyms, which naturally have different contexts of use and frequency patterns, are pronounced differently based on these varied profiles. Therefore, the more frequent homonyms are more likely to participate in a sound change. Her results show: "adjective, conjunction, discourse marker, and preposition forms of LIKE are in lockstep throughout the entire course of the change, despite order-of-magnitude differences in their within-dataset frequencies"(Tamminga 2014: 464). Furthermore, verbal LIKE behaves differently than all other types, indicating a grammatical involvement rather than a pure word-chunk mental representation (Tamminga 2014). Her most meaningful criticism of the paradigm is that exemplar theory is incapable of explaining stability in the sound system.

As outlined here, there are two central issues with the concepts of regular sound change and lexical diffusion. First, regular sound change rests on the theoretical assumptions made from a diachronic perspective and thus a completed change. Lexical diffusion was introduced as a concept when those patterns of progression theoretically expected were unable to account for the patterns of change and residual forms at hand. Both concepts, however, were not developed to account for individual differences in a change in progress. Nonetheless, researchers looking at change in progress have generally attempted to use notions of how change progresses that are based on these concepts. In a change in progress one would expect variability in production on the basis of individuals and generations, neither just based on words nor necessarily phonologically gradual.

Not all words with the phonetic or phonological context in question will categorically be produced with the changed sound. If there are indicators that the switch is more categorical and abrupt in nature, it would point towards lexical diffusion. However, the complex nature of the acoustics of many sounds involved in language production does not generally provide us with a clean cut-off point between a switch from phonetically gradual to categorical. Especially with respect to the potential sound change under investigation here, categorical patterns are unlikely. I therefore assume phonetic gradualness, much like Bybee.

While quantal theory or dispersion as described above could be used to establish categorical switches, changes in place of articulation are likely to show at least some degree of gradualness. Realistically, the most meaningful distinction to account for the type of progression of a change will therefore be its patterning in words. Following the analyses of Labov and Tamminga, an evaluation of the nature of a change would follow the default assumption that a change is regular in nature and only if a pattern emerges from the types of words being affected first, be it frequency based or random, can we consider an ongoing change a case of lexical diffusion. Furthermore, homonyms would have to appear with categorical differentiation in the midst of an ongoing change, especially if one extends

lexical diffusion to the frequency of a word in the lexicon. The question of sound change progression will therefore be reduced to two hypotheses to be answered by the data. On the one hand, I evaluate whether a categorical differentiation between /s/-based and /ʃ/ -based /str/-production can account for the variability in production. In a second step, I establish whether any word in the data set shows a meaningfully different retraction pattern than other words in the lexicon. If neither is the case, both characteristics of regular sound change would be met.

3.4 State of the Art in /str/-Retraction

The retraction of /s/ as described in the introduction is at the same time a locus of linguistic debate and an under-researched phenomenon. Early in the debate, scholars were mostly concerned with its potential for language change, its social characteristics, its regional profile and its phonological nature. However, current research is more interested in its progression through various speech communities.

The voices in this linguistic debate commence with Labov, who was one of the first sociolinguists to mention the variability of /s/ in consonant clusters as a "sound change in progress" (Labov 1981: 50). Within the description of the process, he first addresses the specific variable as the variation between a "hissing and a hushing sibilant" (Labov 1981a: 50). He defines the context in which the variation occurs, namely "before /tr/, though it also extends to /st/-clusters without a following /r/ and across word boundaries." (Labov 1981a). Later, he describes four variants, starting with a "hissing [s], used only by cultivated speakers", moving to "a normal sibilant with considerable hushing quality", which extends to "a fully hushing sibilant equivalent to the [š] in sheet" and ends with "an even more extreme form with distinct rounding" (Labov 2001: 206). There are both sociolinguistic and phonological factors involved in this understanding.

From an articulatory phonetic standpoint, he describes the sibilant space as starting at a fronter alveolar production, a "normal" alveolar with hushing quality, a fully post-alveolar production that is indistinguishable from /ʃ/ and lastly an even backer version that is amplified by lip rounding. Acoustically, he describes that the range of sibilants starts at noise frequencies reaching as low as 1200Hz for the rounded post-alveolar and as high as 4000 Hz for the "hissing" sibilant (Labov 2001: 206). Phonologically, this results in a gradient process. He argues that the retracted pronunciation is "triggered by the palatal quality of aspirated [r]" (Labov 2001: 206).

Without directly describing the social characteristics of this change in progress, Labov further specifies the hissing sibilant as a feature of "cultivated speakers", which leaves the other extreme to appear the opposite. He studies the variable amongst other variables in a Subjective Reaction Test. For this test, four speakers produce a reading passage including the target items *streets*, *straight*, *strips* and *strain*. Only one of the speakers is coded as using the retracted "hushing" sibilant and is significantly downgraded on the job suitability scale in comparison to her own rating. This speaker is a waitress from "an area of Kensington even less prosperous than Wicket St." (Labov 2001: 207), indicating lower socioeconomic status.

Shapiro mentions the sound change chronologically between Labov's first and second description. He addresses its lack of recognition: "The recent history of American

English includes a sound change that seems to have gone unattested in the scholarly literature" (Shapiro 1995: 101). Based on his own experience with speakers in broadcast media, he describes the social profile of the change as "neither dialectal nor regional" (Shapiro 1995: 101), which means that there is no apparent language external influence on the variable "gaining ground" (Shapiro 1995: 102).

Shapiro's description of this process involves fewer increments than the description by Labov. He states that regular users of this pronunciation replace the alveolar with the post-alveolar fricative, although he argues that it "can stop short of the full-fledged "phonetic power" of American [ʃ]". In further descriptions, he stresses the "peculiar acoustic character"(Shapiro 1995: 102) of the sound as visible in Olive et al. (Olive, Greenwood & Coleman 1993). He uses their spectrograms to interpret that in /str/, the friction rapidly changes its frequency patterns before the onset of /r/ voicing, as if the stop disappeared entirely. This disappearance of the stop informs his argument that the visible frequency profile permits an interpretation of the sibilant as retroflex alveolar sibilant. He supports this argument by comparing the spectrograms of retroflex alveolar and post-alveolar sibilants and finding almost no difference in the acoustic profile of the two.

The precursor for assimilation is a phonetic similarity between the sounds. Based on the description of the consonant cluster in Olive et al., Shapiro argues that the /t/ is not affricated and thus shows no qualities prone to assimilatory processes (Olive, Greenwood & Coleman 1993). This "relevant phonetic similarity" (Shapiro 1995: 103) does exist between the post-alveolar fricative and /r/ in being produced toward the palatal place of articulation, resulting in palatal productions of sibilants in words such as Sri Lanka (Lawrence 2000). This is the phonological argument he delivers for "assimilation at a distance", supported by the additional finding that /st/-clusters would also be affected if we were dealing with a case of adjacent assimilation (Shapiro 1995: 103). He finally theorizes that this phonetic change connects to markedness theory and thus reveals the marked features present in the particular environment of /str/:

> "More concretely, this pronunciation is an implicit "assertion" on their part [the retracting speakers] that in their grammar /r/, being [+ interrupted], is marked with respect to the opposition INTERRUPTED versus CONTINUOUS- exactly the same way that [ʃ], being [+ compact], is marked with respect to the opposition COMPACT versus DIFFUSE. The marked character of /r/ is prompted and confirmed phonetically, of course, by its high degree of retroflection in American English." (Shapiro 1995: 104).

He adds that /s/ remains stable in /spr/, because /p/ is not marked as ACUTE the same way that /t/ is. Through this theoretical explanation, he describes the change both as phonetic on the surface but phonological in its involvement of markedness.

Lawrence rebuts Shapiro's interpretation by arguing in favor of direct assimilation to affricated /t/. He identifies the phonological environment of the sound change as "not only word internally (*straight, strategy, stronger, instruments, Australia*) but also across word boundaries (*race track, last race*)" (Lawrence 2000: 82). Lawrence' argument against assimilation at a distance rests on various impressionistic findings for speakers in New Zealand. First, he clarifies that the analysis of affrication based on spectrograms is invalid, because the spectrograms in Olive et al. are based on an alveolar production (Olive,

Greenwood & Coleman 1993). He then draws a connection between innovative /ʃ/ in /stj/ clusters and /str/ clusters, illustrating that a possibility of retraction in the former environment negates the argument that aspiration may be the cause. He then argues that a /ʃ/ need not assimilate to a place of articulation of /t/ because /t/ is underspecified for this feature. This is followed by the analysis that affricated /t/ is considered one phoneme such that it must not be considered a linear progression from /t/ to /s/. He finally addresses that retractors cannot produce the pattern when asked to produce rolled /r/ in response to Shapiro's markedness argument. Since all versions of /r/ must be marked for [+INTERRUPTED], this makes the markedness argument questionable according to his logic. He concludes: "I have argued that the process that appears to be /str/→ /ʃtr/ is in fact an instance of /stʃ/→ /ʃtʃ/" (Lawrence 2000: 84). The fact that American English shows fewer instances of /stʃ/→ /ʃtʃ/ due to yod-dropping is not addressed.

Janda & Joseph report about this pronunciation in their big bang theory of sound change, stating that it "is getting to be strikingly rampant, and spreading to ever more contexts" (Janda & Joseph 2003b: 10), thus alluding to both a relatively quick diffusion of the change and an increase in permissible linguistic environments. They argue that neither of the described processes can fully account for the change, because it has spread to contexts lacking the /r/. However, this environment could be at the origin of the innovation of retracted sibilants in /str/-clusters.

Durian is the second sociolinguist to scrutinize /s/-retraction, this time in speakers in Ohio. His description is more centered on rounding than on the acoustics when compared to Labov's 2001 description, although it also addresses gradience of the phenomenon:

> "In Columbus, (str) has three gradient realizations. There is the standard variant [str]; an intermediate variant that is typified by an /s/ that shows retroflexion without pronounced rounding [ʂtr]; and the rounded variant [ʃtr], which is the more prototypical vernacular realization." (Durian 2007).

This description again attaches a notion of vernacularity to the phenomenon. He bases his findings on multiple approaches to data collection. In a rapid and anonymous survey of 120 white store clerks, he finds that in a logistic regression, social class, age, sex, and speech environment were significant factors in modeling the retraction of /s/ in word-initial (str)-clusters. He then analyzes 32 recordings from his dialect survey by taking into consideration tokens from two conversational styles as well as word-initial and word-medial position. Every speaker contributed 10 tokens. In contrast to Shapiro and Lawrence, he works with recordings and visual acoustic analysis. He rates the word position and morpheme structure of thirty-two study participants based on the three-point distinction quoted above. Where auditory judgment remains inconclusive, he carries out a visual judgment of the items in Praat, based on the following scheme: "high concentration of spectral energy at or below 2500Hz were coded [ʃtr]; those having a concentration of energy between 3000 and 3500 HZ were coded intermediate; and those with energy at or above 4000Hz were coded [str]." (Durian 2007: 70).

He finds that in the linguistic conditioning of the change, outside of the assimilation debate led by Shapiro and Lawrence, word position of the sibilant may also play a role, because he observes a slightly increased degree of retraction in medial consonant clusters. In greater detail, he stresses that these are for the most part instances in which the cluster

appears across morpheme boundaries. He argues "[ʃ] initially occurs either due to feature spreading of the [+high] feature from a preceding segment (as in the case of a high front vowel) or possibly lenition (triggered by a morpheme boundary) in word medial position" (Durian 2007: 71). The highest percentage of alveolar production of /s/ occurs word initially, while the highest percentage of non-alveolar /s/-production seems influenced by high front vowels. He therefore argues that the point of origin of /s/-retraction in triple consonant clusters with /t/ must be feature spreading from [+ high] vowels. However, these observations are not statistically significant and not based on acoustic measurements. He interprets the findings from both studies as evidence for an involvement of urbanization in the change, because he connects the local origin and the age of speakers with the cultural history of Columbus.

Rutter researches /str/-retraction within a smaller group of speakers, but is the first to measure all items acoustically. In contrast to Durian's rapid and anonymous survey approach, he carries out an experiment in which he controls for age, gender and ethnicity and expects a large proportion of categorical retractors. The participants in his study are ten eighteen to nineteen year old female speakers of Southwest Louisiana American English, who were recruited on campus without prior screening. When he describes the phonetic details of the possible change, he divides the gestures into *tongue placement*, *tongue shape* and *lip rounding*. The relevant phonetic similarity first mentioned by Shapiro can thus be defined in a more nuanced manner. Rutter notes that the post-alveolar sibilant and the rhotic share a backer *tongue placement* and *lip rounding* (which has never been reported for alveolar sibilants). He considers this evidence for a "harmonizing process" (Rutter 2011: 39).

His declared aim is to evaluate the gradualness of the sound change, and he argues that involvement of either one of these gestural features may create intermediate realizations:

> "If intermediate forms do exist, it may well be because the parameters associated with [s] are shifting toward [S] at differing rates. If tongue shape and position adopt the articulatory settings of [S] before lip shape, an intermediate, palatalized but non-labialized allophone would occur." (Rutter 2011: 31).

He decides to quantify the features acoustically using spectral peak measurements, which is a decision based on Jongman et al.'s findings as described above (Jongman, Wayland & Wong 2000). In contrast to Durian's focus on the /str/-cluster, he compares /str/ to /s/ and /ʃ/ per individual. This process considers the individual sibilant space and may account for differences in dispersion. As such, it enables Rutter to conclude that most of his subjects have a production of the clustered sibilant fully within their range of /ʃ/ and only six out of four tokens were located somewhere outside the production of /ʃ/. None of the outside tokens fall anywhere within the range of their alveolar production. While Rutter argues that these six intermediate realizations may be the result of a lack of lip rounding in the three gestural factors involved, he cautions against judging the sound change as fully gradual. He finds that low vowels (such as in the word STRUT) correlate with lower spectral peak measurements for the sibilant cluster and thus concludes that these coarticulation mechanisms may be the cause for auditory intermediate forms reported by Labov, without being representative of a gradual process. He further shows that the

standard deviation as a measurement of variability was greater for /str/ than for /ʃ/ both across speakers and per speaker. Rutter's study is thus the first study to confirm the existence of a sound change through a quantitative acoustic measurement technique, though on a rather limited sample.

In a later study, Rutter investigates the sound change in children and child directed speech. He interviewed eight mother-child pairs and found that of those children whose mothers were inconsistent adopters of the sound change, one showed features of lexical diffusion by retracting the word that the mother was retracting (Rutter 2014).

Mielke, Baker and Archangeli investigate /s/-retraction from two different perspectives in two journal articles. In the earlier one, they analyze the articulatory data of 20 speakers (Mielke, Baker & Archangeli 2010). They find that /str/-environments are generally produced with bunched /ɹ/. Within the acoustic centroid frequency measurements, they look at similarities between the sibilant in the cluster /st/ and the approximant in /str/. They directly compare the effect of being judged as retractor by the researcher and the effect of /r/ presence on the lowering of the centroid frequency and find both significant in an ANOVA model. To better account for the individual differences in the sibilant space, they develop a sameness-based score for retraction. Their so-called retraction ratio assigns the respective sibilant one of ten increments between zero and one, where one is equal to the participant's post-alveolar production and zero is equal to the participant's alveolar production. This retraction ratio shows no significant predictors in a linear regression model with following vowels, but a significant interaction between the predictors Judgment (retractor) and Articulatory Difference (between /s/ and /r/ tongue position). When they divide the data according to retractors and non-retractors, the centroid frequency of /s/ in the cluster is significantly affected by articulatory difference in non-retractors but not in retractors. They argue that this shows that "/s/ retraction has a phonetic motivation, since a phonetic tendency in the direction of /s/ retraction can be observed in non-retractors" (Mielke, Baker & Archangeli 2010: 720). This analysis is the first to consider a difference between the cognitive representation (assimilation) and articulatory effects in the motor activity (coarticulation). The fact that Articulatory Difference cannot successfully model retraction in retractors means that they may be assimilating due to partial phonologization of /ʃ/ in /str/ clusters, while non-retractors are only changing the frequency profile of /s/ in reaction to coarticulatory forces.

Baker et al. then use /s/-retraction as a variable in an investigation of the actuation problem. Much like Rutter, they also take the entire sibilant space per participant into consideration in their study of twenty-eight retractors and non-retractors. In contrast to Rutter, who describes a group of people who have fully completed the change, they only judge eight of their twenty-eight speakers to be retractors. Notably, they take a preliminary judgment to carry out this division by a binary rating of all items. They report to have few categorical retractors in their data set based on these ratings. They make a distinction between retractors and non-retractors based on a break in their data where no speakers are located between the 35.48% and the 62.50% rates of retraction. They argue that for a change that is phonetic in its origin, significant coarticulation patterns should be present in those speakers that have arguably not phonologized a new form yet. Indeed, their results show that a following consonant lowers the production of the sibilant and the presence of

/r/ lowers it further. The acoustic analysis they carry out is based on center of gravity measurements for the sibilants in the recording. They identify sibilant boundaries visually in Praat and measure the CoG of the spectrum created from the middle of each sibilant. Through this acoustic analysis they also find speaker-specific coarticulation patterns that correlate with diverging pronunciations of (bunched) /r/. "Specifically, more similarity between a speaker's /s/ and /r/ tongue shapes was found to be associated with more coarticulation between those segments when they co-occur in /str/ clusters" (Baker, Archangeli & Mielke 2011: 359).

In line with their expectations for the effects of coarticulation, the retraction ratio of retractors is not only greater but also not affected by these articulatory similarities. They find "three important facts about the phonetic motivation for s-retraction in American English" (Baker, Archangeli & Mielke 2011: 360): Non-retractors are coarticulatorily "biased toward s-retraction", there are parallels in the coarticulation patterns between non-retractors and retractors, and the bias for non-retractors is "variable across speakers" (Baker, Archangeli & Mielke 2011: 360). They conclude that the actuation of /s/-retraction is the complex interplay of coarticulatory forces in the phonetic environment and a lack of normalization in listeners due to great interspeaker variability (Baker, Archangeli & Mielke 2011). "We claim that sound changes are initiated when there is a spurious correlation between a phonetic variable and a social variable, and a speaker adopts the change into his/her own speech" (Baker, Archangeli & Mielke 2011: 365). They conclude that the unlikelihood of change relies on the interplay of three factors, namely: phonetic variability being attached to a social factor, individuals changing their speech patterns accordingly and then promoting these same changes throughout their social network (Baker, Archangeli & Mielke 2011).

Scudieri is one of the first to investigate perception aspects in the change. She argues that the lack of contrast for the sibilant in the consonant cluster results in perceived similarity as compared to sibilant contexts that do contrast such as following vowel environments. However, the differences in similarity judgments and speeded discrimination tasks included all non-contrastive environments fairly equally, with no increased pattern for /str/. Interestingly, she finds that individuals generally vary in their sensitivity for these detections. (Scudieri 2012).

Stevens and Harrington in turn investigate both the perception and the production of retraction. To assume a phonetically conditioned change, they argue that within a speech community that has not yet shown any indication of s-retraction as ongoing change (Australia), sibilants in /str/ clusters should show a lowered center of gravity or M1 (first spectral moment) than pre-vowel sibilants. Furthermore, if retraction were present in the production of these clustered sibilants, it should become meaning-distinguishing in spliced words (artificially inserting the sibilant produced in a cluster into a pre-vowel environment of a minimal pair). Methodologically, they employ dynamic measurements of the sibilant in contrast to the static sibilant measurements used by Rutter and Baker et al. Within these dynamic measurements, they can show that the M1 trajectory of the sibilants in /str/-clusters are not necessarily more retracted at the onset of the sound but rather that the M1 is stable for a longer period of time than other clustered sibilants. They further show that the log distance between the respective sibilant and /ʃ/ is greater for s/Consonant/r

sequences than for s/Vowel sequences.

In this first part of their study, both genders produce the various sibilant groups in similar patterns. In the perception part of their investigation, they crosssplice sibilants taken from /str/, /st/ and pre-vowel /s/ and /ʃ/ segments from the words *stream*, *steam*, *seam* and *sheep* into minimal pairs to create a forced choice environment. They further include listener response and talker gender as conditioning factors. Twenty-two participants rated the spliced items as either *sheet* or *seat*. The control items with the pre-vocalic alveolar and post-alveolar sibilants were overwhelmingly identified correctly, although fifteen of 440 items of /s/ were misidentified as /ʃ/. They interpret this finding to indicate that "/s/ can resemble /ʃ/, but not the other way around" (Stevens & Harrington 2016: 127). The other two environments are judged as *sheet* 7% (*steam*) and 26% (*stream*) of the time, with an effect of gender showing that male speakers are more often interpreted as post-alveolar in their production. This indicates that even in a speech community that cannot be considered as changing, there is potential for speakers to misinterpret the sibilant as post-alveolar, especially in male speech (Stevens & Harrington 2016: 128).

Gylfadottir examines a group of 225 white speakers in Philadelphia, who are a combination of two different corpora: The Philadelphia Neighborhood Corpus (136 speakers with birth years from 1888-1992) and the project "Influence of Higher Education on Local Phonology" (eighty-nine speakers), creating a birth year range from 1888-2004. She only includes speakers with more than four /str/-items. She standardizes her data by calculating a z-score based on the per-speaker overall sibilant center of gravity mean and standard deviation. Like Rutter, she finds significant effects for the lowering of the sibilant in /str/ especially for -tr but also for /k/, /p/ and just slightly for following /t/ (Gylfadottir 2015). Gylfadottir's analysis is the first work on /s/-retraction that takes a larger set of speakers in a naturalistic speech environment as its basis. She establishes that younger speakers are more likely to produce lower z-scored measures of center of gravity. She finds no effect for gender and does not analyze any other social factors. The retraction of /s/ in her corpus does show some correlation with the abandonment of the short-a system by younger speakers in Philadelphia (Gylfadottir 2015).

Wilbanks uses a sample of 140 white speakers from Raleigh in North Carolina to investigate further language internal factors in the change. He is the first to compare medial productions of the cluster with initial productions of the cluster in a corpus study. His findings suggest that word-medial productions of /str/ are driving the change, with a complex interaction of word position, gender and birth year. He further finds that the distance between /s/ and /ʃ/ is smallest in older speakers and increasing over time (Wilbanks 2017). In contrast to Gylfadottir, Wilbanks finds a strong female lead in his data (Wilbanks 2017).

A first comparative study of corpora from various regions was discussed by Stuart-Smith et al. at the New Ways of Analyzing Variation conference 47. They compared sibilant productions in corpora from the US, Canada and Great Britain. Their data indicates that /s/ in /str/ clusters shows differences in progression (with Raleigh and Columbus being the most retracted) and variability across speakers. The clustered sibilant is further removed from /s/ than from /ʃ/ in all the corpora, showing that none of the communities have completed the change just yet. Furthermore "Western US" is one of the few corpora for

which there seems to be little to no distance between /s/ and /str/.

Phillips, like Stevens and Harrington, looks at the temporal development of the sibilant in order to establish whether coarticulation or assimilation lies at the center of this change. He expects a coarticulatory pattern to result in a dynamic temporal development of the sound while assimilation would result in categorical productions. He finds, indeed, that coarticulation is present /s/-retraction with gradient temporal dynamics. He also considers prosodic position as a language internal factor in the change and finds that, in line with Wilbanks' observation on word-initial position, phrase-initial position is less affected by retraction than phrase-internal position. In line with Stevens and Harrington, he also finds great interspeaker variability (Phillips 2018).

In a rating task with nonce words, Jacobs and Resnick find that a stepwise incrementation of /ʃ/-like sounds in /str/, /skr/ and /spr/ environments is more likely to be rated categorically in the latter two environments. While speakers, due to the phonotactics of English, are more likely to rate the sibilants as alveolar, the threshold for choosing the post-alveolar sibilant in /str/ is lower than for the other two clusters, meaning that this choice is less categorical. They show that younger speakers are slightly more likely to make this categorical choice, indicating a change toward categorical judgment in apparent time. This could be interpreted as increased tolerance for coarticulation. The study further shows that higher toughness ratings for the voice in the stimuli correlated with more /ʃ/ -ratings (Phillips & Resnick 2019).

Nichols and Bailey discuss the same process debated by Lawrence based on an investigation of eight speakers and argue that their experimental data from both acoustic measurements and tongue tracking supports the hypothesis that /str/ and /stj/ clusters are both retracted in Manchester speakers. They further show that speakers are either categorical or gradient in a comparison of the spectral slices. Gradient to them are speakers whose averaged spectral slice aligns neither with the alveolar nor the post-alveolar sibilant. They conclude that the driving force is a similar /t/-affrication in both clusters. The tongue tracking is also interpreted as showing more than just a tongue retraction as the cause for the acoustic lowering of the sibilant (Nichols & Bailey 2018)

3.5 Research Aims of the Present Study

The status of research in /s/-retraction delivers only a small amount of definite answers on the sound change that this study investigates. From a phonological point of view, studies have indicated a phonetic initiation of the change in the specific combined environment of /str/. For the differentiation between coarticulation as a phonetic and hence a motoric process and assimilation as a neuro-motor and thus phonological change in pronunciation, the findings seem to indicate coarticulation at the onset of this change (Mielke, Baker & Archangeli 2010; Stevens & Harrington 2016; Phillips 2018). Even if coarticulation is agreed upon as causing retraction, and thus the initiation of sound change for the phenomenon, there are few indicators as to why the process presents as a gradual shift in pronunciation and how it could progress in so many communities in the English-speaking world. Furthermore, the datasets that assumptions are based on are quite different. At its beginning, the research on /s/-retraction was a purely theoretical endeavor (Labov 1981a;

Shapiro 1995; Lawrence 2000). Starting with Labov, the phonological nature of the change was investigated in laboratory experimental studies generating substantive amounts of sibilant items for relatively small numbers of speakers (Baker, Archangeli & Mielke 2011; Rutter 2011; Stevens & Harrington 2016; Nichols & Bailey 2018). Durian's study is considered the onset of corpus studies concerning the phenomenon of /s/-retraction, although he treats retraction as a categorical variable and his results are largely statistically insignificant (Durian 2007). Gylfadottir's study of /s/-retraction marks a change in data interpretation because she is the first to investigate retraction as a continuous variable, similar to Rutter, Baker et al. and Stevens & Harrington, but in a corpus that was not created for the investigation of sibilant clusters. Gylfadottir's and Wilbanks' investigations of /s/-retraction are thus far the largest studies concerning /s/-retraction in the US in terms of participants and therefore the first to successfully investigate change in progress as well as gender as a language external factor (Gylfadottir 2015; Wilbanks 2017). The largest global approach to /s/-retraction is currently undertaken as part of the SPADE project (Stuart-Smith, Sonderegger, Mcauliffe, et al. 2018). The SPADE project is the only one to use this corpus data for a comparison of different (arguably white) regions in the English-speaking world. While all studies have established patterns hinting at a change in apparent time, many aspects of this ongoing change are under-analyzed.

The data presented in this dissertation is created differently than those previous studies in multiple respects. First, it is the largest dataset collected for the singular purpose of evaluating /s/-retraction in /str/-clusters to date. As such, it includes both more items per speaker and more stylistic variation than previous studies. It is also the only specific dataset that includes speakers of various ethnic backgrounds. Furthermore, metapragmatic commentary allows for a unique qualitative analysis of discrepancies between the speakers and their production. It thus provides a new ground for an analysis of the assimilatory patterns that have been described in other studies, by combining many of the strengths of previous research: I treat retraction as a continuous acoustic variable as established in experimental approaches. Like the corpus studies with continuous dependent variables, I account for interspeaker differences in pitch through standardization and elicit on average ten tokens per speaker, which exceeds the cutoff point (five) set by Wilbanks and Gylfadottir. I am able to include further language external factors such as ethnicity and style and other language internal factors such as measures of word frequency, allowing an investigation of earlier theoretical accounts on the gradualness and progression of the change. The influence of the yod-cluster production is not considered since it seems an unlikely cause for the phenomenon in American English and thus, even if contact assimilation were the case, an unlikely explanation. The hypotheses that are (to some extent) addressed before and will be (re)evaluated here are:

Phonological Aspects:

1.) /s/ is produced with /ʃ/ -like qualities in the triple consonant cluster /str/ (Labov 1984; Shapiro 1995; Lawrence 2000; Labov 2001; Durian 2007; Baker, Archangeli & Mielke 2011; Rutter 2011; Gylfadottir 2015; Stevens & Harrington 2016; Wilbanks 2017).
2.) /s/ is most retracted in /str/ as a result of coarticulation (Baker, Archangeli & Mielke 2011; Stevens & Harrington 2016; Phillips 2018)
3.) The production of /str/ is changing toward a more retracted pronunciation in apparent time.

4.) The change will present as phonetically and lexically gradual, with frequent words significantly affecting Centre of Gravity measurements (Rutter 2011; Wilbanks 2017).
5.) Some individuals have become retractors with full phonologization of the post-alveolar sibilant in the /str/-environment (Rutter 2011).
6.) Sibilants with longer duration are more likely to be retracted.

Sociolinguistic Aspects and Variation Patterns

7.) Females are not leading the change (Gylfadottir 2015).

Activity based style has thus far not been controlled for in any of the experiments. Although both White and African American Speakers (Rutter 2011) have been investigated in this change, there is no acoustic study that compares these speakers within the same sample as of yet. The strong presence of /str/-retraction in Rutter's data as well as a rapid and anonymous survey carried out by Hinrichs et al. informed my assumption that ethnicity will have an effect on the change (Hinrichs et al. 2015). Although Phillips and Resnik have tested the perception of masculinity through /str/-retraction, no research has such far gauged other aspects of social evaluation. I therefore include three hypotheses in my investigation that none of the previous studies have sufficiently addressed before:

The hypotheses that none of the previous studies have sufficiently addressed before were:

8.) Activity-based speech style will show differing levels of retraction.
9.) Ethnic identity will have an effect on retraction such that White speakers are less likely to retract.
10.) For retractors and non-retractors alike, the change in progress shows little social evaluation.

Due to the theoretically problematic assumptions about social class and education as factors in language change, I include questions about this background in my survey without hypothesizing about an effect. These social factors will be explored but not included in confirmatory statistics.

The following chapter describes the methodological approach before providing the results and their interpretations to deliver data on each of these points.

4. Data and Methods

This chapter illustrates and discusses in detail the steps taken to evaluate the research questions outlined above. After a preliminary introduction to the methodological concerns, the resulting study design is outlined. The sample and procedures are then described in detail.

4.1 Methodological Approach

The methodology chosen to analyze sibilants in the present study is dualistic in multiple respects. In the first sense, as outlined in the previous chapters, the combination of sociolinguistic and phonetic features in this study marries the methodological aims of both disciplines. While the focus of phonetic research may often be the most accurate detection of a certain phonetic feature and its exact evaluation, phonetic features are only valuable to sociolinguistic researchers at the level on which they become a meaningful feature in natural speech variation and/or social behavior. The resulting dichotomy in sociophonetics calls for data which is, on the one hand, of sufficient sound quality to differentiate variability on the phonetic level. On the other hand, it must also be representative of the natural linguistic behavior of the subject and not merely a constructed product of the restrictive laboratory environment (Di Paolo & Yaeger-Dror 2011: 8).

In the second sense, the methods employed in this study must not only estimate the sound change presumed in /str/-clusters on a quantitative level, but also evaluate to what extent this change has already risen to the level of consciousness and become indexical. To achieve this dual understanding, the quantitative approach taken by researchers such as Gylfadottir is complemented with a qualitative interview. As a result, the present study must be regarded as a mixed methods multi-discipline approach to /str/-retraction.

4.1.1. Sociolinguistic Methodology

The study follows sociolinguistic standards by conducting interview activities in the natural environment of each subject and eliciting relatively natural speech through various activities described in the methods section. The underlying theory of variationist sociolinguistics is described above in greater detail. Methodologically, sociolinguistic inquiry is aimed at representing the 'vernacular'. This is the case for two reasons. For one, anecdotally, most variability in language seems to exist when people use language in everyday life situations. Secondly, it has been defined as the "style in which the minimum attention is given to the monitoring of speech" (Labov 1972: 208).

In the present study, the vernacular is treated not as one end on a one-dimensional language spectrum, but rather in line with Trudgill's dialect triangle (Trudgill 1996). The vernacular is understood here to include all socio- and dialects in the speaker's linguistic repertoire, which they use in familiar environments. This means, in more concrete terms, that even at the level of minimal language monitoring I cannot assume that the samples I obtain are entirely congruent with the speech patterns participants exhibit among friends and family. It is thus the least monitored interview speech with a community outsider that is realistically being evaluated in this study.

Monitoring is exactly the concept described by the "observer's paradox" (Labov 1972: 209); the problem that language data can only be gathered through observation and yet an observed speaker is likely to perform differently than a speaker in a common life situation. It is therefore important to conduct the interviews in a familiar environment with minimal devices. The Rutledge effect further influences the nature of natural speech elicited through sociolinguistic methods. It refers to the influence the perception of varying interviewers have on the linguistic performance of an interviewee (Bailey & Tillery 1999). I chose to control for this effect by conducting all interviews alone. While I cannot deny being perceived differently by speakers of varying age groups and social backgrounds, it brings the benefit of not imposing any expected outcomes on the data as would be the case when different interviewers were to be chosen based on social categories.

Stylistic variation in an individual speaker is understood to be much more complex than Labov's initial single dimension of attention paid to speech (Labov 1972, 208), but the concept of style remains a concern in experimental design. Though the complexity of style and stance (Kiesling 2009) exceeds the possibilities of the scale of data in the present study, an awareness of stylistic differences is important to account for possible variation.

The present study further aims to evaluate the influence of larger social categories such as gender, socioeconomic status, age and ethnicity. Age is a necessary measurement to develop any theories of language change.

As discussed in Chapter 1 in connection to the waves of variationist sociolinguistics, the remaining categories are questionable in the sense that they are large enough to overestimate effects and may mask smaller scale patterns such as communities of practice. However, as discussed in Chapter 1, these categories are categories because they are meaningful to western societies in their identity formation processes. Social positions are shaped by gender and ethnicity to this day, regardless of whether this is a pleasant political reality or not. For a process such as /s/-retraction, I therefore find it necessary to look at the larger patterns before establishing to what extent these subgroups may be divided into more meaningful, smaller subgroups. In other words, the first step is to find those communal groups that are valuable to a closer analysis. The triangulation approach I take should therefore not be misunderstood as an outdated analysis of overrated sociolinguistic patterns, but rather a first and necessary step in the social profiling of an under-researched linguistic phenomenon that will deliver both quantitative and qualitative pointers for future research foci.

It must be assumed that certain language stereotypes may alter the performance of participants on top of stylistic and situational differences. In order to prevent such bias, researchers may choose to use any range of vagueness and deception in line with the expectations of informed consent (Eckert 2013). In the present study, the linguistic focus was therefore entirely masked and was instead guised as tasks used for memory research.

The debriefing section was used to establish first insights into the social perception of the phenomenon. I chose to have an open conversation about the phenomenon rather than a matched-guise task to evaluate perception, because it puts the participant in control of the situation. Having just learned that they volunteered their speech for analysis, I wanted to grant an opportunity to have a say in the matter. While providing some good starting

points to gauge the overall opinions, if any, and have an insight into folklinguistic explanations, the debriefing left participants with a sense of value and empowerment. Another experimental task would have done the opposite.

Beyond representing various social categories in my research, I also designed my interviews with the people in mind. It was especially important to me that they considered the activities as playful and fun in order to create an environment where I was not so much the interviewer asking them to perform tasks, but rather in a position to encourage them to use their memory well or comment on a story. Such designs are beneficial for preventing the observer's paradox effect.

4.1.2. Phonetic Methodology

The aim of the phonetic part of this thesis is to better characterize the nature of the sound change in both its gradualness within the production of the sound as well as environmental intra-linguistic factors in its appearance. The present study therefore relies on acoustic measurements of the characteristics of the speech sounds being analyzed. This is not only in line with newer research on the phenomenon, but also methodologically important as it allows for two things that binary judgements such as the ones discussed by Shapiro cannot account for:

1.) An analysis of the gradualness of productions
2.) A non-rater-biased quantified measurement

The recording procedure, transcriptions in Praat and forced-alignment with FAVE, as described in greater detail below, yield data that can not only be analyzed in this quantitative manner, but also deliver more appropriate results on the factors involved in the change as they are less prone statistically to overestimate effects.

To evaluate whether differences in the linguistic environment, speech rate and cognitive processes were involved in producing items (be it reading or free speech), the present study also included various tasks in item production. While this is, on the one hand, creating different styles as interpreted above in the Labovian sense, it also creates output variability from a phonological point of view.

4.1.3. Methodological factors in quantitative and qualitative aspects

Studies on /s/-retraction thus far may be considered sociophonetic, but can clearly be divided into qualitative and quantitative approaches as outlined above. Crucial aspects of how change functions and progresses however lie at the crossroads between quantitative and qualitative approaches. How quantitative acoustic results relate to theory and how much of the initiation and propagation relies in fact on indexicality and as such, can meaningfully be evaluated by qualitative means. A dually mixed methods approach is therefore necessary. Not only are multiple tasks needed to provide a stylistically complex picture of the phenomenon at hand, but the complex progression of retraction must also be questioned from a qualitative point of view. This can be considered triangulation in the sense that we have the surface representation of retraction on the side of the countable instances used in speech and the underlying mental representation of retraction on the opposite side. To deduce the true complexity of the appearance of retraction, research methods must aim to evaluate the social occurrence and phonetic nature of the phenomenon

below the level of consciousness as well as views and accessible knowledge about its social meaning on the other hand.

4.2. Study design

The state of the art section above has described several different data sets that have been used to make claims about /s/-retraction. They can be divided generally into theoretical papers (Shapiro 1995; Lawrence 2000), corpus studies (Gylfadottir 2015; Wilbanks 2017; Stuart-Smith, Sonderegger, Mcauliffe, et al. 2018) and experimental studies (Durian 2007; Baker, Archangeli & Mielke 2011; Rutter 2011; Stevens & Harrington 2016; Phillips 2018; Phillips & Resnick 2019). All corpora that have been used for retraction research thus far are based heavily on sociolinguistic interviews. While corpus studies provide insights into a large group of speakers, they include relatively few per speaker examples in such an infrequent phonotactic environment as /str/. They are also often based on communities that are easily accessible for undergraduate students, resulting in very little ethnic and socioeconomic diversity within the data sets. The present research, therefore, is based on an experimental design with much larger numbers of participants than previous experimental designs (the largest number of recorded participants thus far are the thirty-two undergraduates recorded by Baker et al.). It thus combines the statistical power of corpus-based investigations with the control of experimental designs. It is also the first study carried out in a guised manner with a varied number of elicitation tasks, where others employed "two conversational contexts" (Durian 2007) or simple target word elicitation through a carrier phrase (Rutter 2011; Stevens & Harrington 2016).

The details of the sample, the material and the procedure of this research project are described below.

4.2.1 Sample

The aim of the present study was to represent larger social categories present in the city of Austin. I therefore set out to find 100 participants roughly balanced for gender, age, education and ethnicity. The overarching goal was to create a sample of Austinites without the classic college bias present in many of the datasets that /s/-retraction was analyzed in (e.g. Rutter 2011). Overall, the sample was designed to deliver enough data to evaluate /s/-retraction in native English speakers in Austin in apparent time.

As an outsider to the community, my sample must be considered a convenience sample in multiple ways.

1) I had no institutional or financial backing that could grant me an additional incentive for participation. I had to rely on the graciousness and free time of participants.
2) Though I aimed to visit multiple diverse neighborhoods in Austin, I was limited both by concern for my safety as well as finding a communal space in which my approaches would not be considered solicitations.
3) The snowball principle was in some cases a helpful way to reach a bigger number of participants. Yet, I was not able to apply this principle to all data due to my lack of connections.

I ended up interviewing eighty-six participants in roughly six different neighborhoods in Austin. Many of the participants grew up in Austin or other parts of Texas. However, fourteen participants only moved to Austin from out of State after 2010.

Since research thus far strongly indicates a non-regional change, I included these participants as well. Nonetheless, the social dynamics of Austin as a city are expected to have an impact on the social structure of the change in my dataset. I therefore elaborate on interview locations and the history of the city below. In the following, I first describe the demographics of the City of Austin, then briefly introduce my interview locations and illustrate the specifics of those neighborhoods in greater detail.

Austin is the capital of the State of Texas. According to the most current population estimates, the city spans over 326.33 square miles and has a population of 981,035 people (City of Austin, 2019). Austin is one of the proportionally fastest growing cities in Texas and has grown by 150.000 inhabitants in just seven years. It is a so-called "no-majority" city, in which the proportion of white inhabitants is currently below 50%. The neighborhoods are rapidly changing as property prices have gone up by as much as 80% between 2010 and 2015. Two areas well represented in the dataset and crucial in the current development of the city are Central East Austin and Central South Austin. While East Austin is still inhabited by large goups of African Americans and Hispanic-Latinxs, current population trends show that the percentage of white inhabitants in all central areas of Austin is sharply increasing, which restructures the diversity of central neighborhoods drastically. The city of Austin reports:

> "Many community leaders talk today of how many of these families are still returning to churches in east Austin on Sunday morning. However, many of these same community leaders fear that the newly-suburban African American population will eventual build suburban churches closer to home, leaving the original houses of worship somewhat stranded. The potential impact of the loss of these churches and their community outreach and community care programs on the African American households left in east Austin could be devastating." (Robinson 2016).

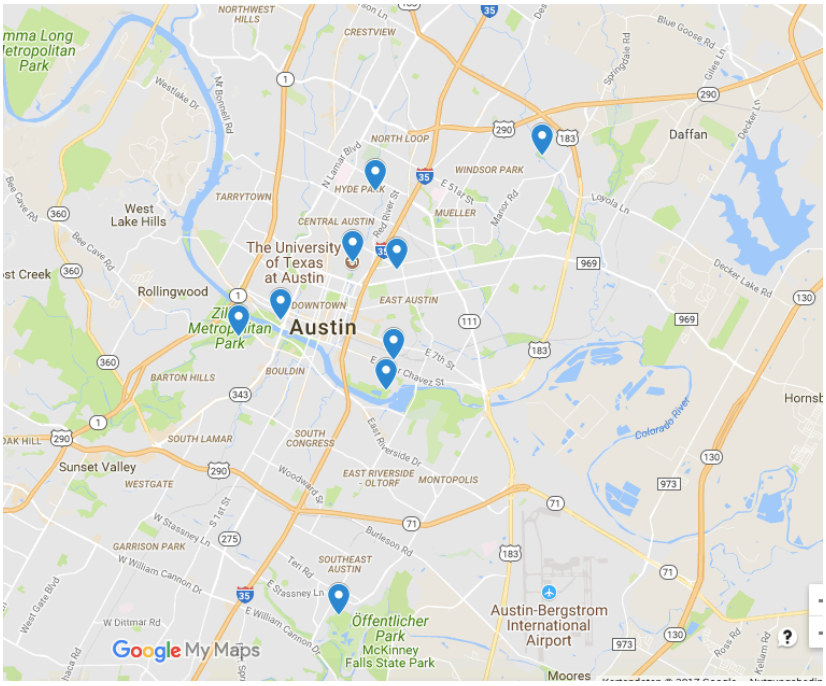
The Hispanic-Latinx population is increasing especially fast in areas of Lower East Austin and Central South Austin, both of which are represented in the dataset. The clustering of Hispanic-Latinxs, however, is currently strongest for lower income households.

I allude to these general trends in greater detail in the sections below.

4.2.1.1 Interview Locations

The map in Figure 4.1 provides an overview of all of the interview locations. All of them are parts of the metropolitan area of Austin. These locations are spaces where two or more interviews took place. Most of the locations are public places such as recreation centers, which have an influx from the larger surrounding neighborhood.

Figure 4.1 Interview Locations



4.2.1.2 Population

The demographic maps below illustrate the general characteristics of Austin and its neighborhoods.[6] To aid in the interpretation of these maps, it can generally be stated that I35 (visible in Figure 4.1) is the main dividing line between lower and higher income families. East of I35 (although in recent city planning, gentrification is quickly changing the nature of East Austin) and south of the Colorado River/Lady Bird Lake are, generally speaking, the areas in which minority groups and lower income families can be found.

4.2.1.2.1 Ethnicity

Though Austin still has a large group of white inhabitants, the Hispanic-Latinx population makes up more than two thirds of the ethnic set-up of the city. The Asian and African American population is almost equal ranging around 7% of the total population. The Hispanic-Latinx population has increased in the past seventeen years while the percentage of African American inhabitants is slowly declining. This trend is visible in Figure 4.2 and 4.3 below, which show that in 2014, the youngest generation has a greatly increased percentage of Hispanic-Latinx inhabitants and decreased percentages of white and African American inhabitants.

[6] All demographic maps were used with written permission by Ryan Robinson, city demographer for the city of Austin. In line with their legal notices, the material was not changed, although some images were cropped to the relevant areas.

Figure 4.2: Ethnic Age Structure

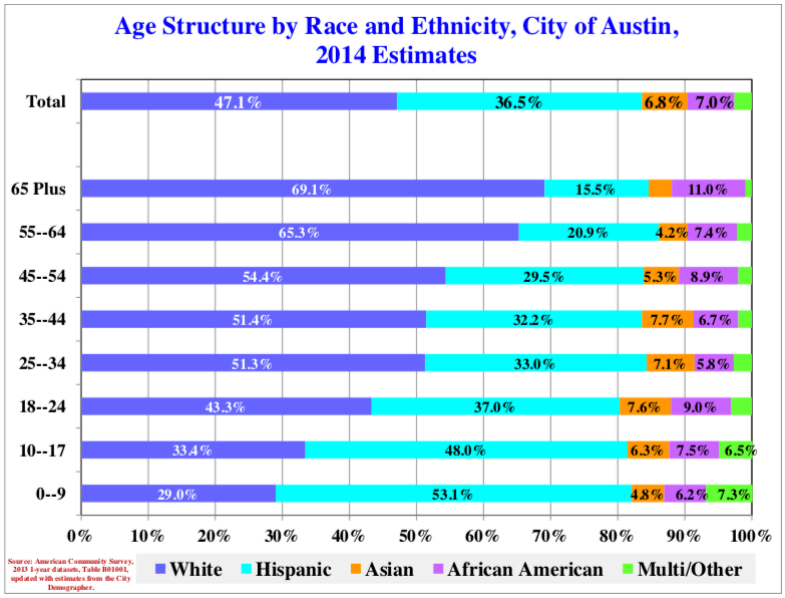


Figure 4.3: Ethnic Share of total Population

Race and Ethnicity, 2014

City of Austin, Travis County, the Austin--Round Rock MSA and Texas (2)

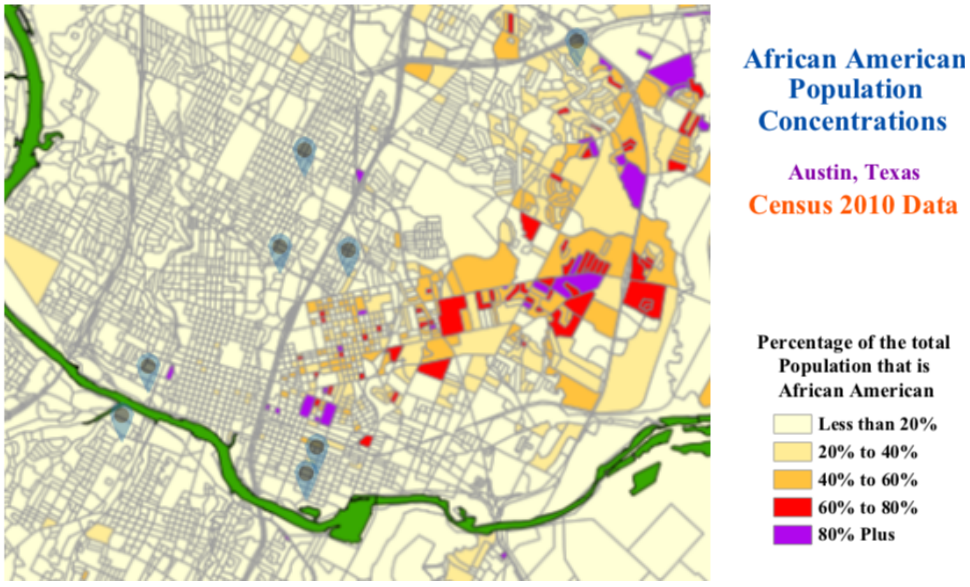
Data Source: City of Austin estimates plus information from the US Census Bureau's American Community Survey.

	---City of Austin---		---Travis County---		---MSA---	
Racial or Ethnic Category	Totals	Share of Total Population	Totals	Share of Total Population	Totals	Share of Total Population
White--Anglo (non-Hispanic)	407,665	47.1%	547,106	48.0%	1,011,674	52.5%
African American (non-Hispanic)	60,585	7.0%	88,833	7.8%	132,963	6.9%
Hispanic--Latino	315,909	36.5%	387,221	34.0%	632,055	32.8%
Asian (1) (non-Hispanic)	58,854	6.8%	65,540	5.8%	107,912	5.6%
American Indian (non-Hispanic)	2,529	0.3%	28,472	2.5%	5,588	0.3%
Other (non-Hispanic)	813	0.1%	2,050	0.2%	3,221	0.2%
Multi-Racial (non-Hispanic)	19,017	2.2%	19,662	1.7%	32,813	1.7%
Totals	**865,504**	**100.0%**	**1,138,884**	**100.0%**	**1,926,998**	**100.0%**

Notes:
(1) Asian includes native Hawaiian and Pacific Islander
(1) Total population estimates are as of April 1, 2014 with the exception of the state of Texas for which the as of date is July 1, 2014.

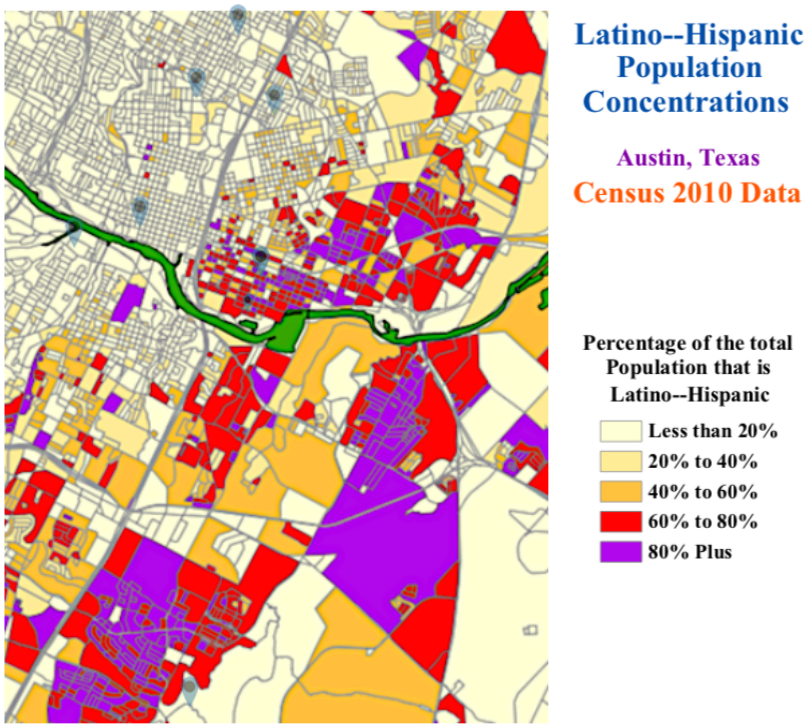
Figure 4.4 from 2010 underlines that two of the community centers where interviews took place are situated in neighborhoods with higher percentages of African American inhabitants. These representative neighborhoods are called East Austin and University Hills. Compared to Figure 4.5, it is apparent that African Americans are minorities in almost all Austin communities with only two remaining purple (>80%) areas in the metropolitan center of East Austin.

Figure 4.4: African American Population Concentrations



Most communities with any Latinx-Hispanic inhabitants are located east of I35. Within this part of the city, most neighborhoods in Austin show high concentrations of Latinx-Hispanic inhabitants in Figure 4.5. In this sample, the interview location in Southeast Austin and the location in South Austin are situated in especially Hispanic-Latinx Neighborhoods.

Figure 4.5: Latinx-Hispanic Population Concentrations



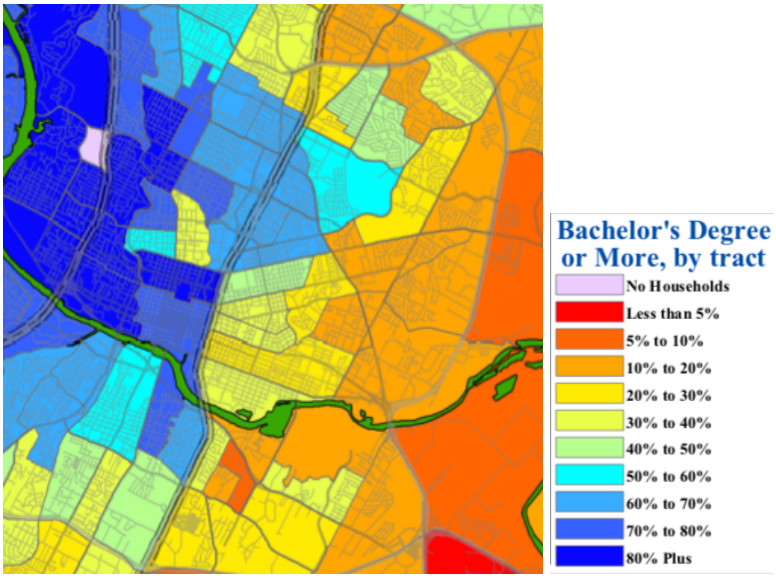
The large number of Hispanic-Latinx inhabitants visualized in Figure 4.5 is not representative of the proportion of native English speakers with Latinx-Hispanic ethnic affiliation. This is in contrast to the African American population, where population estimates are interchangeable with possible participants for the study. It is important for the

research presented here that transfer effects between languages with differing phonotactics are being kept to a minimum. However, according to the American Community Survey 2017, a large proportion of Austin's Hispanic-Latinx population speaks mainly Spanish at home. The American Community Survey reveals that in the City of Austin, 38,1% of the Hispanic-Latinx population who speaks Spanish at home rate their own English as less then very well (City of Austin, 2019). This was an issue when finding participants in Hispanic-Latinx neighborhood. Many speakers in Hispanic-Latinx Rec Centers were unable to participate, because they could not be considered native speakers of English.

4.2.1.2.2 Education

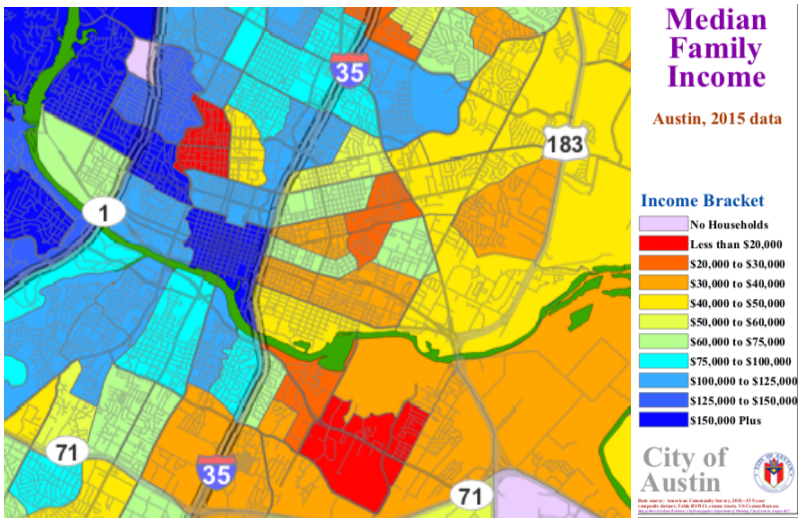
The rise in the tech industry in Austin is in line with the general rise in adults with Bachelor's (BA) degrees or higher (Kerr 2018). Compared to the national total of 7% of jobs in the high-tech industry, Austin currently has 14% of employees working in said industry. From 1990 to 2010 the percentage of inhabitants with a BA has increased drastically from 34,4% to 49% in 2017 (Census 2017). The east-west division present in the ethnicity-based population maps is in line with the dispersion of Bachelor's degrees reported in the census data.

Figure 4.6: Bachelor's degree or more



Similar patterns appear in the median family income as presented in Figure 4.7. I35 divides areas where the income ranges between $20,000 and $50,000 from areas with $6,000 income and more. Note that the low-income area west of I-35 is located around the UT campus and should thus not be considered a regular residential area.

Figure 4.7: Median Family Income



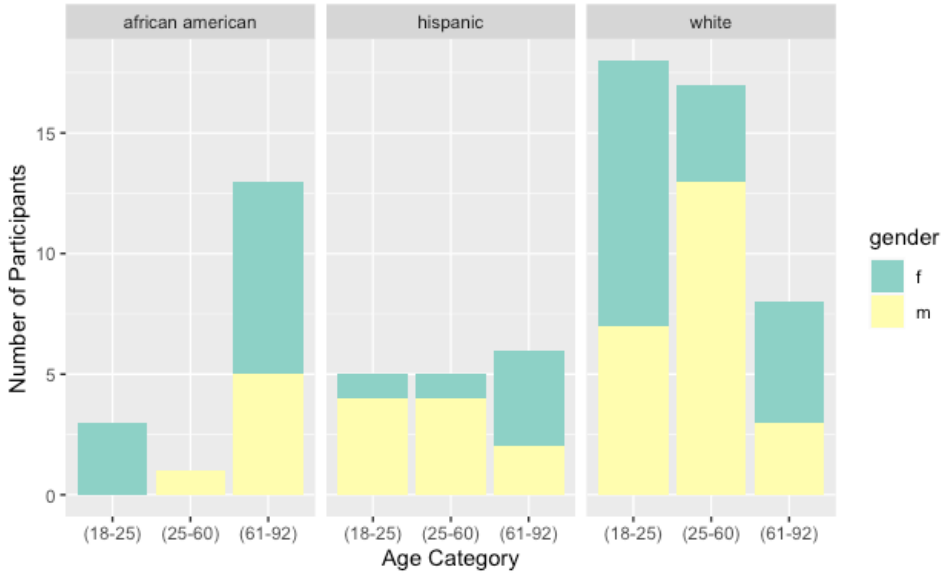
Other trends present in Austin's current development are the rapid growth and resulting urban sprawl, the decreasing proportion of families with children in the urban core, the rising proportion of Asian inhabitants and the increasing social gap between affluent and working class inhabitants. This gap is, however, visible in all the maps as the strong East West division. Although the total numbers of African Americans in the city are still increasing, their proportional share of the total population is changing rapidly as mentioned above. This creates a unique environment for linguistic inquiry, because the city remains segregated between the rich and the poor, but former communities of African Americans in the East are dissolving. It appears to me that the dividing line is no longer one of just black, brown and white but rather between white and non-white, at least for the younger visitors in the community centers I frequented. There are two centers with almost exclusively African American seniors in their senior programs but very mixed participants in the parent and children generations. The next section illustrates how these general demographics are represented in the sample created for this study.

4.2.2 Participants

The distribution of age, gender, ethnicity and education are described in Figure 4.8. This graph is based on the subsample of speakers identifying as belonging to one of the three main ethnicities that will also be used for regression modeling. While there were eighty-two participants overall, I will explain in the results section which subsamples of these speakers were used for which analyses.

It is visible in Figure 4.8 that gender is fairly evenly dispersed, while the other social categories vary in numbers. The numbers of young and old participants are comparable, which is especially important for the analysis of an apparent time change. The unfortunate lack of middle-aged participants, partially due to their inaccessibility as full-time workers, may mask some details of the progression of the change without impacting the overall trends.

Figure 4.8: Histogram of main participants according to age, gender and ethnicity



The problems with representativeness in the sample of ethnic groups present in Austin has been outlined above. The Census does not differentiate between generational differences and English language proficiency within the group of Hispanic-Latinxs. While the number of speakers of African American ethnicity in the sample is greater than the portion present in Austin's population, it allows for a more powerful evaluation of the trends present within that group.

Education is not displayed above because it is the most unbalanced factor in the sample and yet representative of the differences in education noted in census data. The problem here lies both in the nature of quantifying education as well as in the distribution of education in Austin. As mentioned above, I had little access to the true blue collar workers (or unemployed people) in Austin. For those that I could reach, a college degree is relatively normal in certain generations. I refrained from asking more about the degree.

The great variability amongst universities, colleges and community colleges that exists throughout the USA renders an undergraduate degree an almost meaningless distinction. In chapter one, I mentioned the difficulty of determining valid judgments of social class. I decided against threefold measurements or other factors, because I did not want my participants to feel judged in any way and thus decided to only record the biographic factors job and education. I am aware that a combination of factors such as education, job and household income could create relatively reliable measurements of social class, but I will leave this possible social factor for more restricted types of field work to analyze.

4.2.3 Instruments

The mixed methods approach described above was realized through various steps of consent and social information as well as activities and interviews.

4.2.3.1 Consent Procedure

Before the interview, participants were informed about the entire process of activities with a consent form. They had time to read all the information provided on the form and ask any additional questions. I also verbally went over the important points with them before they

signed. Since the linguistic aspect of the study was concealed, the consent form was focused on the types of activities they would perform. In order to not lose their trust after the debriefing, I made sure to keep goal-related answers to a minimum and thus lie as little as possible when it came to my own training as a researcher and my research hypotheses in conducting the interview. The participants then filled out a paper copy of a survey with social information. In an open format, it recorded the birth year, gender, (former) job, highest level of education, interview location and self-identified ethnicity.

4.2.3.2 The Interview

I will refer to the entire range of activities performed by my participants as the interview, because they were naturally embedded in the interview situation, even though some sections were activity-based, while others were comprised of free speech. The entire interview was recorded on a zoom H2 recorder.

4.2.3.2.1 Word Naming

The first task performed by all participants was a word naming activity. For this activity, I developed two slideshows with twenty pictures[7] each. The first slideshow was a random array of various pictures. The second slide show grouped the items into 6 groups with 4 items each to improve recall ability. The groups, as indicated on the slides, were: Berries, Beverages, Transportation, Pets, Dining and Jobs. Slideshow A included the following targeted words (the capital letters being target words, the rest fillers):

Raspberry	STRAWS
Mouse	Lion
Forest	STREAM
Baby	Steak
Peach	Light bulb
Hammer	Dolphin
Pen	Ice cream
RESTAURANT	Bird
Cake	Dice
Apple	STRING
Moose	Horse
Fork	

Slideshow B was grouped in the following way, with target items in capital letters:

Blueberries Rasberry STRAWBERRY Blackberry
Tea, Coffee, Orange Juice, Soda
Path, STREET, Trail, stairs
Cat, Dog, Rabbit, HAMSTER
Café, RESTAURANT, Bar, Diner
Nurse, CONSTRUCTION WORKER, Teacher, Police Officer

Pictures instead of word lists were chosen in order to elicit free speech rather than read speech. The downside of pictures is the fact that not all pictures will deliver the desired

[7]All pictures were taken from the website pixabay.com to avoid copyright issues.

target productions, with some participants calling a strawberry a raspberry and the like. I still chose pictures over reading passages because I wanted to have a more varied range of language production activities than just reading and interviews. These slideshows were presented to the participants on a laptop. The switch from one picture to the next was based on their own naming speed.

4.2.3.2.2. Reading

The second task performed by the participants was a reading section. I gave them a children's book in an unaltered format and asked them to read the story "Lotta on Troublemaker Street" aloud. I chose the short story because it included multiple renditions of the word STREET as well as other rare sibilant patterns such as in the word SWEATER, SCISSORS, and STAIRS. I decided not to alter the story in order to not make my subjects suspicious of the unusual print-out format of the story. They read the story directly from the children's story book collection. The story itself is also well suited, because it was translated from Swedish and thus it is much more likely for the participants to be unfamiliar with the story than it would be with American children's books. I chose to use a children's book for two reasons. First, these stories, as they are designed for children, are designed to be easy to read, understand and memorize. Moreover, they offer content that is accessible to anyone, no matter what their literary preferences might otherwise be. An interest in and understanding of the story not only increases comfort and ease, it also is crucial to the open interview that followed.

I handed participants the book and asked them to read it to me at their own speed. The book is about a five-year-old girl named Lotta and her family. In the story, Lotta has a bad dream about her siblings mistreating her stuffed toy and gets into a bad mood over it. As a result, she refuses to get dressed, sulks at the breakfast table and eventually decides to cut the sweater she was supposed to wear into pieces.

Content-wise, it is well suited because it is relatable to anyone that children may show exactly this kind of behavior and yet, the ending is so extreme that it is likely to have people react and comment on it. This content is relatable on two levels; they can either still relate to being the child in a bad mood or to the guardian having to deal with irrational behavior.

The story included the following target items: TROUBLEMAKER STREET, STRIPED, STRETCH. Further words of interest included are STAIRS, SWEATER, SCISSORS and START.

4.2.3.2.3 Narration

After the reading activity, I asked my participants to retell the story with as much detail as they could remember. This provided a second possibility to pronounce some of the target words in a different cognitive and articulatory activity. This activity and the last produced a varied range of output per participant both in terms of length and in terms of what type of reaction they had to the story. While some speakers actually re-narrated the stories, others provided comments or evaluations of the story. Within this section I also asked for details to elicit further target items. Frequent questions included: *Do you remember where the family lives*? and *What did the sweater look like?* Since no priming effect had been indicated before, I also asked: *Would you say Lotta was frustrated?* and

Would you say the mother was stressed out? I elongated the interview further by asking participants what they thought of the story and how realistic they thought it was.

4.2.3.2.5 Debriefing

Once the comments on the story exhausted the content of the book or the willingness of the speaker to further analyze the story, participants were debriefed. I verbally explained to them that my research focus is /str/-retraction and that I will analyze these patterns. Once they agreed to further participate in the study, I provided examples of /str/-retraction and asked them if they heard a difference. Depending on the flow of the conversation, I then proceeded to simply listen and have a conversation with them or asked further questions. Commonly asked questions included:

1. Have you ever noticed this pattern before?
2. Do you have a certain perception of this pattern?
3. Do you have a preference for one or the other?
4. Does it sound wrong to you? Does it sound like a speech impediment?
5. Do you know a person who pronounces /s/ this way?
6. Can you think of a particular group of people that would use this pattern?

4.3 Procedures

Participants were recruited in various cultural and community centers as well as the public sphere. Some were encountered through friend of a friend recruiting. Another part of the data stems from both visitors and workers at various recreation centers, whose locations are described in section 4.2.2 above. I interviewed seniors in different activity groups there, but also parents within the junior programs. Furthermore, I was able to record interviews at the cartography firm owned by a neighbor as well as at the University of Texas. Most of the college students, who were interviewed at the University, were undergraduate students recruited in a class on introductory linguistics. Although this may arguably cause some bias, the fact that the variability in pronunciation is assumed to be below the level of consciousness and was well guised in the study, allows me to regard this bias as negligible.

Once a participant agreed to do the interview with me, I asked them into a separate room, made sure to create a welcoming and relaxed atmosphere through casual conversation and then presented them with the consent form. I first gave them time to read it and then verbally repeated the important points, before I asked them to sign the form if they were still willing and able to participate.

Next, I had them fill out the social survey. For the duration of this, I interfered as little as possible to avoid influencing their choices. I set up the slide show as soon as they completed the social information survey. They first watched the slide show and were asked to provide a name for every concept presented to them was they saw it. This naming during the slide show was an idea I came up with only after participant ten. I therefore have only recalled naming not visual naming in the data set for participant 1-10.

After the slide show, I asked the participants what they remembered from it. To boost their confidence, I made sure to call the unordered slide show A the “hard one” and was very reassuring during their recall activity. Once they said they couldn’t remember

anything else, I started Slideshow B and reminded them that they would likely do better on this one due to the grouping of items. Once again, they had to name the items during the slide show and were then asked to recall them. When they stated they had remembered everything they possibly could, I handed them the booklet with the story and asked them to read it aloud. I also reminded them that they would have to narrate the story afterwards in as much detail as possible.

I started the narrative portion of the interview by asking participants what had happened in the story. Depending on how detailed this narration was, I asked them multiple questions about the story as described in the instrument section. I then gradually led them into an interpretation of the story, especially of the behavior of both protagonists.

Once this part was completed, I verbally explained what the study was about and why I had not told them about it. I presented the participants with the debriefing form and asked them for continued consent. Their agreement to continue was then used for a meta commentary on the phenomenon of /str/-retraction. Due to the fact that many participants were entirely unaware of the phenomenon, these comments varied greatly in length and depth.

4.4 Measurements

All the personal information per participant was collected in a survey that they filled out by themselves. This aids in self-categorization rather than a discourse approach, which may be influenced by perceptions of desirability.

The interviews themselves were partially transcribed by hand in text grids in Praat. Partially transcribed, in this case, means that I transcribed only those breath groups that had the desired sibilants. The text grids and corresponding audio files in .wav format with 44.1 kHz sampling rate were then chopped into those breath groups and force-aligned in Praat to create automatic phoneme boundaries for all orthographically transcribed phonemes in the data. A Praat script was then applied to the resulting force-aligned file pieces to measure the Center of Gravity for each sibilant in the dataset. This script recorded the participant, the file name, it grouped all sibilants into the defined Items "sk" "skr" "st" "str" "s" "sh" "sp" "spr" "sht", it recorded the word in which the item occurred, the style in which it was uttered (based on the categories naming, reading and narration indicated during transcription), two phonemes before and after each target phoneme (both /s/ and /ʃ/), the start and the duration of each individual sibilant. It then calculated the center of gravity for the midpoint of each sibilant in a Hamming window with a pre-emphasis filter set to 750Hz[8]. The pre-emphasis is different from recent acoustic analyses, but was determined to cope well with noisy recordings without excluding certain frequency ranges entirely, as was done by Wilbanks (Gylfadottir 2015; Wilbanks 2017).

In line with Jongman's findings and most of the studies on retraction thus far, assumptions about place of articulation for the sibilant will be based on Center of Gravity

[8] The entire script that was used to perform all these steps was developed in cooperation with Anke Dittmer and can be found in the electronic appendix. All of the executive decisions were mine. Further data wrangling and all data analysis was entirely performed in R and by me.

(CoG) measurements (Jongman, Wayland & Wong 2000). This results from the belief that the most characteristic trait of sibilant noise is its concentration of acoustic energy at varying frequency ranges with fronter productions showing the most energy in higher frequencies and backer sibilant productions showing most energy at lower frequencies. The CoG is the mean of all energy across a spectrum as described above.

The social information from the survey was finally added to the quantitative measurements provided from the analysis of the audio chunks. The resulting data is taken as the basis for all the descriptive and inferential statistics described in the following.

4.5 Classification

Chapter 1 above has introduced the complexity of social factors in language change and the theoretical assumptions we attach to it. In describing Austin and the sample in particular, I have used many of the terms from previous research and demographic descriptions on official information. There are two classifications of identity in the following analysis that are crucial and need some explanation. First of all, I use certain ethnic classifications for speakers. While African American and Hispanic-Latinx are heritage-related categories, White Anglo or European American speakers are simply called White. This appears to be the meaningful distinction in everyday life, questionable as that might be. Being White creates a difference in how you are being treated and thus perceive and are perceived in the world, no matter whether you strongly identify as Texan or Southern or West Coast. This type of treatment is what the ethnicity category is trying to gauge as a quantitative approach to language.

I have described the importance of identity, stance, social salience and individuality above. The argument for calling speakers African American instead of Black, is that they have a particular history in the US and especially in Austin that distinguishes them from Black people, who are not African Americans. The category of Hispanic Latinx contrasts with this approach and with the many experiences of different generations as immigrants or citizens, it is even harder to argue for a unified category. Nonetheless the interaction with other ethnic groups is based on being perceived as Hispanic-Latinx, especially with the bilingual speakers interviewed here. It is not simply being “Brown” but the expected connection to a certain culture that makes the label Hispanic-Latinx more valid. Thus, I use ethnically informed labels for speakers in this analysis not because I believe they should be treated differently, but because these categories have been shown to meaningfully influence both the reality and identity formation of speakers and their linguistic patterns.

Similarly, I classify all my speakers as male or female and do not include sexuality as a category. Non-binary classification was considered, but none of the participants appeared to identified as non-binary or addressed any concerns with this identity category. There are a couple of participants in the data set who may not be heterosexual. However, I valued their comfort in the experimental situation and positive experience of participating over including this further category in my data. I think it takes a different type of experimental set-up and a more qualitative approach to provide the right environment to talk about sexuality. I did not want it to be something they felt forced to disclose.

Nonetheless I believe that this is an interesting and important further dimension in any research relating to /s/ and I think future research should attempt to include this factor.

5. Results

The data gathered with the methods above yielded a corpus of 8.425 observations of sibilants uttered by 82 participants. To cope with the imperfections in the acoustic recordings and the fact that one speaker did not yield any interpretable measurements, I deleted the data of one speaker and performed Tukey's method to those measurements per speaker that were expected to be more stable. This means that I deleted all items of /s/ and /ʃ/ per speaker that were 1.5 times out of the interquartile range of the overall dispersion of data points. I also deleted all items with a duration shorter than 0.4 seconds in line with Wilbanks and Gylfadottir (Gylfadottir 2015; Wilbanks 2017). For certain analyses of specific factors described in other subchapters, I also excluded non-representative speakers and will describe these subsamples in greater detail then.

Including both the encouraged words and the spontaneous usages of the /str/-cluster, participants uttered the words listed in Table 5.1 with the corresponding frequencies. To simplify, I only included those words in the table that were uttered more than 5 times.

Table 5.1: Words with a frequency >5 in the corpus

Word Context	n
CONSTRUCTION	184
FRUSTRATED	19
FRUSTRATION	9
RESTAURANT	72
STRAW	29
STRAWBERRIES	44
STRAWBERRY	95
STRAWS	72
STREAM	6
STREET	270
STRESSED	21
STRETCHING	77
STRING	7
STRIPED	168
STRIPES	14
STROKED	79

The following chapters describe the patterns found in these observations with a focus on /str/-clusters. The analysis starts with a general description of data dispersion and the nature of sibilants in the data set. It addresses the normal distribution of the data and focuses on the differences between pre-consonantal /s/, pre-vocalic /s/ and the extent to which specific consonant combinations have a significant influence on the retraction of the sibilant. A further question to be established in this section is how the distance between /s/ and /ʃ/ can be described. Following this definition, the focus is on internal factors of language change. Gradualness, as expected by Bybee in both phonetic development as well as word frequency differences, is then evaluated based on findings of the sibilant space analysis (Bybee 2002).

The following section of the analysis focuses on a descriptive statistical account of further social factors in the description of /str/-lowering. Together with the linguistic factors, the final mixed-effects regression model is then be developed and fully evaluated.

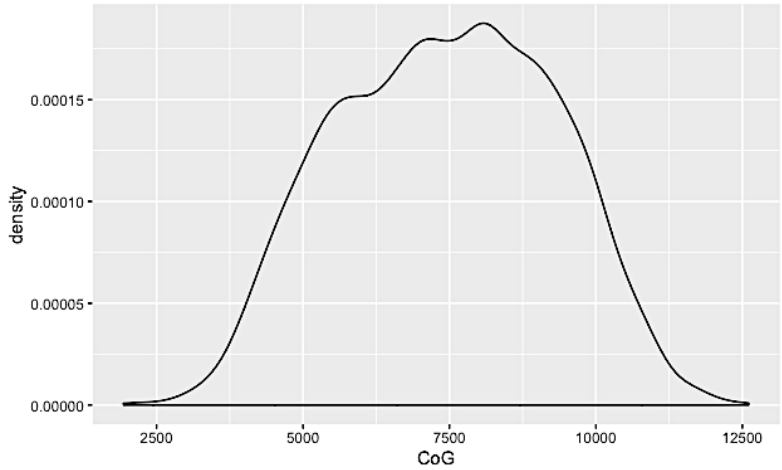
In a qualitative extension of the insights gained through the regression analysis, a brief thematic analysis lastly summarizes the patterns found in the commentary part of the interviews. This aids in drawing final conclusions on how to treat the observed lowering of /s/ in the triple consonant clusters in terms of an ongoing change in the sibilant landscape of American English.

5.1 The sibilant space

The first question in evaluating the behavior of sibilants in this dataset is how the overall sibilant space is shaped and where most of the data is located in this distribution. Figure 5.1 shows that most Center of Gravity (CoG) measurements lie between 5.000 and 10.000 Hertz and are relatively evenly spread between these thresholds. This exemplifies the meaningfulness of CoG as measurement for sibilants and the threshold for sibilants when measured through CoG. The graph in Figure 5.1 shows that there is a certain shape in the frequency ranges- even without the standardization procedure used to establish z-scores. Z-scores in this analysis are calculated based on the per-individual overall sibilant mean and standard deviation.

Although descriptions of patterns in the following sections are mostly based on the calculated z-scores and not the raw CoG values, it shows that the normalization procedure does not under- or overestimate the shape of the data. The pattern visible, although somewhat spread out, is also an indicator for a normal dispersion of values. This is exactly the expectation for a continuous variable such as a sibilant's CoG, because there is a meaningful threshold within which the sibilant is produced, but this threshold can never be exact, resulting in this bell-shaped curve.

Figure 5.1: Dispersion of all sibilant data points measured



The spread at the top further shows two plateaus that are indicative of further meaningful subsets in the data which peak at different points. The first plateau occurs at roughly 5500 Hz, the second at roughly 6500Hz with a final peak at 8000Hz. The appearance of these plateaus is a first indicator that there are two meaningful distributions within this larger dataset, connected through some more gradual values.

Outside of this threshold, there is visible skewness at both ends. The graph provides first evidence that the pattern is not well described when all data points are treated as one continuous dependent variable, but that there is a level of gradualness involved. The

following subchapters will introduce further distinctions both between subgroups of phonetic environments and their dispersion.

5.1.1 Group-wise analysis of sibilants

The previous graphs took the entire sibilant space under consideration, including all measurements of sibilants in the corpus. However, there are certain meaningful grouping factors to be expected in the data set. While it is technically possible to produce sibilants on a continuum from alveolar to post-alveolar tongue positioning, a difference between /s/ and /ʃ/ is expected based on quantal theory and English phonotactics as described in Chapter 2. This expected difference creates meaningful minimal pairs in pre-vocalic position.

Much of the research thus far has hypothesized mainly about the involvement of /r/ in the process and shaped the effect structure accordingly (Gylfadottir 2015; Wilbanks 2017; Phillips 2018). There are several hypotheses that need to hold for this analysis of /str/-retraction. First of all, the difference between productions of /s/ and /ʃ/ needs to be meaningfully differentiated. If there is no significant difference between the measurements of /s/ and the measurements of /ʃ/, the resulting assumption would be that CoG measurements cannot quantify differences in place of articulation or that there is great and unordered heterogeneous variability in sibilant production as a whole.

However, if these sounds are significantly differentiated, but show differences in variance, it is worthwhile to continue to hypothesize about the factors that cause these differences. Once they are established, a secondary question is whether the phonetic environment of the phonemically alveolar sibilant shows greater variability than the post-alveolar sibilant, which is restricted to pre-vocalic position. My assumption rests on the notion that certain environments allow for the /s/ to be retracted to an extent that makes it /ʃ/-like. If the variability is greater for the alveolar sound, I can assume very generally that the phonotactic ability to appear in conjunction with other consonants allows for harmonic processes. Once this variability is established, I will try to account for it by finding meaningful patterns in the phonetic environment of the sound.

Greater variability is to be expected for those sounds that appear in clusters, as they are not parts of minimal pairs in the way that was described above. What is interesting then is to see if a binary split of the data set suffices to create the best model or whether there are more than two meaningful groups within the dataset.

5.1.2 Differentiation between /s/ and /ʃ/

The underlying assumption of my analysis is that there are significant acoustic differences between the two structures we distinguish as alveolar and post-alveolar sibilants. If the variability overlapped to an extent where means would not be significantly different, it would falsify the assumption that CoG measurements suffice to detect the expected pronunciation differences in sibilants. In a first step, I thus needed to establish whether all items with the phonotactically expected phonemes /s/ and /ʃ/ were significantly different from each other.

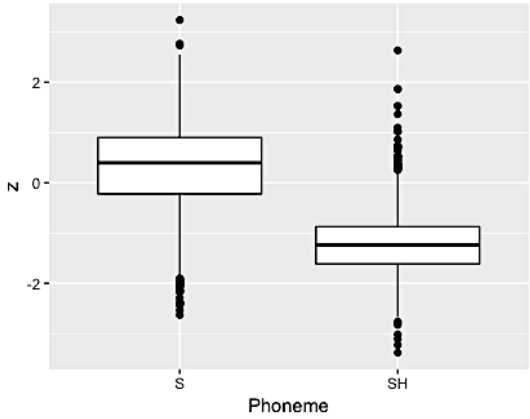
Table 5.2 and Figure 5.2 indicate that this difference in dispersion is existent in this dataset, since mean and median of the z-scores are different. Furthermore, the variance of /s/ is almost twice as high as that of /ʃ/.

Table 5.2: Characteristics of Phonemes

Phoneme	n	mean	median	variance
/s/	6215	0.248	0.363	0.770
/ʃ/ (sh)	1328	-1.162	-1.208	0.372

The boxplot further visualizes that the interquartile range of the data shows no overlap, indicating that /s/ and /ʃ/ can indeed be differentiated based on standardized CoG measurements.

Figure 5.2: Interquartile Range of z-scored CoG measurements per phoneme



In order to establish whether there is a significant difference between the measurements of both sibilants, I conducted an analysis of variance using the lm() function of *lme4* (Bates, Mächler, et al. 2015). The graph shows that the medians of both groups generally seem to be significantly different from each other, but a number of outliers occur outside of the general dispersion of each. The model outcome indicates that z-scores change significantly as a function of phonemes, with a difference in means of -1.514. This is significant at $p<0.001$. The model has an Adjusted R-squared of 0.292 and thus describes 29% of the variability in the data. The residual standard error is fairly high at 0.837, which means that the model with the current parameters does not predict the z-scored CoG outcome well.

I have established through modeling that the measurements of the sibilants are good for the differentiation of groups within the data, but the variability in the alveolar sibilant causes an insufficient model and needs further exploration.

I therefore include only those items in further analyses that stem from expected alveolar lexical items. Model comparisons establish which patterns best describe the data and thus provide first insights into the effects of surrounding sounds.

5.1.3 Phoneme Context and Effects

There are several possibilities of how the following segments could influence the z-scores measured for alveolar sibilants. Effects can hypothetically range from differences created by different places of articulation in the following segment, different places of articulation in the immediately following segment and the sound following it, show no pattern according to groups of sounds or be affected by specific sets of sounds. In the most extreme

case assumed by Janda and Joseph, namely the strongest influence being that of the triple consonant cluster /str/, a strong effect for this combination, some effect for other phonotactic triple consonant clusters and possibly an effect of the type of following sound would be expected (Janda & Joseph 2003a). The assumed model parameters would then show strongest effects for:

t*r > p*r, k*r > t > p,k > pre-alveolar consonant > (post)-alveolar consonant > back vowel > front vowel

If the assumption is that /s/-retraction is not a case of consonant harmony but rather a case of direct assimilation to affricated /t/, an effect based on the place of articulation of just the following segment would be the result. This would also result in an effect of consonants other than those phonotactically able to include /r/, such as the word SWEATER.

The second phoneme in the segment can also be investigated to evaluate consonant harmony. A pure case of consonant harmony would successfully model frequency lowering with just the presence of /r/ in the third position of the cluster. If a model in which the interaction of /t/ and /r/ becomes more significant than separate fixed effects for Phoneme.1 /t/ or Phoneme.2 /r/, the conclusion is that neither assimilation at a distance nor direct assimilation creates retraction.

There are several models that were used to test these hypotheses. All of the following models were carried out as mixed-effects linear models in R using the function lmer() in the lme4 package (Bates, Mächler, et al. 2015). This type of statistical modelling was chosen, because I assume speaker-based and item-based variability that simple linear models would not be able to elucidate.

In a first trial of modeling phoneme context in sibilant production, many of the models yielded a singular fit. Since I am focused on language internal factors in this section, I decided to test the significance of the random effects structure based on an average grouping structure (ST, STR, SP, SPR, SK, SKR, Sconsonant, Sfrontvowel, Sbackvowel). This test was carried out with the rand() function of the lmerTest package(Kuznetsova, Brockhoff & Christensen 2017) in R(R Core Team 2019). Results are shown in Table 5.3:

Table 5.3

Random Effect Significance Test

Random Effect	Log likelihood	AIC	P-value
none	-6904.6	13835	
Participant	-6905.8	13836	0.115
Word	-7023.4	14071	<0.001

Table 5.3 reveals that the inclusion of individual differences does not significantly improve model fit. The result is the same in a direct likelihood ratio test of the models with and without the participant effect. This is surprising as I expect individual differences in the production of at least /str/. However, the lack of significance can be accounted for by the fact that the variability in /s/-production is affected to a greater extend by other factors and the random effect for word than by random effects of the individual. It is further

possible that there are not enough items per participant to support such a detailed analysis, as the research design does not represent all sibilant environments equally per participant. The z-scoring also masks larger individual differences. My hypothesis rests on the importance of the /str/-cluster, which appears in the dataset in more comparable numbers across individuals. I thus treat the close analysis of all sibilants as a stepping stone to establish the linguistic importance of /str/ within the sibilants space and then proceed to evaluate the social aspect of this particular cluster in the following chapters.

Once the singular fit of models was accounted for through these significance tests, I therefore proceeded to calculate models with a random effect for word at this stage. Once I designed all possible models from lowest to highest possible grouping structure, I aimed to establish the goodness of fit that the grouping structures provide.

There are some issues with model comparison at this stage. Both likelihood ratio tests and information criteria are designed to evaluate the goodness of fit for a particular dataset and nested models. They are particularly useful because they penalize overly complex effect structures (Müller, Scealy & Welsh 2013). A revised definition of fixed effects parameters, as created in this design, is however not nested, even if the same dataset is used. The only measurement of goodness of fit that may be used to compare non-nested models or even models based on different data sets is R2. Methods to calculate R2 for mixed effects models have only recently been developed (Nakagawa & Schielzeth 2013; Nakagawa, Johnson & Schielzeth 2017). However, R2 does not penalize overfitting and complexity to the same extent as information criteria such as the AIC (Stauffer 2007: 267).

The model comparison presented below therefore has to be understood as a close reading of model parameters and criteria that are not necessarily designed for this type of comparison. Models expect parameters to be theory driven and not variable. Instead of assuming significant improvement of model fit due to adding effects, I attempt to gauge the above predicted effect hierarchy through this model comparison under the assumptions that greater variance explained (R2) can reliably identify better structures and magnitude of AIC difference between models and the minimal model can be used to further evaluate which models may be overfitted. All models presented here are significantly improving the fit compared to the minimal model based on likelihood ratio tests. Each AIC value presented in Table 5.4 was calculated based on comparing the respective model to the null model. The results are to be read accordingly in the table.

I thus focused on finding the model with the best possible fit according to AIC and R2. The AIC and BIC differ in their assumptions about the desirability in a statistical model. On a basic level, BIC identifies the model "with the highest probability of being the true model for the data, assuming that one of the models under consideration is true"(Kuha 2004: 216) while the AIC "explicitly denies the existence of an identifiable true model and instead uses expected prediction of future data as the key criterion of the adequacy of a model" (Kuha 2004: 217). Kuha underlines the positive aspects of using both criteria in model selection, arguing that they differ in their assumptions about the desired model and penalize the choice of too many variables differently. He describes that "agreement between them implies that the choice is insensitive to quite dramatic changes in the informativeness of the priors of the model parameters."(Kuha 2004: 224). However,

Burnham challenges the theoretical idea that data in life sciences can ever be modeled in one true model and thus advises to use AIC for model selection (Burnham, Anderson & Huyvaert 2011). The best models will thus be ones with low numbers for both the AIC and the BIC, although some models may be preferred by one of them.

Generally speaking, I expect models with more nuanced parameters to outweigh less complex models until the number of parameters is greater than the meaningful patterns in the dataset. If the expected factor hierarchy held true, I should thus find the biggest increase in goodness of fit in a model that includes the first three stages (str>spr,skr >st, sp, sk) with decreasing significance in goodness of fit for other factors of surrounding elements. This would mean that the patterns described more meaningfully model the data at hand than additional factors.

From smallest to biggest possible model, I regrouped the environmental factor as follows. Model VC included the following environment as binary segments, differentiating between vowels and consonants. Model TRI and model ITEM grouped the findings as one factor with the differing levels from the first three stages (ITEM) and only those involving the cluster I focus on (TRI). Model PoA1 grouped the following environment by place of articulation as labial, alveolar, post-alveolar, front and back, Model PoA12 did the same including an additional fixed effect for the adjacent sound (second after /s/) and Model PoA12I included an interaction between sound 2 and 3. Model PoA12RbinI did the same with interaction, but grouped the second sound as binary variable classifying the sound as /r/ or not /r/. Model PoA12R differed only in the fact that the second sound was classified as either front vowel, back vowel, /r/ and consonant.

All consonants other than /r/ indicate that the following sounds form a new word. Since Janda and Joseph have argued that the retraction even happens across word boundaries, these sounds are not regarded different from those within word boundaries and are thus included in this model (Janda & Joseph 2003a). The largest possible models were models Phoneme1, Phoneme12 and Phoneme12I since they include all phonemes. These models are expected to be overfitted. However, they were calculated to show the entire spectrum of possibilities in search for the best possible model. All models are described in Table 5.3.

Table 5.4: Mixed effects regression models per dependent variable structure[9]

Model	Fixed effect description	Model formula	df	AIC	R2
0	Overall mean value	z~ 1 + Duration + (1|word)	4	13158	0.00
VC	Binary variable vowel/ consonant following	z~ vc + Duration + (1|word)	5	13118	0.04 0.08
TRI	S/ST/STR	z~ TRI + Duration + (1|word)	6	13101	
PoA2R	Place of Articulation of adjacent (as vowel(front/back) consonant(R/not)	z ~ PoA2R + Duration + (1 |word)	6	13107	0.07
ITEM	S, ST, STR, SK, SKR, SP, SPR	z~ ITEM + Duration + (1|word)	10	13086	0.11
MoA1	Stop, fricative, nasal, approx., vowel (2),	z ~ MoA + Duration + (1 |word)	10	13113	0.06
PoA1	Place of Articulation following sound	z ~ PoA + Duration + (1 | word)	11	13124	0.05
PoA12R	PoA following and adjacent(as vowel(front/back) consonant(R/not)	z ~ PoA + PoA2R + Duration + (1 |word)	13	13063	0.15
PoA12RbinI	PoA following and adjacent(as binary) with interaction	z ~ PoA * PoA2Rbin + Duration + (1 |word)	14	13076	0.13
PoA12	PoA following and adjacent	z ~ PoA + PoA2 + Duration + (1 | word)	19	13067	0.16
Phoneme1	Phoneme following	z ~ Phonemesurrounding.1 + Duration + (1 |word)	44	13141	0.08
PoA12I	PoA following and adjacent with interaction	z ~ PoA * PoA2 + Duration + (1 | word)	50	13081	0.19
Phoneme12	Phoneme following and adjacent	z ~ Phonemesurrounding.1 + Phonemesurrounding.2 + Duration + (1 |word)	92	13078	0.22
Phoneme12I	Phoneme following and adjacent with interaction	z ~ Phonemesurrounding.1*Phonemesurrounding.2+ Duration + (1 |word)	250	13145	0.26

I will now describe the most important findings from Table 5.4 and draw first conclusions. Keeping the expectations described above in mind, there are several hypotheses as to what has an effect on the data. First of all, all models improve the minimal

[9] All regression tables were created using the stargazer() function from the stargazer package (Hlavac 2018).

model without a specified effect for phoneme groups. The significant improvement in fit of the model VC including just the binary distinction of vowel and consonant for the first phoneme after /s/ shows that there is some information present in this distinction. Some of the theories assume an assimilation of the place of articulation of affricated /t/ or other sounds to cause retraction. The significance of both the simple manner (MoA1) and place of articulation (PoA1) models for the first following sound indicate that these types of direct assimilation may explain some variability. However, the magnitude of improvement in AIC is greater in all smaller models that include information about both the first and the second phoneme after /s/ (TRI, ITEM, PoA12R, PoA12RbinI, PoA12, PoA12I, Phoneme 12). Judging from the AIC difference magnitude, PoA12R is the model that best fits the patterns in the dataset. R2, as expected, shows a similar tendency to rate those models that include information about both phoneme positions best, but prefers those models with the largest degrees of freedom. This is at risk of overfitting as cautioned against by Bates et al. (Bates et al. 2018). Since comparing goodness of fit parameters in the manner presented here is extending the context within which these parameters are usually employed, I will refrain from choosing one best model. Nonetheless, several useful observations can be made. First of all, "good" models are representative of long-distance assimilation, because they all include knowledge about the second phoneme, even if this phoneme is just a binary distinction between /r/ and not /r/. In contrast to Gylfadottir's analysis that assumed the structure in the model ITEM, this approach provides a more nuanced picture of the phoneme combinations.

While the mixed-effects model is certainly suited to cope with the great imbalance of some of the groups created in the various models, these models can only serve as indicators of possible patterns and many of the more complex models warned about rank deficiency. More balanced data sets should be evaluated with the same technique to yield more reliable results. Nonetheless, I have gone beyond the hypothetical assumptions used for modeling by Gylfadottir and Phillips (Gylfadottir 2015; Phillips 2018) and was thus able to provide some evidence in the debate concerning the big bang of this sound change. Neither the effect of /r/ as the third phoneme in a cluster nor a division between all triple consonant clusters (STR, SPR & SKR) best predicts retraction. Rather, it is a complex interplay of post-alveolar consonants, /r/ and the combination of the sounds, as supported by the better ranked models.

I will now continue to evaluate this particular cluster and how well it and other clusters account for the patterns at hand. I will first elaborate on the dispersion of the data and then describe in detail how the cluster /str/ (per word) behaves differently than other instances of the alveolar sibilant in the dataset.

5.1.4 Item-based analysis of Retraction

A phoneme-based and an item-based quantile-quantile plot in Figures 5.3 and 5.4 show that normality is still present in the dataset, at least for the range that is present in the data (1st quartile -0.75, 3rd quartile 0.77). Not only do these graphs provide visual indication that there is normality in the overall shape of the dependent variable, but the previously established groups in the data are also normally distributed individually. I include the post-

alveolar phoneme here that was not included in the analysis in Chapter 5.1.3 to allow for a comparison with /str/.

Figure 5.3: Phoneme-based dispersion

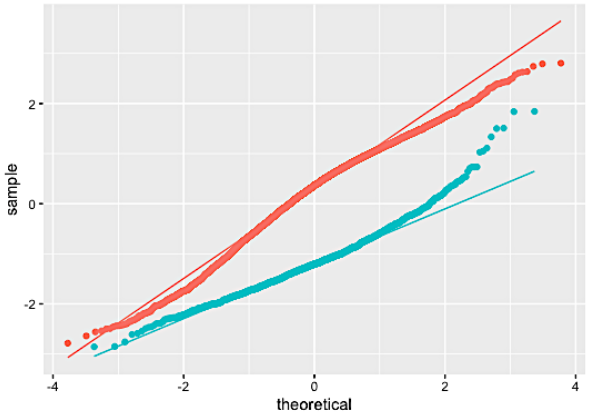
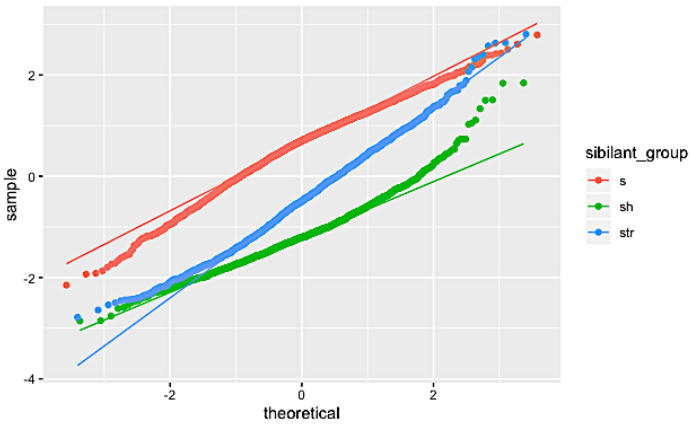


Figure 5.4: Item-based dispersion

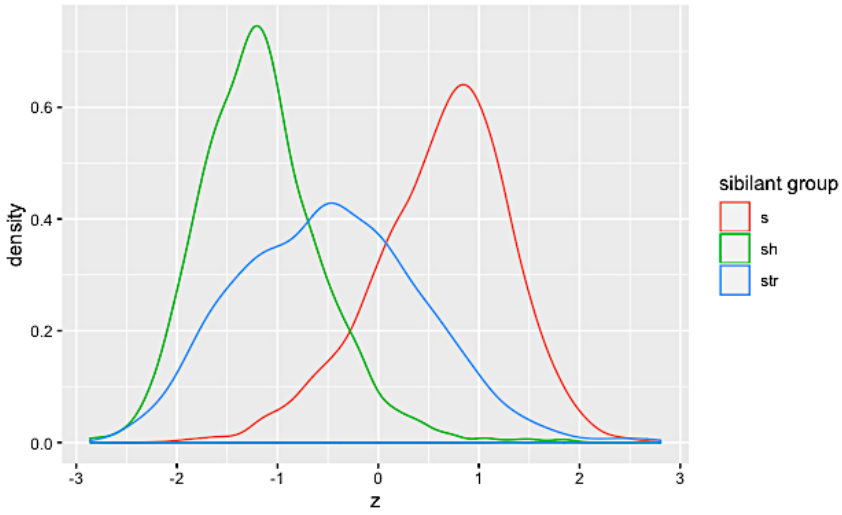


In direct comparison of Figures 5.3 and 5.4, there is some indication that /s/ is better aligned with the theoretical values where /str/ is accounted for. These patterns will be statistically evaluated below, after some further investigation of the dispersion.

Other indicators of data dispersion and per-item variability are delivered by mean and median. They are almost the same for each of the samples as portrayed in Table 5.5. Too many outliers or skewness of the dispersion would result in greater differences between mean and median. Somewhat more problematic for the visual inspection of normality in the qq-plots above is the knowledge of outliers. Considering that the interquartile ranges per sibilant type are between -1.7 and 1.1, the slope of points above 2 can be disregarded for marginality. It also shows that there is a meaningful frequency range within the dataset that differentiates the sibilants on the phoneme and possibly even the item level.

While the qq-plots above visualize the normality of the data in direct comparison to the idealized normal shape, the following density plot in Figure 5.5 visualizes the dispersion of the data in the normal bell shape.

Figure 5.5: Density of item dispersion



If a fully alveolar production of /s/ existed in /str/ clusters, all of its variability would mimic that of alveolar /s/ in all respects. However, this is not the case. The group-based Figure 5.5 shows that the production of /str/ covers almost the entire range from alveolar to post-alveolar production, which manifests itself as z-scores that range between -2 and 1 for the bulk of the data.

This is also visible in the structural characteristics of the dataset. Table 5.5 shows: While /s/ and /ʃ/ are fairly equal in their variance at 0.447 and 0.395 respectively, the variance of /s/ in /str/ is almost twice as large at 0.715.

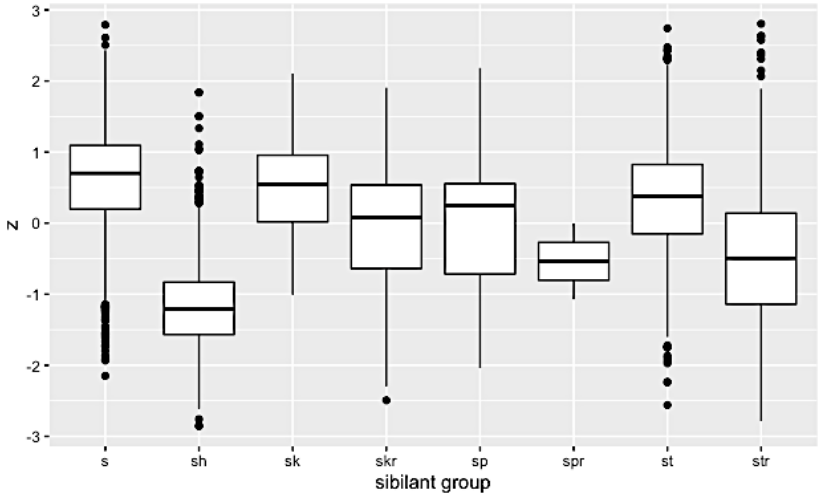
Table 5.5: *Variance of Items*

cluster	n	variance	mean	median	25th percentile	75th percentile
/s/	2920	0.447	0.628	0.696	0.199	1.086
/ʃ/ (sh)	1625	0.395	-1.207	-1.240	-1.614	-0.875
/sk/	134	0.493	0.458	0.528	-0.031	0.928
/skr/	429	0.527	0.039	0.089	-0.475	0.551
/sp/	19	1.071	0.111	0.329	-0.452	0.705
/spr/	3	0.391	-0.730	-0.946	-1.082	-0.485
/st/	1645	0.550	0.309	0.363	-0.176	0.824
/str/	1223	0.715	-0.371	-0.399	-0.973	0.185

Contrary to all other frequent items that have a variance that ranges around 0.4, /str/ is by far the item with the greatest variance ranging around 0.72. Furthermore, the mean and median of /str/ lie in between /s/ and /ʃ/. Considering that the distance between the mean of the post-alveolar and the alveolar sibilant is 1,835, the midpoint between these ranges would be -0.29. In reality it thus is slightly closer to the production of /ʃ/.

Interpreting additionally the following boxplot in Figure 5.6, it is visible that the variance for the item /str/ is generally much greater than the variance observed in /s/ and /ʃ/.

Figure 5.6: Boxplot of sibilant groups



It further shows that the interquartile range of /str/ does not align with the interquartile range of /s/ and slightly overlaps with the upper range of /ʃ/. /sk/ and /st/ have slightly lower IQR than /s/ and /skr/ shows both more variability and lower overall z-scores than all the other alveolar groups except for /str/. Due to the great differences in number of items, /sp/ (n=19) and /spr/ (n=3) should be disregarded.

Table 5.5 presents the results of a mixed-effects linear regression model used to evaluate the differences between the sibilant groups as presented above excluding the post-alveolar phoneme. The reference level is /s/. Apart from the visual inspection of the boxplot, this model establishes how comparable each of the groups is to one another. Baayen et al. argue that p-values calculated on the basis of the t-statistic in mixed models are questionable, because these statistics assume balanced data (Baayen, Davidson & Bates 2008). Luke shows that the Satterwhaite approximation that was used to calculate the values displayed in the table is at least the most robust method to lower the type 1 error rate of model assumptions (Luke 2017). I will thus proceed to interpret the significance levels displayed with caution.

Table 5.6: Sibilant group model with random effects for word and participant

Sibilant Group Model

Model: z ~ sibilant_group + (1|WordContext) + (1|Participant)	*Dependent variable:*
/sk/	-0.133 (0.150) t = -0.887
/skr/	-0.572 (0.113) t = -5.066***
/st/	-0.178 (0.061) t = -2.924***
/str/	-0.860 (0.072) t = -11.903***
Constant	0.543 (0.038) t = 14.316***
Observations	6,187
AIC	13,591.250
BIC	13,645.090
Note:	*p<0.1; **p<0.05; ***p<0.01

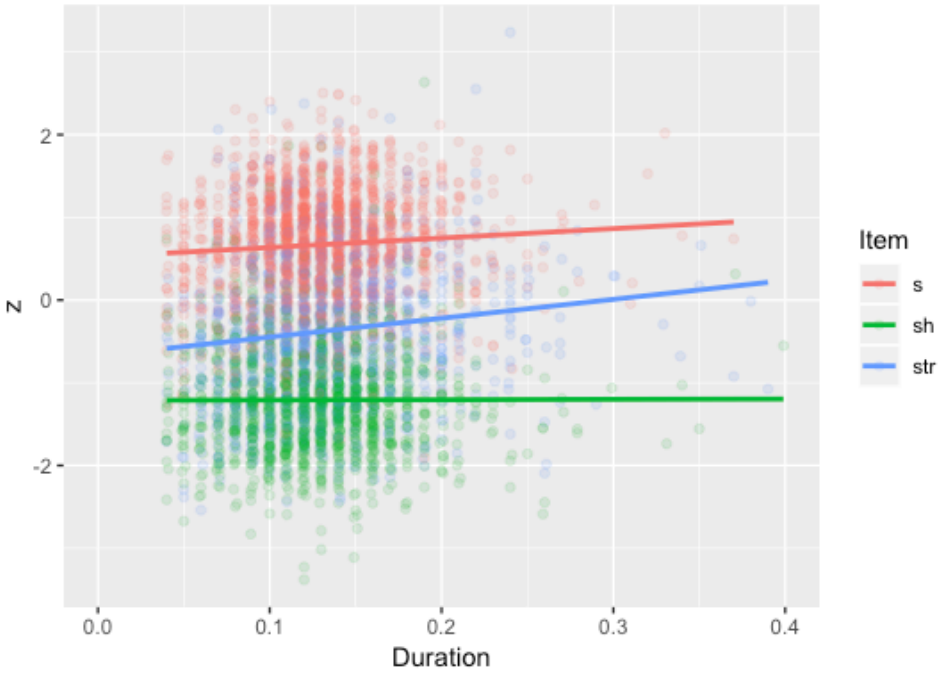
The strongest effect in Table 5.6 is arguably that of /str/ as expected, because it lowers the intercept by -0.860 whilst providing by far the largest t-value and thus significance. Note that, in contrast to the boxplot, this model indicates that only /st/ differs significantly from /s/ but not /sk/, /sp/ and /spr/ will again be disregarded in this analysis for group size reasons. All parts of the analysis above are evidence that the triple consonant cluster /str/ not only patterns meaningfully, but also has the most significant impact on the observed retraction of /s/. In the following, I analyze some more language internal factors i.e. duration and word context in the overall sibilant space. The remaining analyses in this chapter will then focus on retraction in the specific cluster environment that this research project was designed for, namely /str/.

5.1.4 Other linguistic factors

To further characterize the process of /s/-retraction linguistically, I included effects for the factors duration and word context of the sibilant. These factors were also controlled for in the models reported in Table 5.3, but this section will look at their effects in greater detail. I begin with a visualization of the effect of duration on the sibilants and /str/.[10]

The following scatterplot in Figure 5.7 further illustrates the variability of /str/: alveolar and post-alveolar productions are relatively meaningfully patterned across the entire duration range. While productions of /s/ lie between z-score levels 0 and 2, and productions of /ʃ/ mostly lie between 0 and -2, /str/ points are centered around 0 but spread across the entire sibilant range. A longer duration moves the production slightly further into the direction of /s/.

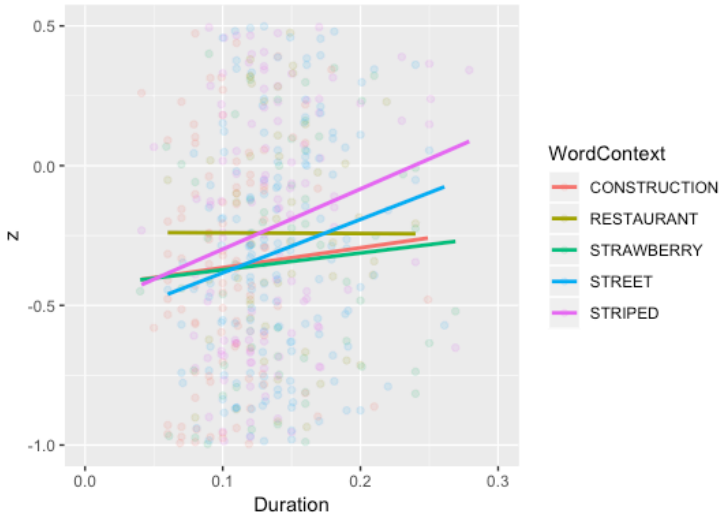
Figure 5.7: Z-score measurements per Item across duration



The necessity to include an effect per word is visible in the plot below. It plots the z-scores as a function of the duration for 5 of the most frequent words in the /str/-subset.

[10] A standardized number per participant would be preferable to model this relationship. Since estimates are calculated in mixed effects models, the random effect per participant does account for some of these expected individual differences.

Figure 5.8: Z-score across duration for most frequent words



Although the data points are fairly spread out, the regression lines illustrate that duration has almost no effect on *restaurant*, which is centered at -0.25, while it has a strong effect on the words *street* and *striped*. Although this plot clearly shows the differences in retraction patterns per word, we cannot include word context as a fixed effect, because this sample underrepresents the variability expected in reality. As Baayen et al. argue:

> "two core principles of the structure of language, the arbitrary (and hence statistical) association between sound and meaning and the unbounded combination of finite lexical items, guarantee that a great many language materials must be a sample, rather than an exhaustive list. The space of possible words, and the space of possible sentences, is simply too large to be modeled by any other means. Just as we model human participants as random variables, we have to model factors characterizing their speech as random variables as well." (Baayen, Davidson & Bates 2008: 390).

The inclusion of word context as a random effect accounts for this characteristic of words as non-exhaustive. Figure 5.8. supports the meaningfulness of this effect and also portrays its heterogeneity.

Table 5.7 below compares the model created for Table 5.6 with the maximum model that includes a further fixed effect for duration.

Table 5.7: Model comparison with and without duration effect

Sibilant Group Model Comparison

	z ~ sibilant_group + (1|WordContext) + (1|Participant)	z ~ sibilant_group + Duration+ (1|WordContext) + (1|Participant)
/sk/	-0.133 (0.150) t = -0.887	-0.128 (0.150) t = -0.849
/skr/	-0.572 (0.113) t = -5.066***	-0.565 (0.113) t = -4.981***
/st/	-0.178 (0.061) t = -2.924***	-0.173 (0.061) t = -2.815***
/str/	-0.860 (0.072) t = -11.903***	-0.858 (0.073) t = -11.821***
Duration		0.456 (0.210) t = 2.173**
Constant	0.543 (0.038) t = 14.316***	0.480 (0.048) t = 10.018***
Observations	6,187	6,187
Log Likelihood	-6,787.623	-6,785.268
AIC	13,591.250	13,588.540
BIC	13,645.090	13,649.110

Note: *p<0.1; **p<0.05; ***p<0.01

The direct comparison of the fixed effect structure shows that duration does significantly improve the model fit, which is visible in the improvement of the AIC and confirmed by a likelihood ratio test (p=0.03). The inclusion of duration does not lessen any of the effects in the simpler model, which provides further support that these patterns are quite robust. The complexity of the word context as a factor in retraction will be further evaluated for /str/ in the following subchapters. This first visual glimpse at its complexity was just included to underline its importance as language internal factor overall.

5.1.5 The importance of /str/

All of the models above have shown that /str/ performs significantly differently than other environments in which /s/ may appear. To this extent, it confirms findings of other acoustic studies (Gylfadottir 2015; Wilbanks 2017). What is different, however, is that the alveolar production was considered in greater detail. Rather than just testing the assumed clusters for significance, the detailed analysis of different structural possibilities for the fixed effects has shown that all clusters remain to have an effect, even if there is a further differentiation between these groups, sibilant-vowel and sibilant-consonant variation. Since the best model includes fixed effects for both places of articulation and their interaction, the complex interplay of phonetic context was further underlined. Further language internal elaboration keeps pointing to /str/ as by far the most significant environment for retraction. For the remainder of this chapter, I will therefore evaluate the ongoing sound change in this particular cluster.

5.2 Mechanisms of Language Change in /str/-retraction (Internal Factors)

In his book on internal factors in linguistic change, Labov characterizes the features of regular (sound) change and lexical diffusion. While his binary categorizations and the additional social descriptions as "change from above" and "change from below" seem to fall into place overly easily, I will establish how gradualness is at play in the change observed in my dataset. This will be done by evaluating lexical and phonetic gradualness, without taking any of Labov's categorizations as a given (Labov 1994).

5.2.1 Lexical Diffusion or Gradualness

Chapter three explained in detail how lexical diffusion as a concept rests on the understanding that frequency has an influence on how linguistic structures change. Bybee predicts that "changes that affect high-frequency words first are the result of the automation of production, while low-frequency words change first when the change makes the words conform to the stronger patterns of the language" (Bybee 2002: 271).

While the general idea of this frequency effect is appealing, there are several problems with describing word frequency in greater detail.

The first problem is frequency itself as a parameter. How can frequency be measured? Generally speaking, there are several levels on which frequency matters. It can be assumed that words have varying frequency levels within a language but also vary per individual. Further consideration is necessary to establish whether the speech samples gathered are good representations of frequency both within and across speakers or whether a global measure derived from a corpus is necessary. Erker and Guy (2012) caution that local measurements may come with sampling issues while global measurements may be problematic because they are often based largely on written language.

These problems are rooted in the nature of word frequency, which Guy and Erker describe based on Zipf's law. According to Zipf's analyses of several speech entities, the most frequent items behave differently than less frequent items (Zipf 1936). To Erker and Guy, this means that the distribution of words is "a power law distribution: there are very few words with very high frequencies and many words with very low frequencies" (Erker & Guy 2012: 529).

Therefore, it not only matters which measurement is used to describe frequency but also what type of variable this parameter is classified as. If I treat word frequency as a continuous variable, I underestimate the effect that this incongruent distribution may have and also negate a possible threshold between the groups of frequent and infrequent words. A continuous variable would assume that there is a linear relationship, while we are expecting a threshold based on theory (Erker & Guy 2012,). It is further quite difficult to estimate patterns of low frequency words due to their small numbers. This is generally the case in the present data set because some words, even if they technically are high frequency types, appear in very small numbers. While not necessarily a problem for modelling, it is a problem for assuming multiple patterns of variability based on frequency types.

The second problem with quantifying word frequency is the question which word forms to take into consideration. Both lemma-based and surface-based approaches are functional, but if we follow Bybee and Pierrehumbert in assuming an exemplar-based

model of phonological structure in cognition, surface forms need to be considered rather than lemmas (Bybee 2002; Pierrehumbert 2002; Erker & Guy 2012).

5.2.1.1 Modelling lexical frequency

The general problem with modelling these types of frequency-influenced changes are rooted in the continuity otherwise assumed for linear models. Linear regression models provide only two options: pooling data as if every word equally affected the outcome variable or not considering different words at all and thus modelling the notion that a change affects all words to the same extent. The latter would overrepresent the patterns in the most frequent words and thus overestimate those changes that are related to the automation of production, while it would underestimate those changes that analogically favor the stronger patterns of the language. The former on the other hand would possibly find no effects where there are some, because it would fail to predict the patterns in low frequency items. Mixed-effects models aid in this process, because they provide the opportunity to partially pool words. Yet accounting for the unbalanced words in a data set and taking it into consideration does not explain the effects of frequency that were described above (Barth & Kapatsinski 2018).

For the present dataset this would mean that the model accounts for the fact that STREET appears in the dataset more often than STRESS but it could not account for the fact that language users are far more likely to use the word STREET on a regular basis than the word STRIPED. It further could not predict that STRESS and STREET might present with one pattern while STRIPED appears in another and thus the model would expect different behavior from low and high frequency patterns. I therefore need to include item identity as a random effect but I also need to find a way to quantify word frequency.

Erker and Guy show that frequency can be measured both locally and globally. A global measurement is a corpus-based measurement. They argue that many corpora rely on "heterogeneous texts" and question "how relevant are these to the usage of a specific speech community or individual" (Erker & Guy 2012: 530). At the same time, they challenge the representativeness of local frequency norms. They wonder whether small chunks of data can provide a realistic sample on an individual's language use. Nonetheless, they decide to base their own research on local frequency measurements in pronoun usage (Erker & Guy 2012). Since the cluster investigated in this study is rather rare, I cannot properly predict an intra-speaker usage from the data at hand. Furthermore, appearance in the corpus is partially artificial due to the activities performed in the interview.

However, I can use appearance in other corpora as an estimate of how frequently the particular word is used in general by speakers of English. To improve the relevance that Erker and Guy criticized, I will base my frequency estimates on the SUBTLEXus corpus. In recent years, this corpus has been established as a good indicator of frequency in natural speech, because it is based on spoken language rather than written texts, is larger than other copora of spoken language (Fruehwald 2016) and has been used in some studies on sound change as a frequency measure (Phillips, Jacob 2016; Wilbanks 2017). Brysbaert and New compared the SUBTLEXus corpus to other corpora such as the BNC and the Michigan Spoken corpus and showed how much better SUBTLEXus performs statistically in explaining variance in lexical decision tasks. Although larger corpora of spoken language

such as the respective section in the COCA exist, many of these are based on very limited social environments such as talk shows and frequencies thus only poorly relate to everyday usage. Therefore, SUBTLEXus frequencies are best suited for frequency related approaches to language. Fruehwald describes this as "good psychological validity" (Fruehwald 2016: 386).

The SUBTLEXus corpus is made of American films and television shows and contains a total of 51 million words, of which 16.1 million are taken from television series, 14.3 million from films before 1990, and 20.6 million from films after 1990 (Brysbaert & New 2009).

There are several measurements of frequency available in this corpus. The ones of interest here relate to differing conceptions of frequency. While the above distinguished token frequency (as opposed to type frequency) is of interest, it also provides the opportunity to show the versatility and thus everyday frequency of the word with a measurement of context variability. This versatility is parametrized by the percentage of films in which the word appears described as contextual diversity ($SUBTL_{CD}$).

Erker and Guy work with three different measurements of frequency. All of them locally measured within their corpus, they test the effect of what they establish as raw frequency, logarithmic frequency and discrete frequency. The former two are continuous variables with log frequency being shaped in a way to better represent the dispersion hypothesized by Zipf's law. In their study, the frequency predictor that best described the data is the latter one, which is a binary distinction based on a threshold that differentiates between words that amount to more than 1% in the corpus and the rest. To visualize this split in my data, I created an indicating line in Table 5.8.

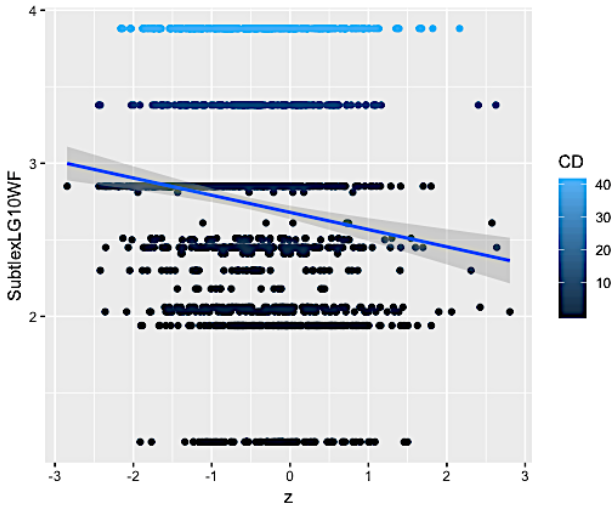
I also included both of the continuous standardized measurements provided in the corpus, which are $SUBTL_{WF}$ (WF) and Lg10WF. The former describes the frequency per 1 million words, the latter is based on log10(FREQcount+1) (Brysbaert & New 2009). To establish the best model, I further included the described measurement of word versatility, namely the contextual diversity (CD), which was expected to interact with the word frequency measurement.

Table 5.8: Three SUBTLEXus
Frequency measures for /str/-words in thus dataset

Word	WF (per 1 million words)	Lg10WF	CD (percent of films in which word appears)
street	148.14	3.88	40.67
restaurant	46.35	3.38	14.83
Threshold 1% of the corpus			
construction	13.84	2.85	5.42
string	12.67	2.81	6.02
stream	8.34	2.61	3.51
stripes	6.31	2.51	2.10
straw	6.24	2.50	2.93
stressed	5.55	2.45	2.72
strawberry	5.53	2.45	1.98
frustrated	5.00	2.41	2.54
strawberries	3.92	2.30	1.26
frustration	2.98	2.18	1.62
stretching	2.22	2.06	1.23
straws	2.10	2.03	0.94
striped	1.71	1.94	0.81
stroked	0.27	1.18	0.15

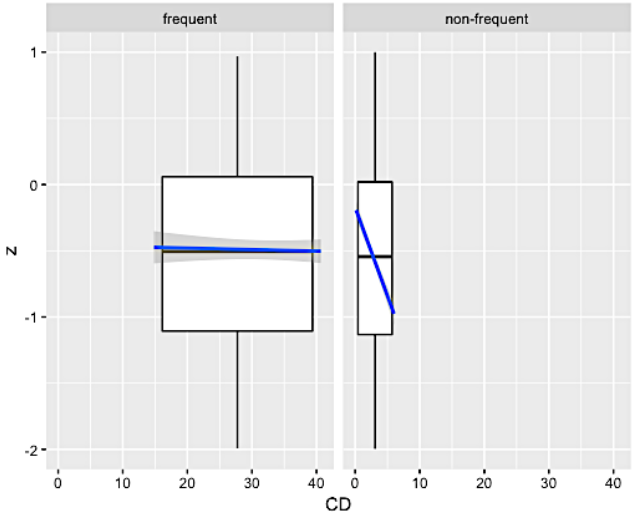
Before I developed the model, I first surveyed any effects that might be visible in an exploratory data analysis. Since the log word frequency is a continuous variable as opposed to the binary distinction, I used this variable as visualization and also included the CD in Figure 5.9. The regression line in the graph illustrates what is invisible to the naked eye when focusing on the points: Words with higher frequency scores and higher CD have a lowered z-score as compared to less frequent words, which means that participants produce /str/-items with high general frequencies on the lower end of their sibilant continuum.

Figure 5.9: Lg10WF and CD across z-score



The boxplot in Figure 5.10 divides the dataset into the binary frequency categorization and shows that although the spread remains similar, non-frequent words are highly influenced by context dependence while the two frequent words show no such patterns.

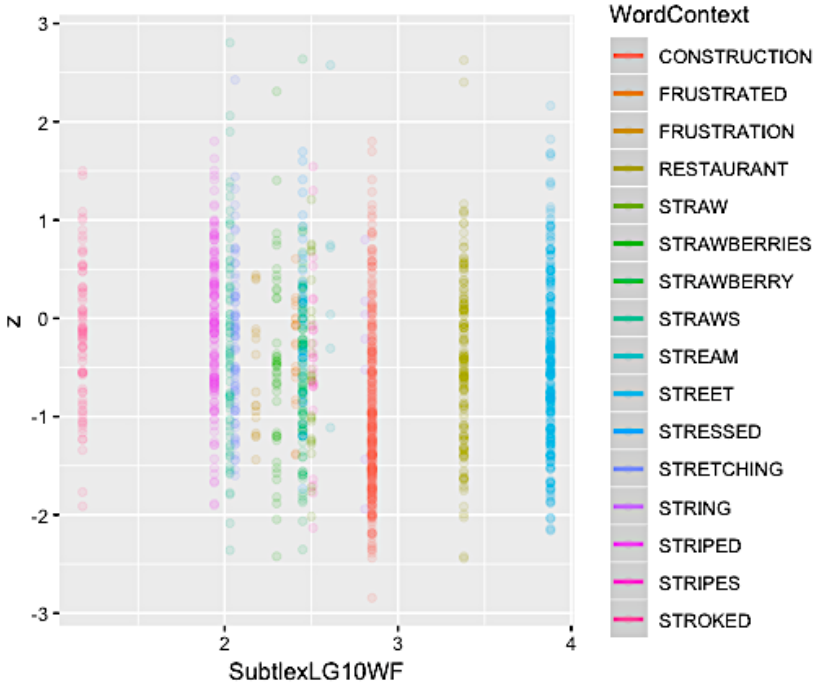
Figure 5.10: Binary Frequency over CD and z-score



The boxplot illustrates that CD and binary frequency are clearly connected. Non-frequent words have values lower than 10 for their CD while frequent words are spread somewhere between 15 and 40. The regression line further indicates that the behavior of frequent words is fairly stable ranging around z-score values of up to -0.5 while the z-score of the non-frequent words is highly influenced by the CD value with lower values having a z-score representative of less retraction while higher values approach retracted z-scored of up to -1.

Figure 5.11 finally portrays words that appear more than 5 times in the corpus. The patterns are very similar to the patterns in Figure 5.9, which is partially due to the fact that the dataset only includes a small number of words with great variability in frequency. Nonetheless, there are a few results that work slightly against the linear trend when we look at the word CONSTRUCTION.

Figure 5.11: *Z-score based on log frequency and word*



It is apparent in Figure 5.11 that both the specific lexical item (e.g. CONSTRUCTION) as well as the word frequency (threshold at WF 2.5) may play a part in retraction, but both are not good indicators of retraction. Arguably, there is a comparable pattern for CONSTRUCTON, RESTAURANT and STREET such that there is a significant amount of values below -1 while the majority of points for all other words lie above this threshold. CONSTRUCTION is the most interesting word here, because it prevents a clear pattern from forming through these points. Here, its largest concentration of values lies below that of the two more frequent words. The theory described above thus holds true in stating that frequent words do behave differently. The amount of spread indicative of the variability however further indicates that lexical frequency is neither perfectly gradual nor general. Such a spread can be explained by great individual differences as well as word position.

5.2.1.2 Results for Lexical Frequency

Modeling lexical frequency as a factor indicating lexical diffusion is problematic in mixed effects linear regression models according to Barth and Kapatsinski. They show that a random effect for item usually steps in to explain the variability at hand when trends could also be explained by the frequency and not just the specific token (Barth & Kapatsinski 2018). I thus followed their recommendations in using simple linear models to establish the effect of word frequency on /str/-retraction first. I tested the effect for the WF measurement, the Lg10WF measurement and the binary division of data points. Out of all of them, the Lg10WF model was the only model with a multiple R^2 above 0(0.017), differing from Erker and Guy's discrete difference predictions (Erker & Guy 2012). Nonetheless, the R^2 indicates that while there is improvement compared to the null model, word frequency cannot predict the pattern of change well.

In a further step, I then calculated the mixed effects models to evaluate these fixed effects. In a simple mixed model with only the respective frequency measurement as fixed effect and a random effect for item and participant, none of the factors reach significance and a likelihood ratio test determines that none of the models significantly improves the fit compared to the null model. Generally speaking, this means that we can conclude that word frequency, if it has an effect on retraction at all, cannot be found to have a great effect on /s/-retraction with this type of modeling. Accordind to R2 measurements, the best model of the three is the model with the log frequency measurement (log: 0,017, cd: 0.004, dis: 0,0008).

There are other approaches to modeling an effect of word frequency however. Hay and Foulkes take a more "rigorous" approach in modelling the effect of frequency on language change. They argue that a demonstration of frequency leading change would result in successfully modelling an interaction of birth year and word frequency (Hay & Foulkes 2016). I took this approach in calculating the models presented in Table 5.9 below. The underlying hypothesis is that word frequency behaves differently through the generations participating in the change, rather than influencing all items equally. Compared to the reduced model used to gauge the best structure for the fixed effect of word above, better fit is to be expected once the factor describing the change pattern (brith year) is included. To fully establish the effect of frequency I included all social factors from the final model. This guarantees that no other effect will step in to explain the pattern described as frequency effect. These social factors will be further explained in the following chapter and are thus excluded from the table below to emphasize the factors being discussed here. Model (1) is the minimal model including only the random effects for word context as well as participant and all other language internal and external factors hypothesized about. Model (2) is comprised of the same effects but further includes an interaction between birth year and the Subtlex log. word.

Table 5.9: Model comparison including social factors

Model Comparison with Birthyear Interaction

	Dependent variable:	
	z	
	(1)	(2)
SubtlexLG10WF		5.016 (2.429)
		t = 2.065**
birth.year	-0.009 (0.002)	-0.002 (0.004)
	t = -5.552***	t = -0.604
ethnicityhispanic	0.042 (0.116)	0.039 (0.116)
	t = 0.363	t = 0.334
ethnicitywhite	0.224 (0.101)	0.223 (0.101)
	t = 2.219**	t = 2.220**
genderm	0.101 (0.074)	0.104 (0.073)
	t = 1.366	t = 1.413
Stylereading	0.101 (0.114)	0.081 (0.111)
	t = 0.884	t = 0.733
Stylestorytelling	0.103 (0.117)	0.093 (0.113)
	t = 0.879	t = 0.823
Duration	1.630 (0.521)	1.664 (0.520)
	t = 3.126***	t = 3.201***
SubtlexLG10WF:birth.year		-0.003 (0.001)
		t = -2.123**
Constant	17.776 (3.326)	4.016 (7.430)
	t = 5.344***	t = 0.540
Observations	1,318	1,318
R2	0,07	0,08
Akaike Inf. Crit.	3,166.457	3,163.848
Bayesian Inf. Crit.	3,223.479	3,231.238
Note:	*p<0.1; **p<0.05; ***p<0.01	

In this first glimpse at the final model we can see that Hay and Foulkes' assumption greatly improves the model fit. The AIC confirms that the added interaction in model (2) significantly improves the fit of model(1). The Satterthwaite t-tests performed in the lmertest package further confirm the significance of the interaction (Kuznetsova, Brockhoff & Christensen 2017). I further attempted to fit a model with random slopes per word frequency and participant, but this model failed to converge.

5.2.1.3 Discussion of Frequency Effects in /s/-retraction

The analysis above established successfully that frequency is a construct that provides challenges in both theoretical and operational realms. There is Zipf's law to take into consideration both theoretically and methodologically, understanding that we are not necessarily expecting a gradual and linear relationship. The categorization of words as frequent and infrequent was certainly prone to error for this specific data set, since only two words are considered frequent. The boxplot generally indicated that there may be a pattern if the data could better accommodate this notion.

The attempt to exclude random effects in order to better evaluate fixed effects was problematic. I assume that this is not a general flaw in the concept, but rather that the importance of social factors and the individual is too great to successfully model this change in a simple linear regression model. The further analysis with mixed effects has shown that frequency effects both theoretically and methodologically only have an effect, when they are closely linked to the changing generations as was modeled by the interaction of birth year and log frequency.

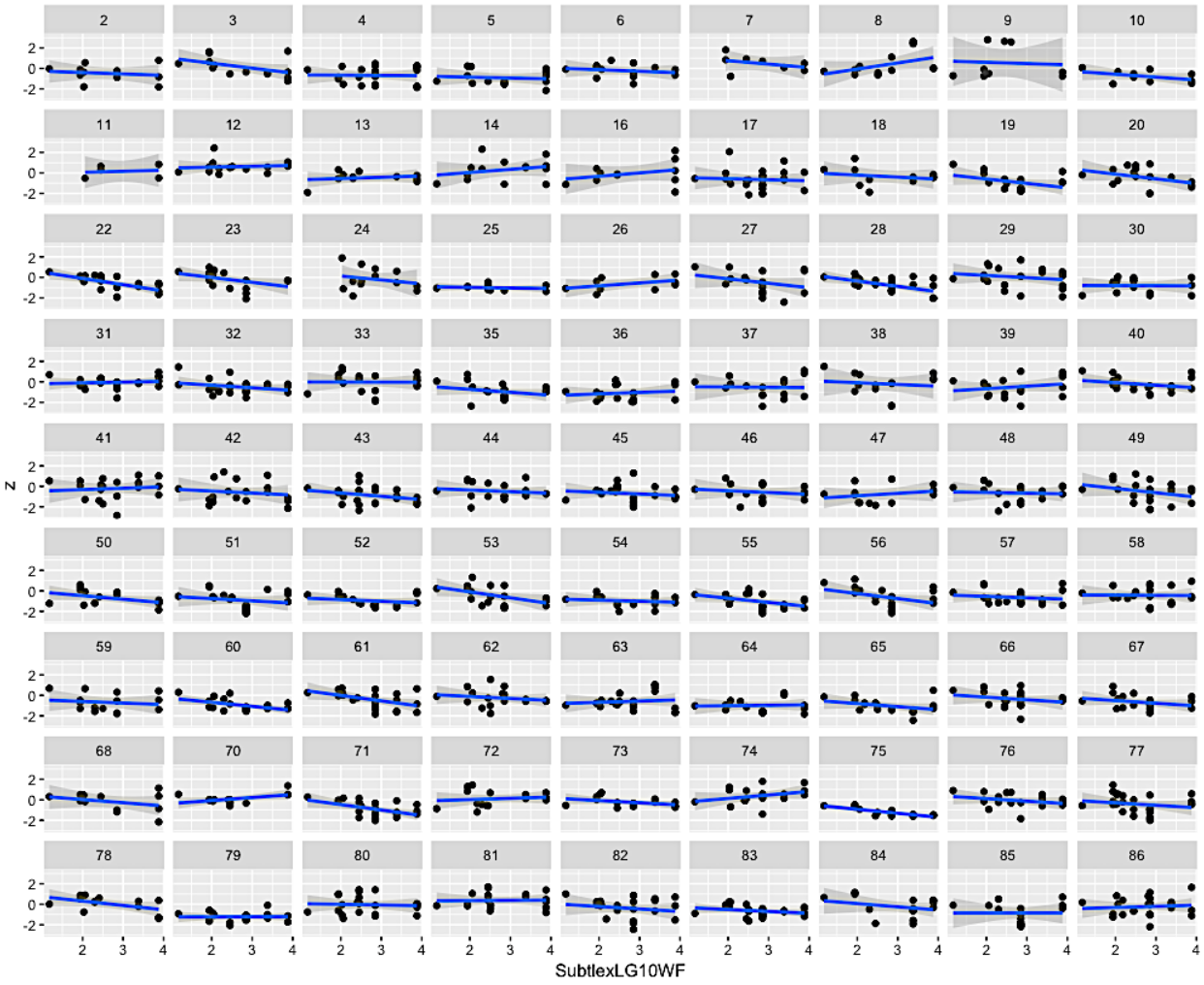
It is necessary to further recognize the special status of /str/-clusters in the language. They are far less frequent overall than any of the sound changes discussed as lexical diffusion changes by Labov and Bybee(Labov 1994; Bybee 2002). Furthermore, this particular dataset only includes two words that lie above the 1% threshold that Erker and Guy have found to be a successful category in establishing frequency differences (Erker & Guy 2012).

Another aspect to consider in understanding lexical gradualness is the individual. The vast difference between modeling with and without a random effect is an indicator that variability is greatly influenced by individual variation. I will further explore this point below.

5.2.1.3.1 Individual Variation in lexical patterns

To estimate whether the pattern seen across speakers here is also observable within speakers, I created Figure 5.12, which shows the z-score and lgWF per speaker as a scatterplot with linear regression lines. The specific pattern to look for is a decrease of z-score with an increase in word frequency and thus a downward slope. Some good examples of speakers showing this pattern in the graph are speakers 3, 10, 22 and 53.

Figure 5.12: Individual z-scores per Lg10WF

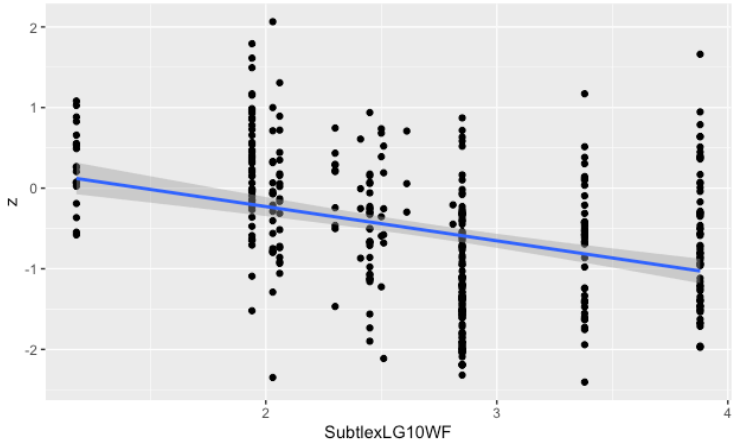


The graph provides a good estimation of why the R^2 is so small and thus insignificant for all of the models created without random effects. It truly depends on the individual both whether there even is a slope for the values as well as whether this slope goes down or up with the increase in word frequency. Therefore, there is a pattern of word frequency for some speakers, while others seem to base their pronunciation on the specific phonetic environment rather than the word or word frequency. I thus conclude that there are three groups in the sample. Some speakers are not showing signs of retracting /str/-clusters while other speakers are retracting almost all of their /str/-clusters. The third group is the most interesting, because they are the ones undergoing a change in their own lexical repertoire. This point will be further elaborated below. The finding of this analysis of lexical factors is that there are some speakers whose repertoire of sibilants currently presents exemplars on the entire spectrum from alveolar to post-alveolar production. For these speakers, lexical frequency may be a factor, while others make categorical production decisions.

In order to better show this pattern of lexical gradualness for speakers currently undergoing the change, I created a subsample from the cases pointed to above. These are speakers 3, 7, 10, 17, 20, 22, 23, 27, 28, 35, 40, 53, 55, 56, 61, 65, 71, 75, 76, 78, 84.

The resulting scatter plot in Figure 5.13 illustrates that the effect is much stronger for this group than the effect overall. Words with higher Lg10WF frequencies are more likely to be retracted with a difference of an entire z-score point.

Figure 5.13: Subsample speakers z-scores per Lg10WF



This brief analysis of a subsample verifies the assumption that lexical frequency is not generally irrelevant for the change in progress being researched here. Nonetheless, this language internal factor is limited in explaining the vast range of variability in sibilant production that is observed here.

5.2.2 Phonetic gradualness

In chapter 3 I introduced the complexity of lexical diffusion. The notion of phonetic gradualness as such is not one that was originally included in the concept of lexical diffusion, but rather was part of the conceptualization of the cognitive process put forth by Pierrehumbert and Bybee (Bybee 2002; Pierrehumbert 2002). Labov concluded that lexical diffusion is "the result of abrupt substitution of one phoneme for another in words that contain that phoneme" whereas he describes regular sound change as "the result of a gradual transformation of a single phonetic feature of a phoneme in a continuous phonetic space" (Labov 2010: 260). The argument of Bybee is not that lexical diffusion changes, such as the ones studied by Wang et al. and Labov, are never categorical, but rather that there is enough evidence to question whether phonetic gradualness is merely a characteristic of regular sound change.

Whether as a feature of lexical diffusion or regular sound change, phonetic gradualness is especially relevant in the case of /str/-retraction, because the place of articulation can be identified as a continuum. In order to establish phonetic gradualness, I have to critically evaluate what gradualness means phonetically and how I establish gradualness statistically, because these are two different notions.

For phonetic gradualness, the understanding is that a sound is produced somewhere between one state and another. In research reviewed by Bybee, examples of phonetic gradualness are both the deletion of final alveolar consonants as well as the deletion of intervocalic /d/ in Spanish. Phonemes in the first example show varying stages through

length of the sound. For the second example, Bybee analyses the different variable stages they discovered as "variants ranging from [d] to [th] to ø as a continuum" (Bybee 2002: 265).

There are two aspects of phonetic gradualness that I will apply to my analysis of /str/-retraction. On the one hand it is the above-mentioned notion of a continuum from the original version of the sound to the fully changed new version. Secondly, I include in my analysis the notion that this state of gradual change will be present to varying degrees in the varying words.

These analyses will be based on the CoG measurements of each individual because the categorization allows for an interspeaker comparison and fine detail per individual will be retained, which is somewhat smoothed over through standardization in z-score analyses. It is also interesting to see if the results differ in any way from the results with the continuous approach through z-scores. In a hypothetical example relevant to my analysis, a gradual production of the retracted alveolar fricative may be produced at a center of gravity of 6000 Hz when /s/ is generally pronounced at 4000Hz and /ʃ/ generally at 8000Hz. There are some problems with this approach. First of all, as I have established above, the pronunciation of sibilants, especially that of highly "combinable" alveolar /s/, is quite variable in and of itself. And while ranges for these pronunciations can be established, variability as high as 2000Hz per sound exists in the sample.

Secondly, it is difficult to establish gradualness statistically. Arguably, the mean and especially the median of production should lie between the alveolar and the post-alveolar sound if these were truly gradual. The problematic part here is that especially the mean is not representative of values lying in the middle but could instead be produced merely by extreme productions on both ends of the spectrum. This is not exactly the same for the median, but it is still going to be a number in the middle of the spectrum, no matter how far the spread is, at least as long as there is no skewness toward one side of the spectrum. For the specific spectrum of sibilants, this would mean that a gradual retractor would produce the same mean as someone who produces similar amounts of categorically retracted and categorically non-retracted values.

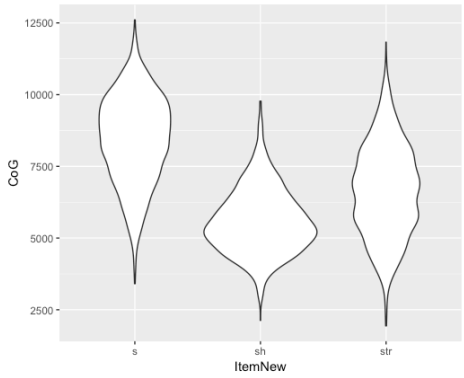
T-tests take both the mean and the dispersion of the data into consideration to show whether a categorical difference between two groups, in this case /str/ and /ʃ/, exists. They can thus confirm or deny a categorical difference. Nonetheless, they can only be used to make a statement on gradualness by combining all items and denying categorical difference, they cannot identify which of the items do not support the categorical choice.

While Rutter argues for categorical patterns based on the mean spectral peak measurements per item and per speaker, the findings in Baker et al.'s study as well as the findings in Phillips' study on timing suggest a gradual pattern (Baker, Archangeli & Mielke 2011; Rutter 2011; Phillips 2018). In contrast to all of these studies that work with medians, I will look at gradualness in my data differently.

This specific dataset has already been used to provide a first glimpse at gradualness above in chapter 5.4. The density plots visualized that productions span the entire range of the sibilant space, resulting in a curve twice as broad as the production of /ʃ/ and /s/.

The violin plot below further illustrates this dispersion. It illustrates that there is greater variability in /str/ than in /s/ and that the largest amount of data points lie between 3000 and 8000Hz.

Figure 5.14: *Item dispersion*



5.2.2.1 Quantifying phonetic gradualness

The method to further quantify gradualness rests on the concept of dispersion. When sibilants are considered a range of values in a particular z-score zone, the shape of this curve can be established by its interquartile range. This range is a measurement of the dispersion of data points. The bigger the difference between the 25th and the 75th percentile of the data, the larger the spread and vice versa. Resulting from this delimitation, the interquartile range of /s/ and /ʃ/ of each individual can be used to establish where on the spectrum of retraction they produce the alveolar sibilant in triple consonant clusters. A fully alveolar /s/ in these clusters must hypothetically lie somewhere between the 25th and 75th percentile of the production of /s/ in other environments. A fully retracted /s/ is expected to be acoustically indistinguishable from post-alveolar /ʃ/ and thus to occur within the interquartile range of /ʃ/. A production that lies between the 75th percentile of /ʃ/ (the sound with the lower CoG) and the 25th percentile of /s/ (the sound with the higher CoG) will then be considered evidence of gradualness.

Since the sociolinguistic collection methods of the recordings are prone to measurement errors, I use the interquartile range, even though using the 10th and 90th percentile as measurements of spread may slightly change the picture.

5.2.2.2 Methods to establish gradualness

In R, I calculated the interquartile range of all non-standardized CoGs for /s/ and /ʃ/ for each individual speaker and set them as the boundaries to rate each CoG in /str/-clusters as either *alveolar*, *gap* or *post-alveolar*. Since some values can possibly occur lower than the 25th percentile of /ʃ/ or higher than the 75th percentile of /s/, I rated the former as *very retracted* and the latter as *fronted*.

The above-mentioned variability of sibilants creates some problems for these judgments. In speech production, only those sibilants that are meaning-distinguishing in their specific environment are under high communicative pressure to behave as quantile theory would predict. Only large numbers of items can somewhat aid in circumscribing this variability. Nonetheless, there is the possibility of variability so spread out that the interquartile ranges of both sibilants overlap. In my analysis, this was the case for

participants 7,9,11,16,19,68, which led me to discard their 61 /str/-productions in the data set for this particular analysis.

5.2.2.3 Results Phonetic Gradualness

For the remaining 75 participants, more than half of the values fall in the above described *gap*. Some values are *fronted* or *very retracted*. This would indicate not only that there is an ongoing change creating a diverging production from the alveolar sibilant in the direction of a more retracted production, but it also shows that the larger portion of retraction is phonetically gradual. As is visible in Table 5.10 below, only 328 cases are produced alveolar or fronted. A larger number of items, namely 523 cases, count as retracted or very retracted. This proves that some items are certainly categorically retracted. Almost the same number of items are uttered somewhere between the individual ranges of /s/ and /ʃ/ (593).

Table 5.10: Sibilant Ratings

fronted	alveolar	gap	post-alveolar	very retracted
64	264	593	275	248

These results need to be understood as additional explanation rather than substitute for other methods, because they are based on the process of creating categorical measures from a continuous variable and as such may overestimate the existent pattern. Nonetheless, it is this type of binning that makes a distinction based on spread rather than means or medians valid. A comparison of medians or means, as was established in section 5.2.2, underestimates not only the importance of each individual's sibilant space, but it is also unfit to account for the spread of the data. In more pronounced terms, a /str/-median between the median of /s/ and the median of /ʃ/ cannot distinguish between someone who categorically switches between similar numbers of alveolar and retracted productions of the sibilants and someone who produces each /str/ in a rather centered position between their alveolar and post-alveolar production. This all is in line with Pierrhumberts "random variation over the exemplar cloud" (Pierrehumbert 2002: 115) that is used to produce varied outcomes of a phonetic pattern in language.

5.2.2.4 Discussion Phonetic Gradualness

I created a binning system above based on the natural dispersion of the data. This system aided in developing a first impression on phonetic gradualness in the change being observed. While it is flawed in the sense that it may mask how exactly the values are dispersed within the *gap* bin, it takes Bybee's theoretical assertions into account and finds a quantitative operationalization to truly measure these results.

I used the 5-way categorization to build multinomial mixed effects logistic regression models and see if any of the language internal or external factors could predict the patterns at hand. However, all attempts to model resulted in a lack of convergence. This may be due to the complexity of the word frequency factor, the oversimplification of the sibilant space as well as the comparatively small number of observations I ended up with in the analysis.

Nonetheless, all 593 values identified as *gap* values above are evidence that points to this change as a non-categorical change.

5.2.3 Conclusion Phonetic and Lexical Gradualness

The section above has illustrated to what extent mechanisms of sound change, namely lexical diffusion and gradual phonetic change, are at work in the present change. I found that lexical diffusion is best described by a combined measurement of cognitive exposure to a pattern in exemplar ways, which were operationalized through word frequency and content dependence. The approach to estimate factors in simple regression did not aid in describing lexical gradualness. The model with the best fit was created by implementing an interaction factor for birth year and word frequency as was pioneered by Hay and Foulkes (Hay & Foulkes 2016).

In the second part of this chapter on gradualness, I further developed a method to establish phonetic gradualness and found not only 593 values confirming gradualness is at work in the current status of the sound change, but further that many speakers produce sounds that are certainly retracted from the alveolar tongue position.

Some of the hypotheses tested in this study have been supported with evidence in these previous sections. I described in Chapter 3 that analyses of social factors for this change have thus far been inconclusive. The following section will explore this aspect in greater detail.

5.3 Social Factors

The analysis has thus far focused on those factors that are inherent in language. I looked at /str/-retraction with a focus on the acoustic structure and the impact of certain words. Although the comments mentioned in the introduction show that there seems to be some sensitivity to social context, research thus far is quite unsure about any social impact factors on the change other than indicators of change in apparent time. These indicators are usually parametrized through age, and the expected pattern is that younger speakers produce a more retracted version of /str/ than older ones. There is a plethora of reasons why this is the state of the art. For one, language internal factors may be at work in any of the varieties of English currently under investigation. The social set-up of every speech community will be more specific however. While the expectation is to see some similar patterns across the broad categories that can be employed to characterize a community, the indexicality in these communities will be impacted by the different social make-up of each community.

The fact that only gender has thus far been found to impact /str/-retraction is therefore not surprising in two respects. For one, the datasets often provide little to no ethnic diversity to begin with. Furthermore, in contrast to language internal phonological and lexical factors, social factors exist on an identity-related spectrum. There are notions of what it means to be a white American female that shape an identity across city, State and even national borders. These will be those aspects of identity that create patterns in datasets from various locations. There will also be notions of what it means to be an African American Austinite that will create patterns in my dataset, while they will possibly not converge with notions of what it means to be an African American man from Lafayette.

The argument I am making here is that my goal is to uncover any and all social factors that create patterns in my specific dataset. It is, however, important to acknowledge

that social patterns come about in fundamentally different ways than language internal patterns and may or may not be reproducible in other environments. They arise because of how humans shape their identity using language as a tool with almost infinite resources.

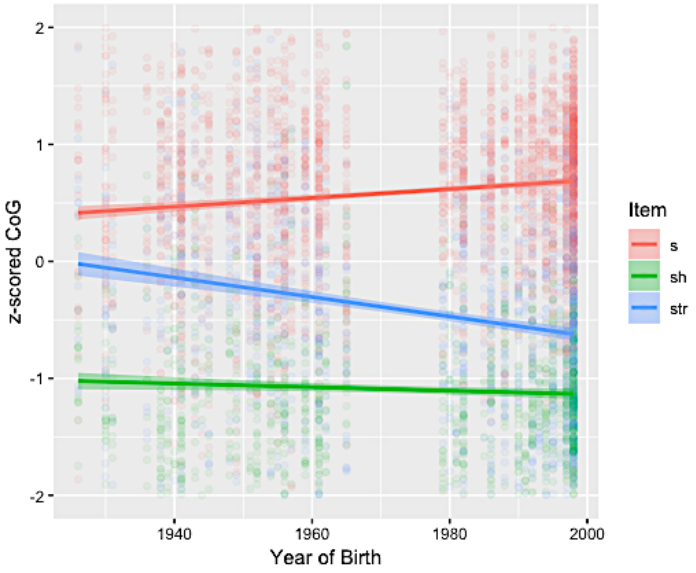
I analyze five social factors in particular here. The first factor is age, parametrized through the birth year of my participants. The second factor is activity-based speech style as was described methodologically in chapter four. Gender, the most-analyzed social factor in retraction research thus far, is then evaluated as a factor in the present data set. I show quickly that my above mentioned criticism of parametrizing social class indeed delivers no additional information and also attempt a first analysis of the social factor self-identified ethnicity.

I start with a descriptive approach to the dispersion of the data at hand. These descriptive statistics aid in understanding how these patterns work together when the final model of the sound change is presented and evaluated. This model is theory-driven and will include factors for all aspects hypothesized to have an impact above. The mixed effects regression model is then compared to a continuous tree model of /str/-retraction in order to better establish the importance of each and every factor for the model and how they may relate to one another. Finally, a random forest will aid in creating a bigger picture of the relationships identified through the tree model.

5.3.1 Age

To establish the age of participants, their year of birth was recorded. Due to the fact that all recordings were made in a timespan of five months, it is safe to expect that there will be no observable changes in real time and birth year can thus be the single time factor analyzed in this change. In contrast to the other social factors, which are tested on the /str/-subset of the data, age is first analyzed based on the entire dataset.

Figure 5.15: Items over time



The regression lines in the scatterplot Figure 5.15 above show two interesting findings: On the one hand, we see that blue dots are scattered across the entire z-score range while green and red dots remain mostly in the upper and lower part of the graph. This

results in a regression line for /str/ that is in between /s/ and /ʃ/ for the entire range of years. In the more recent birth years, there is a visible slope downwards toward /ʃ/. Furthermore, the graph indicates that /s/ is being fronted in apparent time. This is a finding that Wilbanks drew attention to earlier. He found that the sibilant space, which is the distance between /s/ and /ʃ/, is increasing in younger males in his study, while the space is greater for females from the earliest birthyears to the latest overall (Wilbanks 2017).

Without any further analysis, this is visual confirmation that the younger speakers in the dataset are more likely to retract /str/ than the older speakers, which is the strongest indicator for a change in apparent time. Figure 5.15 further underlines that /ʃ/ is stable and /str/ is moving in its direction. Whether the pattern observed for /str/ above is further influenced by any other social factors is analyzed based on the /str/-item subset below. All of the following analyses display the specific social factor across time to show not only whether the factor influences the z-score but also how it is meaningful as a factor in the sound change. Regression lines created in these plots are not representative of the complexity of the entire model, because they neither incorporate random effects nor the other factors. How much of an effect remains for these individual factors once all influences are incorporated can only be presented in the mixed effects regression model.

5.3.2 Style

In close connection to duration as a rather language internal measurement, style or activity-based retraction gauges aspects of ease and formality both in terms mechanical production and speaking style. While all other factors presented here account for interspeaker variation, this is variability that exists within the individual and is thus a unique addition to the current state of the art. The expectation is here that sibilant duration is closely connected to the type of speech activity being performed. Naming and storytelling are expected to be somewhat equal in speed although the cognitive processing of naming functions differently than the cognitive processing in free speech. Certain differences in speed as well as in cognitive processes are expected in the reading section. This is due to the fact that both formal norms of proper enunciation as well as individual fluency in this task are at play here. The less experienced a reader is, the more the fluency and duration of their rendition of written language is affected.

These differences create a hierarchical order of greater and lesser effort in the cognitive performance in speech production. This would result in the following pattern:

lowest expected retraction-rate → *highest expected retraction-rate*

reading - naming - free interview/ re-narration

formal → *informal*

The dispersion of items across activities is visible in Figures 5.16 and 5.17. Since the naming activity could more easily incorporate items with /str/ than the reading section without creating suspicion, it is not surprising that the largest number of items is found in this section of the interview. The pattern illustrated in Figure 5.18 is somewhat more surprising. The outcome for naming is in line with expectations, but the story interview items are higher in z-score than one might expect.

Figure 5.16: *Dispersion of /str/ (activities)*

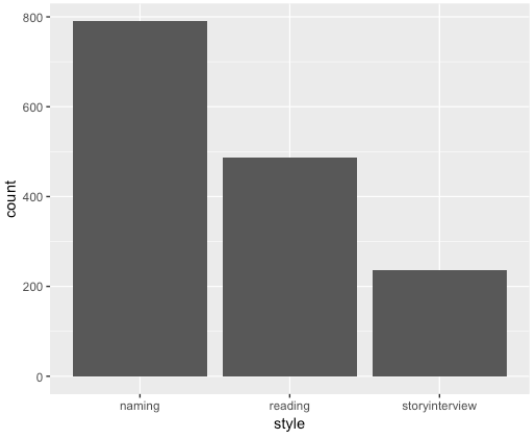
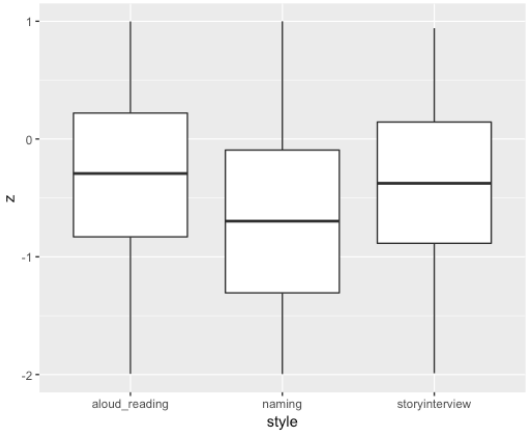


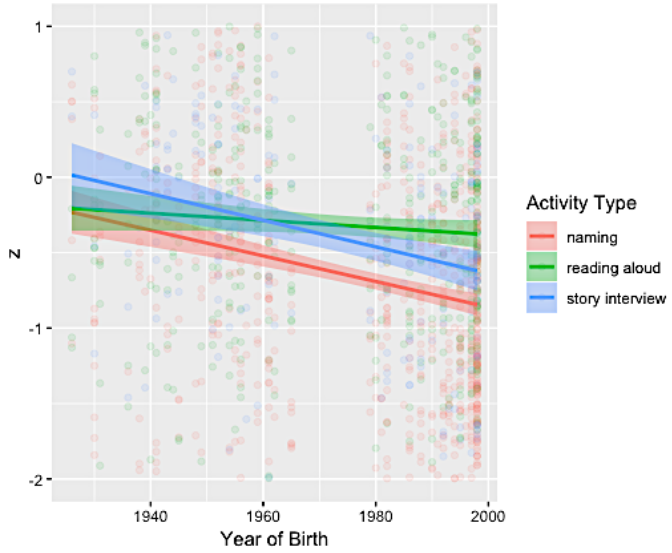
Figure 5.17: *Z-scored CoG per activity*



This pattern could be explained by the fact that the open-ended questions in this section yielded vastly different approaches to the narration task. While some speakers produced very formal recitations of the story, others merely debated the content. The reasons will be further elaborated on in Chapter 6.

Whether the type of speech activity performed has a changing effect in apparent time is visualized in Figure 5.18 below. The scatter plot with regression lines per style indicates that there is certainly some change in apparent time based on these activities.

Figure 5.18 *Activity-based z-score differences over time for (str)*



The scatterplot shows that sibilants in all activities are somewhat retracted in the younger speakers, but the slope of the regression line for both naming and story interview are steeper. Interestingly, the story interview starts out with the highest z-scores for all activities, which created the higher median visible in the box plot. Nonetheless, the slope indicates a pattern of increased retraction in younger speakers and thus a trend in apparent time.

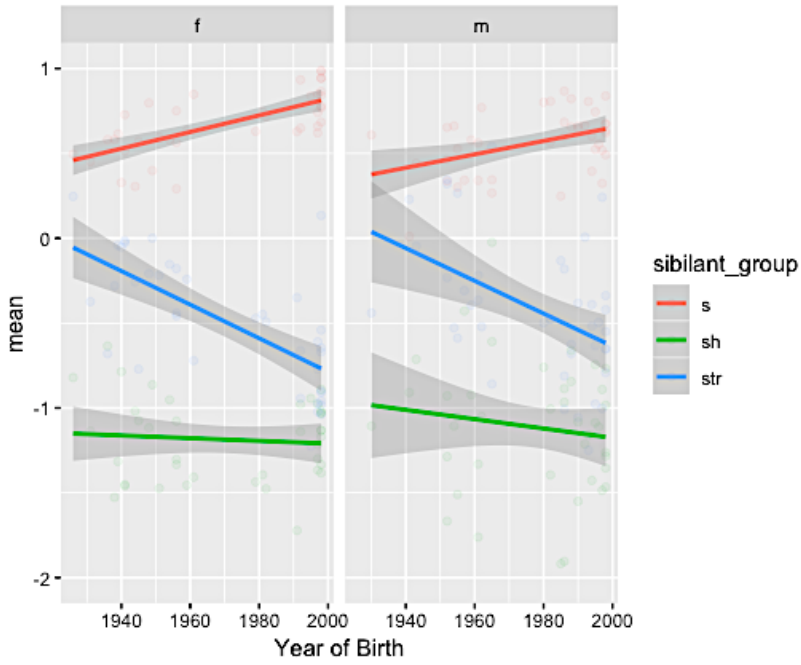
5.3.3 Gender

Gender, as introduced above, is an important variable to establish patterns in variationist sociolinguistics. In older research, female behavior was found to be so predictive in many

types of change that Labov concluded overall principles as described in chapter 1. They predict female behavior in changes from above and changes from below. However, studies on /str/-retraction are thus far unsure of the role of females in the change. Theoretically, a change from below is expected here and thus females are expected to lead the change. This pattern was found by Durian and Wilbanks, while Gylfadottir and Phillips cannot confirm it (Durian 2007; Gylfadottir 2015; Wilbanks 2017; Phillips 2018). Wilbanks looks at the change in /str/-production in initial and medial environments and concludes that females are changing, especially in the medial environment, while men are relatively stable.

This pattern is not reproducible in my data. In contrast to previous plots, I created mean sibilant measurements to provide a better comparison with Wilbanks' plot in Figure 5.19. Since I did not control for environment, it only displays gender.

Figure 5.19 *Gender-based sibilant differences over time (means)*



In contrast to Wilbanks' data, the biggest gender difference here is the variability within the males, indicated by the grey shades. The regression line displayed for males shows a larger standard error, which is a result of variability. A similar pattern to Wilbanks' data appears with females showing a slightly steeper slope toward retraction.

Figure 5.20 displays all values individually rather than per person means. It serves as further evidence that linearity exists in the change toward lowered z-scores for CoG measurements of the sibilant in /str/-clusters, but the change is slightly greater and downward sloping for females. Both figures do not deliver evidence for a female-led change, especially in comparison to Wilbanks' plot.

Figure 5.20: Gender-based z-score differences in (str) over time

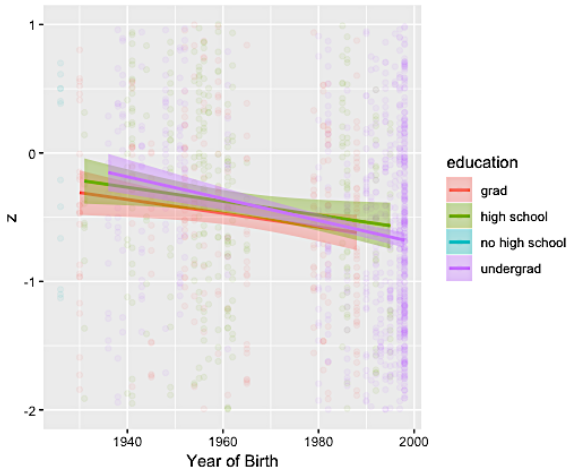
The transparent data points all around the regression lines visualize well that the data is dispersed across the entire range of z-scores from 1 to -2.

This is not in line with any of the expectations for sociolinguistic gender patterns, but confirms the lack of effect mentioned by both Gylfadottir and Phillips (Gylfadottir 2015; Phillips & Resnick 2019).

5.3.4 Education

I have described in great detail how and why quantifying any notions of social class in general and in the specific research setting for my study is difficult. The only measurement of social class included here was the level of education. Not surprisingly, the younger participants are overwhelmingly currently in or have completed undergraduate studies, with some participants throughout having only completed high school and some participants having completed a graduate degree. Such few participants did not complete high school that they do not appear as a regression line below.

Figure 5.21: Education-based z-score differences in (str) over time



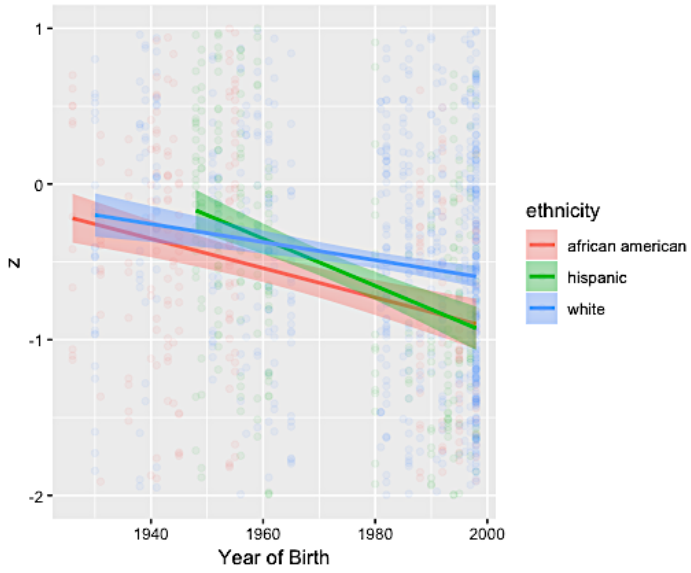
Similar to the findings on gender, the resulting three regression lines indicate that while there is a change toward retraction of the sibilant in /str/-clusters in apparent time, this change is not meaningfully connected to any of the displayed levels of education.

5.3.5 Ethnicity

So far, the pattern of different activities in response to retraction over time was the only figure revealing an effect for these differences in activities. All other figures showed overlapping and thus non-effective regression lines that predicted the change in retraction, but not based on the factor gender or education.

This is different for ethnicity as a self-identified characterization per participant. Figure 5.22 shows how differently retraction functions for Hispanic-Latinx and African American Participants as compared to Whites.

Figure 5.22: Ethnicity-based z-score differences in (str) over time



The figure indicates that there is an effect for all speakers over time. From the very oldest participants to the youngest, African Americans have lower z-scores and a steeper slope of change than white speakers. In contrast, older Hispanic-Latinx speakers, possibly due to bilingualism, appear more fronted than all other groups in the 1950s, but younger Hispanic-Latinx speakers are congruent with the patterns of African Americans in the most recent birth years. The points further show that many of the young participants are White and their values are dispersed once again across the entire z-score range being pictured here, while green and red dots are rare above 0 in the youngest generation.

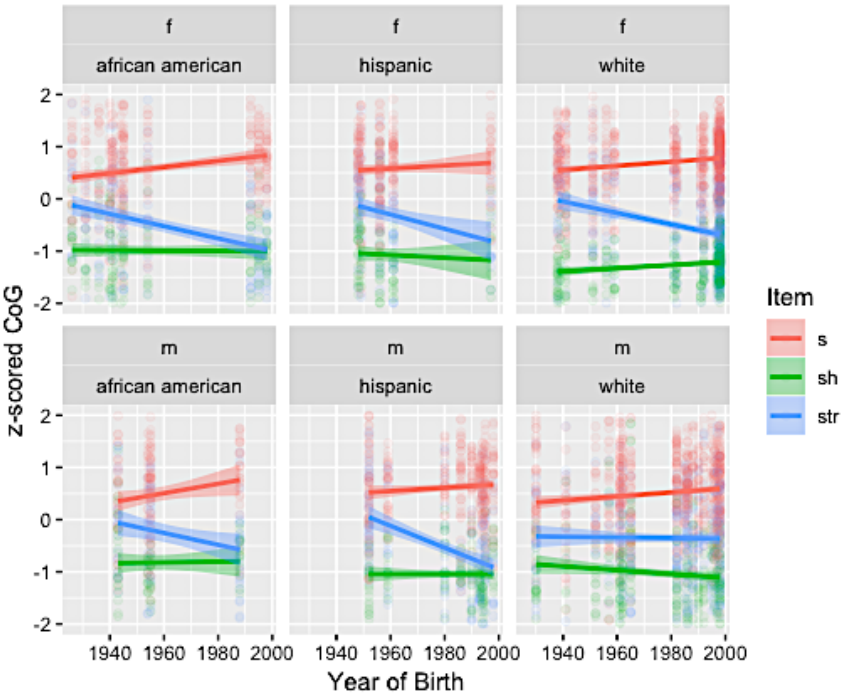
5.4 Social Factors Combined

Those factors that have been involved in pattern formation in the sections above can be divided into further meaningful subgroups, to show in greater detail how they affect the change and which parts of the newly created subgroups are affected the most. Figure 5.23 further illustrates not only the change in /str/-clusters, but also the change in /s/ as opposed to /ʃ/ that was introduced in section 5.3.1. This graph visualizes the relationships of three different social factors: age, gender and ethnicity.

For both the change in /str/ and the change in the sibilant space, the slopes for African Americans show the most extreme fronting and retracting. Hispanic-Latinx speakers change their /str/ drastically over time while the values for /s/ remain somewhat more stable. The patterns in White male speakers are the most stable, while White females

mirror those of African American speakers with less extreme slopes and less overlap between /str/ and /ʃ/.

Figure 5.23 Z-score differences in /str/ by gender and ethnicity over time



The extreme change in Hispanic-Latinxs may be explained by the number of speakers represented in the sample. It is important to keep in mind here that fewer than 20 of the 75 speakers are Hispanic-Latinx. A few extreme cases in the younger speakers are therefore likely to cause a skewness. However, the fact that these extreme cases exist certainly indicates that they participate in this change. To what extent this is applicable to all Hispanic-Latinx speakers needs further research.

This analysis may serve to draw the first conclusion that /s/-retraction in /str/-clusters is a phenomenon of free speech that currently gains ground in Austin. Another current change in sibilants in Austin is the general fronting of the alveolar sibilant. This second change seems to occur in all ethnic groups, but is most visible in African American speakers. Within the current dataset, I am unable to evaluate whether this fronting may be due to the effects of aging or is truly a change in apparent time.

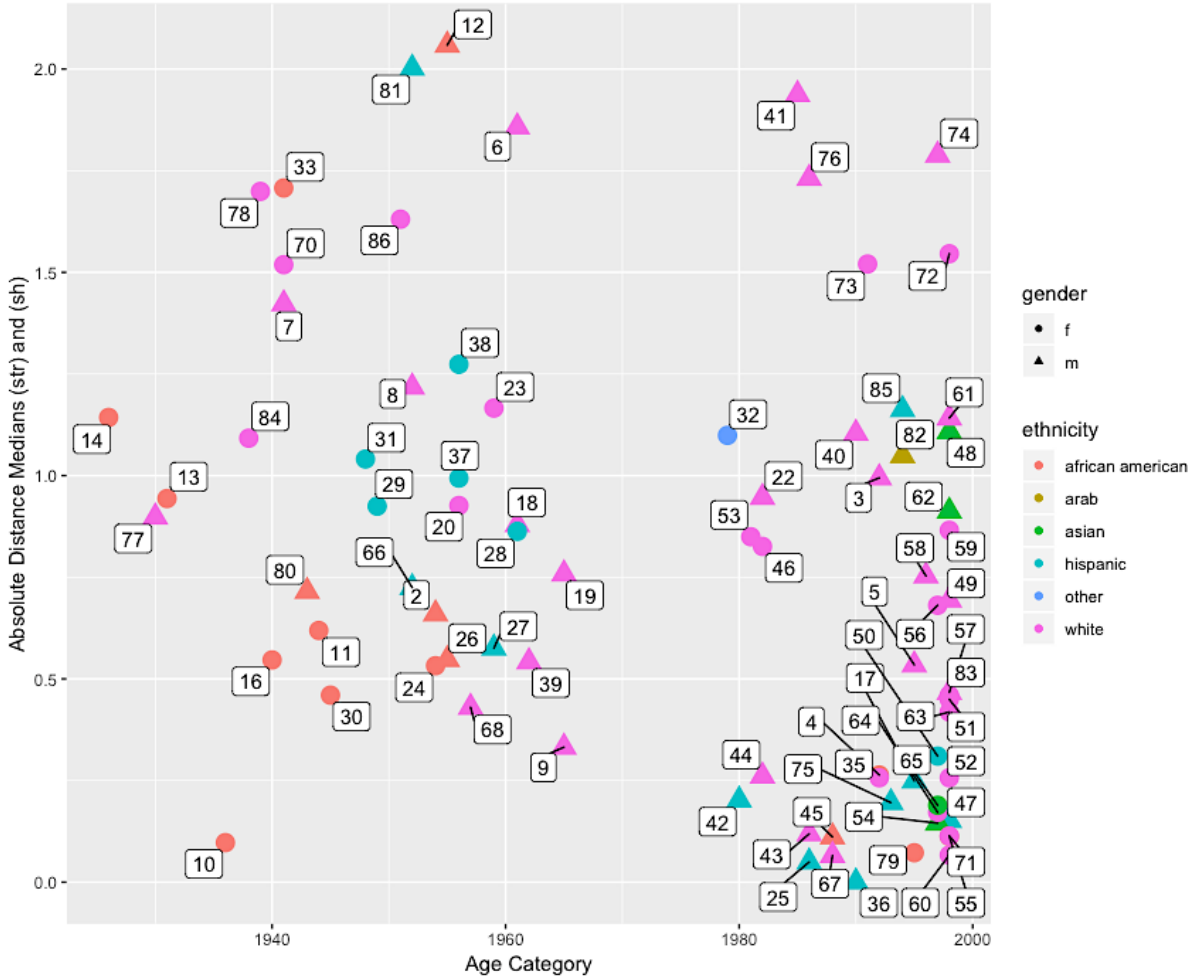
Although all ethnic groups participate in the change in /str/-clusters, there is some indication that the innovators may be African American females. For White speakers, females are certainly in the lead. Before evaluating the importance of these factors beyond this specific dataset and their strength in creating meaningful inference, I briefly introduce the patterns of individual speakers involved in this change. I have visualized in section 5.2.1.3.1 that individuals behave differently in this lexically gradual aspect. The next section will be less specific and just discuss their /str/-performance.

5.4.1 Individuals within the change toward retracted /str/

I have alluded to the importance of individuals within this change in various aspects of this results section. A more direct way to identify the differences between individuals is to slightly simplify the dependent variable by looking at the absolute difference between /str/

and the /ʃ/ medians. Medians are less affected by outliers and therefore preferred over means in this case. The assumption is that the smaller the distance, the more retracted the participants' production of /str/ will be. Across all items, the absolute distance calculated between z-scores of /s/ and /ʃ/ is 1.904. This can be thought of as a measurement of the general sibilant space within z-scores. Figure 5.25 displays the absolute difference between median SH/STR per participant. These differences are displayed over time, with point shapes indicating gender and color indicating ethnicity.

Figure 5.24: Median difference between /s/ and /ʃ/ per speaker



Several trends are illustrated in this plot. First of all, the greatest distance visible in this graph is slightly greater than 2. The upper values are mostly produced by males, especially in the younger age groups. Above a difference of 1, most values are produced by white and Hispanic-Latinx speakers, while the lowest values in older age groups are produced by African American speakers. The threshold for retraction based on my own auditory perception of speaker's utterances appears to be at 0.5. Values around and below this threshold rarely appear before the birthyear 1960. Speakers 9, 19 and 68 are exceptional, because they have very small overall sibilant spaces, possibly indicating some flawed measurements. The two remaining older speakers with retraction are all female African Americans. Speaker 10 is especially remarkable here.

The 46 younger participants are even more spread out in terms of their absolute distance. The change seems to be in the area below 0.5 as described before. Instead of the

3 previously described speakers, there are now 25 retractors in the group, including 12 male speakers. They span all prominent ethnicities, although only one Hispanic-Latinx speaker appears below the threshold.

5.5 Combining language internal and external factors

In the chapters above, I explored both the extent to which the change in /str/-clusters is phonetically and lexically gradual as well as the social factors that may influence this change. In the following, I cannot incorporate measures of phonetic gradualness, because it would force me to change the dependent variable. However, the types of models used to make predictions in the following allow for a combination of both language internal and external predictors to fully evaluate any estimates gleaned from my data.

5.5.1 Regression Modeling

In line with the methodological elaborations by Tagliamonte & Baayen, I start with the mixed effects regression model that best describes the data at hand and then further evaluate the nature and relationships between any effects in a conditional tree and random forest analysis. The final theoretical exploration will consider the differences between gradual and non-gradual individuals guided by a conditional tree (Tagliamonte & Baayen 2012).

The data generally provided the opportunity to establish several language internal and external factors in the retraction of /s/ in /str/-clusters. They have been presented in the previous sections and amount to the following. The social survey and activity types were designed to establish the birth year, gender, education level, self-identified ethnicity and activity style as language external factors. The acoustic analysis further measured the duration of each sibilant.

There is ongoing debate about the pathway to and nature of the best version of a mixed effects model. Barr et al. postulate that the maximal model permitted within the hypothetical constraints is always the best choice (Barr et al. 2014). This idea was recently challenged by Bates et al., arguing "the advice to "keep it maximal" often creates hopelessly over-specified random effects because the number of correlation parameters to estimate rises quickly with the dimension of the random-effects vectors" (Bates, Kliegl, et al. 2015). In order to take both opinions into account, I fit the maximal model first. This first model included all of the independent factors that represent quantifiable assumptions in the hypotheses based on /str/. As such, I included year of birth (H3), style (H8), gender (H7), ethnicity (H9), lg10WF (H4) and Duration(H6) with participant (H5) and word context (H4) as random effects. In section 5.2 on word frequency, I explained the significance of an interaction with birth year for this factor, that was also included in the model. In that same section, I also explored the great impact of individual differences, which is why I included a random slope for word frequency per participant in the model.

In line with Bates et al.'s expectation, this maximal model failed to converge. Table 5.11 outlines the resulting final model that excludes the random slope. Both the word frequency and duration measurements were logarithmically transformed (log10) and standardized to allow for better comparability.

Table 5.11 Final Model

Model Summary

	Dependent variable: z
Birth year	-0.010 (0.002)
	t = -5.692***
Word frequency	3.524 (1.816)
	t = 1.941*
Ethnicity [Hispanic][11]	0.041 (0.114)
	t = 0.356
Ethnicity [White]	0.213 (0.099)
	t = 2.143**
Gender [m]	0.121 (0.073)
	t = 1.665*
Duration	0.099 (0.025)
	t = 4.018***
Style [reading]	0.069 (0.110)
	t = 0.630
Style [storyinterview]	0.101 (0.112)
	t = 0.904
Birth year: WF	-0.002 (0.001)
	t = -1.997**
Constant	18.171 (3.279)
	t = 5.542***
Observations	1,342
Log Likelihood	-1,601.070
Akaike Inf. Crit.	3,228.141
Bayesian Inf. Crit.	3,295.765
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.11 displays all estimates of the model. Several factors are identified as significant in this model with regards to their level-based deviance from the reference estimate per category. Ethnicity (White) is significantly different (namely higher) than the intercept for African American. Duration is also significant with increasing duration

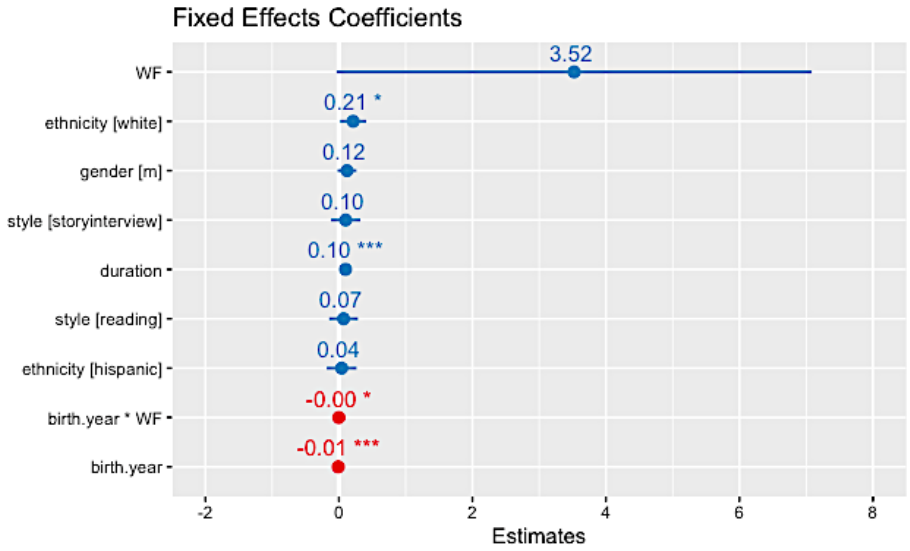
[11] The reference categories for these categorical predictors are Ethnicity [African American], Gender [f] and Style[naming]. The coefficients thus display the difference between these reference categories and the respective category.

increasing the estimate. It further shows a significant effect for the interaction of word frequency and birth year.

This model successfully establishes which factors influence the pattern when word context and individual differences are accounted for. Table 5.11 further shows that the t-values differ greatly when we compare the effects of birth year and ethnicity, meaning that in this model, the pattern is well predicted by age and just secondly by the factors duration and ethnicity.

In Figure 5.25, fixed effects coefficients are presented through the points while the standard error is presented through the lines.

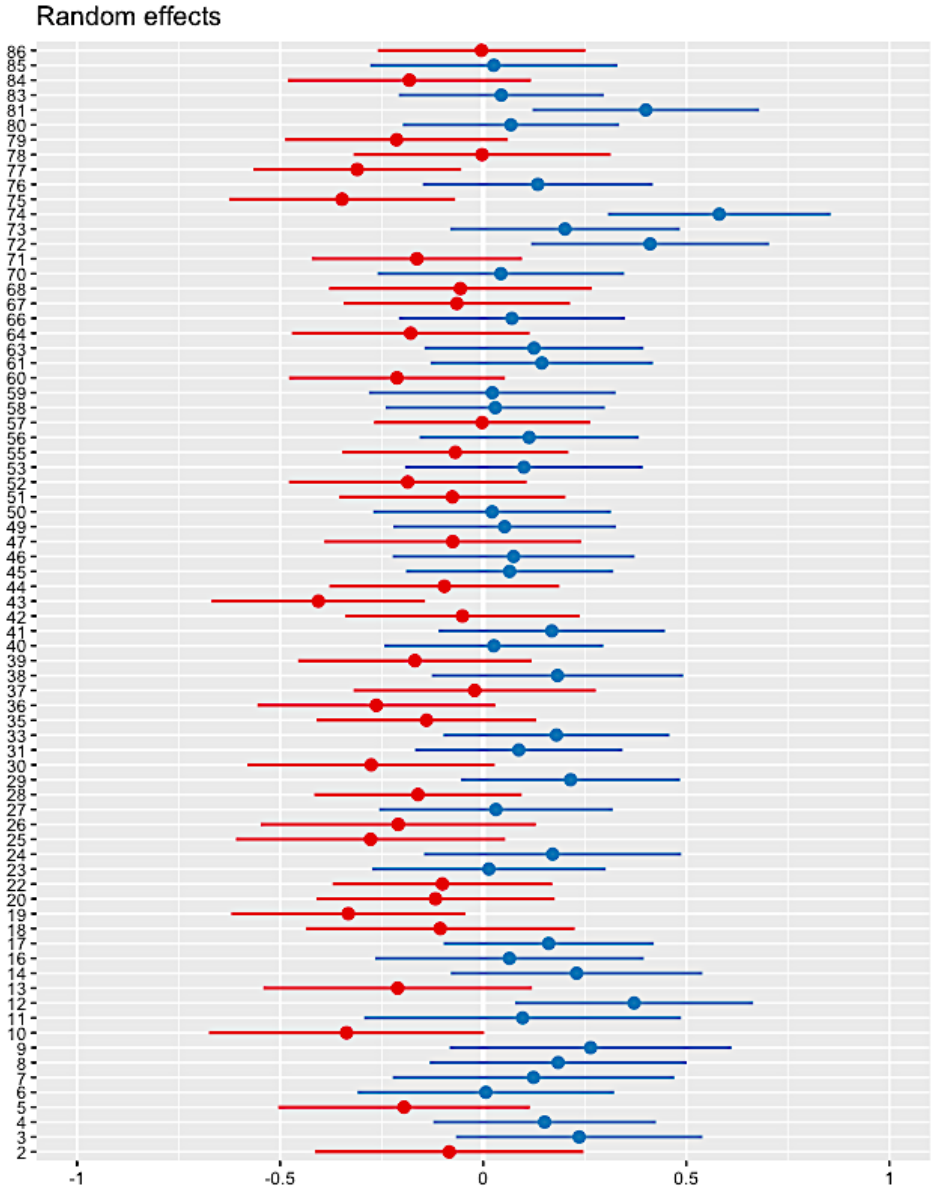
Figure 5.25: Fixed Effects Coefficients



The figure shows that duration creates the greatest movement within the regression line, while also displaying the greatest standard error. Higher birth years and word frequency lower the intercept, while Hispanic and White ethnicity and all activities other than naming increase the intercept. Only the above-named factors have a significant effect however.

The variability in intercepts modeled in the random effect structure is portrayed in Figure 5.26 and 5.27. Figure 5.26 depicts which speakers deviate from the estimate negatively (red) and positively(blue).

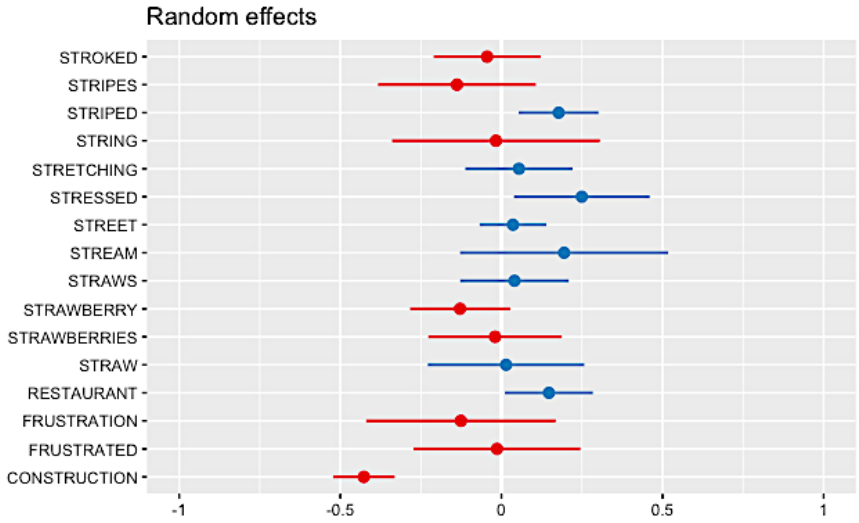
Figure 5.26: Random effects per participant



The most extreme retractors identified through this plot are participants 10, 19, 43, 75 and 79. Overall, 35 of the participants negatively deviate from the overall model intercept. However, the spread of the confidence interval is indicative of great variance in the data and thus imprecision of the model.

The random effect for word is graphically represented in Figure 5.27.

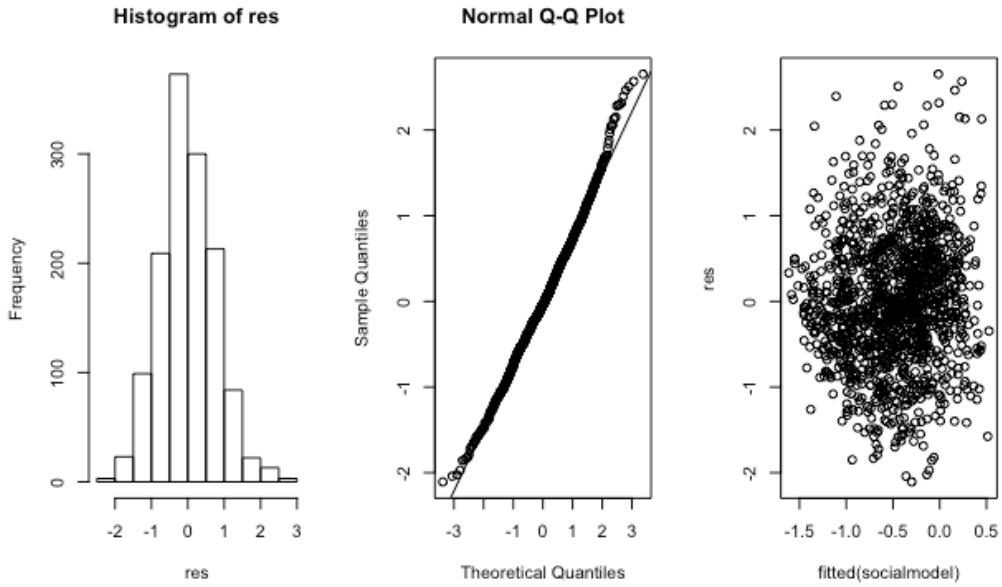
Figure 5.27: Random effects per word



The words that negatively influence the intercept of the model as identified in Figure 5.28 are: STROKED, STRIPES, STRING, STRAWBERRY, STRAWBESSIER, FRUSTRATION, FRUSTRATED and CONSTRUCTION. The overlap of the confidence intervals with the overall intercept of many of these words again points to variance in the dataset.

To visually confirm the assumptions of this model by ensuring normality and homoscedasticity, I plotted the residuals as presented in Figure 5.26.

Figure 5.28: Model diagnostics as histogram, qq-plot and residual plot



The first two plots in Figure 5.27 show that the residuals are dispersed normally around the mean. The third plot confirms homoscedasticity because no pattern is visible throughout the entire range of values.

5.5.2 Random tree

The effects of factors on a given variable may also be estimated by conditional trees and forests. Instead of predicting the impact of a factor on a given outcome by assuming all

factors to have an effect on the slope of a pattern, this type of analysis uses the factors to split the data into meaningful binary groups in an iterative process. Every group is then once again scrutinized for natural grouping based on all the same factors until tests for independence of the dependent variable and a factor yield significant results. Through this method, the most significant predictors are not only established first, but it also provides an opportunity to identify such factors that are only relevant to a subset of the data. In the present case, it would be impossible to distinguish, e.g., whether there is a subgroup in the older speakers in which gender does have a significant effect, because the pattern present in this subgroup is not significant when compared to patterns in the larger dataset that affect more speakers.

A further benefit of conditional inference trees as pointed out by Tagliamonte and Baayen is that they "implement safeguards ensuring that the selection of relevant effects (predictors, variables) is not biased in favor of those with many levels (multiple factors in a factor group), or biased in favor of numeric predictors (e.g., age of the individuals)" (Tagliamonte & Baayen 2012: 159).

Since the visualization of the effects is rather complex, I describe all nodes in the table below and then visualize them with a tree model without edge labels. The tree was created with the following formula in R using the *party* package (Hothorn et al. 2017):

```
Austin.ctree <- ctree(z ~ stylefactor + ethnicityfactor + birth.year + genderfactor + Duration +
SubtlexLG10WF + WordContextFactor + ParticipantFactor, data=final_eth_subset)
```

Gender and birth year do not figure in the tree model, meaning they do not provide a succesful category for classification of any of the subgroups. This is somewhat surprising knowing that age does model retraction well in the regression model. However, the effect for participant may step in to gauge the variability that is otherwise described through birth year here, because the model lacks the ability to differentiate between these types of effects.

Focusing on the node categories and numbers in Table 5.12, the importance of random effects is further very visible. Generally, they are the first and largest grouping factors in this tree. The participant effect splits the participants into two groups, meaning that the left group produces significantly different z-scores from the right group. Within this left group, there is a split between all other words and the word CONSTRUCTION. Comparing the two resulting terminal nodes, CONSTRUCTION shows a mean z-score of -0.723, which is much lower than the mean z-score for all other words at 0.088.

The right group of speakers split by node 1 is further split by word context. In contrast to the left group, the split here differentiates between CONSTRUCTION, STRAWBERRY & STRING and all other words. These three words have a very low mean of -1.079.

All other words in this group of participants are further described by additional factors. The first split of this group is Lg10WF. Words with a frequency above 2.61 produce a terminal node with a mean of -0.715, indicating once again a retracted production. The 7th node subdivides the less frequent words used by participants in the right edge group into ethnic groups. African Americans and Hispanic-Latinxs produce z-scores with a mean of -0.552 in the resulting terminal node. The last split in the White group is based on style,

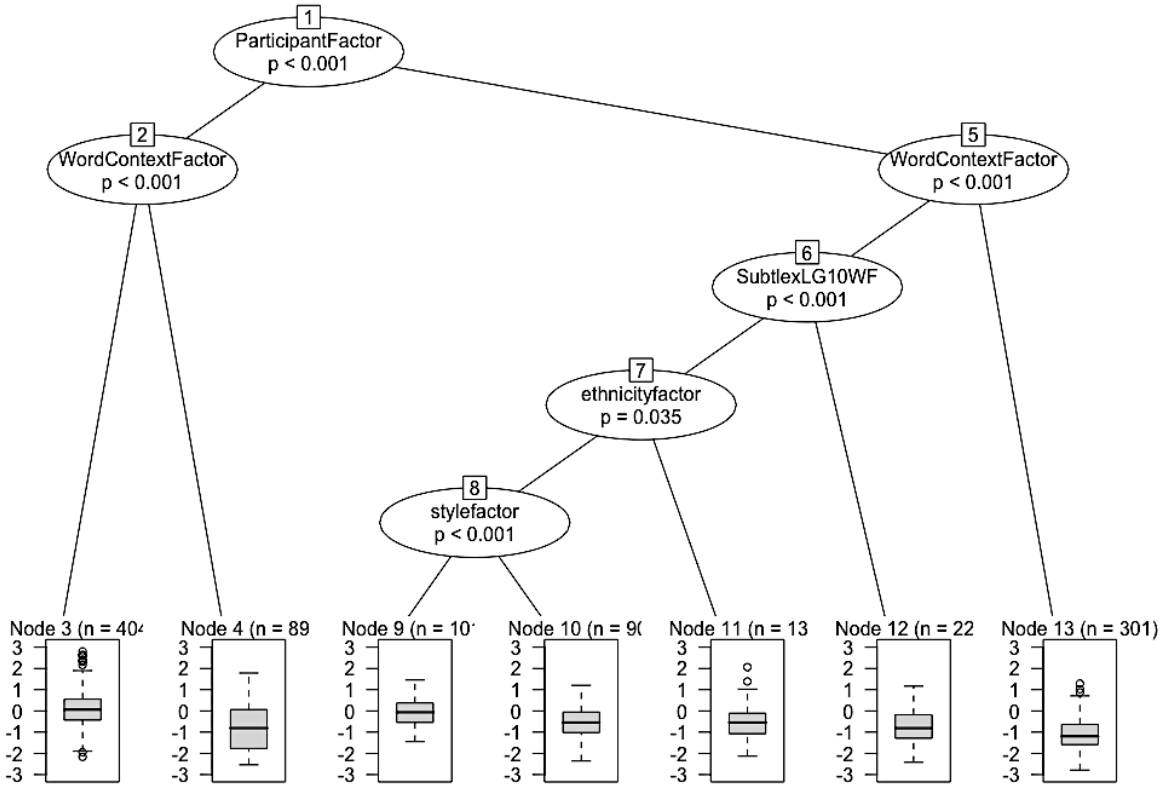
where reading produces a less retracted z-score of an average -0.087, while naming and storytelling are more retracted at -0.522.

Table 5.12 Nodes in Tree

Node Number	Node category	Left Edge	Right Edge
1	Participant	3, 6, 7, 8, 9, 11, 12, 14, 16, 18, 23, 24, 29, 31, 33, 38, 40, 41, 66, 68, 70, 72, 73, 74, 76, 78, 80, 81, 84, 86	2, 4, 5, 10, 13, 17, 19, 20, 22, 25, 26, 27, 28, 30, 35, 36, 37, 39, 42, 43, 44, 45, 46, 47, 49, 50, 51, 52, 53, 55, 56, 57, 58, 59, 60, 61, 63, 64, 67, 71, 75, 77, 79, 83, 85
2	Word Context	FRUSTRATED, FRUSTRATION, RESTAURANT, STRAW, STRAWBERRIES, STRAWBERRY, STRAWS, STREAM, STREET, STRESSED, STRETCHING, STRING, STRIPED, STRIPES, STROKED	CONSTRUCTION
3	Terminal Node, y=0.088	N=404	
4	Terminal Node, y=-0.723		N=89
5	Word Context	FRUSTRATED, FRUSTRATION, RESTAURANT, STRAW, STRAWBERRIES, STRAWS, STREAM, STREET, STRESSED, STRETCHING, STRIPED, STRIPES, STROKED	CONSTRUCTION, STRAWBERRY, STRING
13	Terminal Node, y=-1.079		N= 301
6	Subtlex WF	<2,61	>2,61
12	Terminal Node, y=-0.715		N= 224
7	Ethnicity	White	African American, Hispanic-Latinx
11	Terminal Node, y=-0.552		N=133
8	Style	Naming, storyinterview	Reading
9	Terminal Node, y=-0.087		N=101
10	Terminal Node, y=-0.522	N=90	0.94

Table 5.12 is displayed in Figure 5.29 as a simplified conditional tree. Additional to the visualization of the data in table 5.12, this tree display provides boxplots for each terminal node including further information about the variance in each group.

Figure 5.29: Conditional inference tree of all factors from regression analysis



Variance is greatest in Node 4. Node 4 and Node 13 show the lowest medians for z-scores. The highest z-scores and thus most alveolar sibilant productions occur in Node 3 and 9, namely participants with generally lower retraction rates (apart from the word construction) and white speakers producing less frequent words in the reading task.

5.5.3 Random Forest

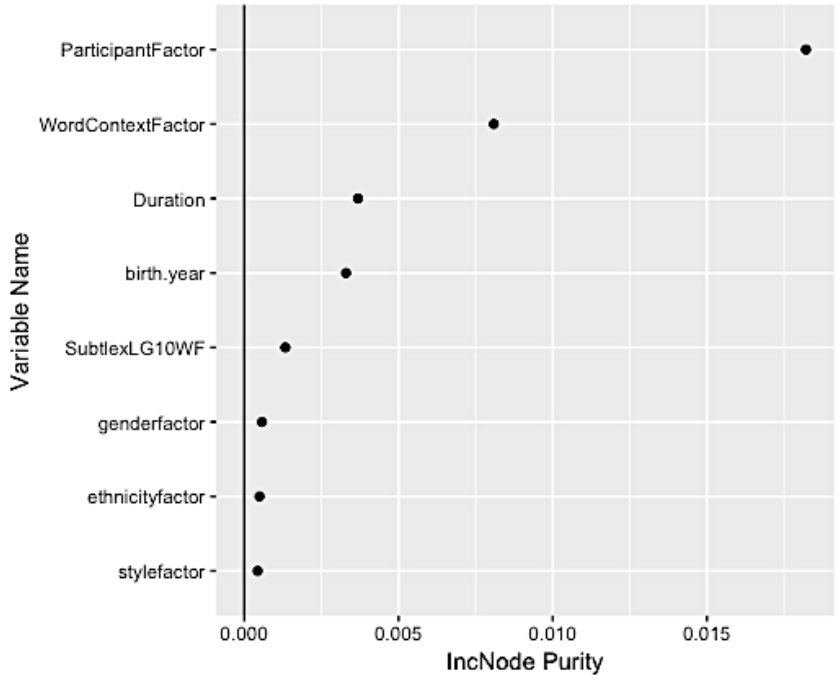
The random forest function I use is also based on findings outlined in Tagliamonte & Baayen (Tagliamonte & Baayen 2012). I rely on the *party* package in R, because it works with the conditional variable importance measure. In contrast to assumptions in other R packages that produce random forests, the underlying structure here is better suited to work with correlated variables and to differentiate better between correlated variables such as the specific word context and measurements of word frequency. Furthermore, the data explored is never the entire dataset, but rather subsets of the dataset that are used as a test set and then compared to the data not used in the test set. The forest of trees created in this way provides a better overview of how strong certain patterns are in the data (Tagliamonte & Baayen 2012). This is especially valuable when compared to the procedure in regression models, where it is more difficult to establish whether a pattern arises merely because some speakers use it predominantly or whether it is a gradual pattern visible in a greater part of the data at hand.

The expectation in this analysis would therefore be that the patterns predicted in the regression model are somewhat similar to the patterns found in the random tree and random forest analysis. In contrast to the regression model however, these analyses directly

compare fixed and random effects and further allow to evaluate whether any factors affect the data more generally than others. For example, an expectation could be that word frequency or duration have a more general effect across all ethnicities, while ethnicity as a factor may not be as strongly predictive, because it applies only to a smaller subset of the data.

The results presented in the Figure 5.30 below support all of the findings and claims previously made in this analysis. /str/-retraction in consonant clusters is overall a phenomenon of individual variation, which means that individual language patterns best predict the pronunciation of the sibilant in this cluster. Further important factors in the outcome of z are birth year and word context. The importance of birth year and duration were both not represented in the random tree above, but do align with the regression model. On the lower end of the nodes, Lg10WF is also represented as a meaningful factor. Lastly, ethnicity and style have an impact on the production of /str/. In contrast to both the regression and the random tree analysis, gender is a significant predictor in the random forest analysis.

Figure 5.30: Random forrest variable importance



5.6 Summary of Statistical Analysis- Internal and External Factors combined

Both the descriptive statistics carried out in sections 3.1-3.3 and the statistical inference in sections 3.4 and 3.5 show that /str/-retraction is a multi-faceted process that cannot be described or modeled as a simple pattern of coarticulation or social change. Only the combination of language external and language internal factors is able to predict the lowering of z-scores in the /str/-environment. Although the importance of factors depends upon the type of statistical analysis, we can generally see that the external factors speech activity, ethnicity, gender, age and individual influence the production of the sibilant.

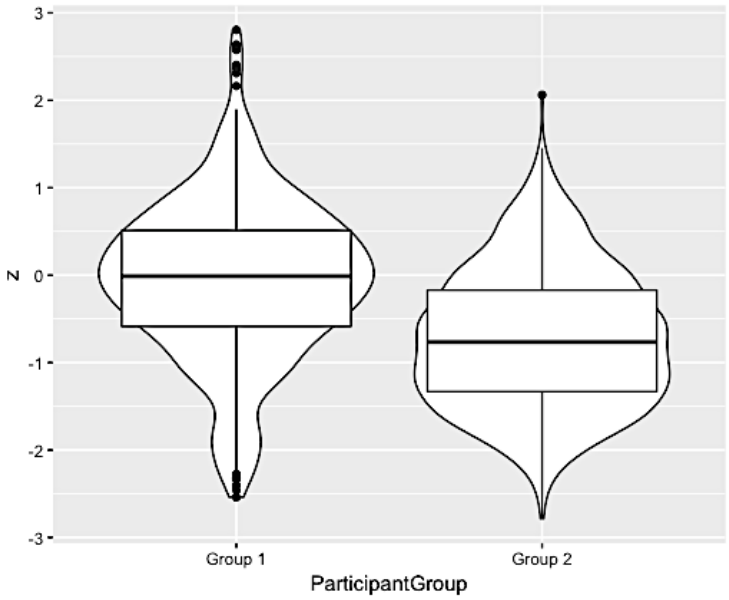
Language internal factors that further describe this variation are duration, word and word frequency.

5.7 Grouping Factors in Interspeaker Variation

The conditional tree above indicates a split in the dataset where certain individuals behave differently than others. An analysis of only those speakers presenting with the variable pattern in question is another additional analysis aspect suggested by Tagliamonte and Baayen (Tagliamonte & Baayen 2012). In the following I explore this group of individuals. Note that these are not entirely aligned with the group of speakers identified above, who are the ones showing patterns of word frequency differentiation. The tentative analysis of the model presented above was that these are the speakers that show patterns of retraction beyond coarticulation.

A first indicator of this could be the overall mean production of /str/ amongst the two groups. To accomplish this, I recoded the data in R and calculated the median as visualized in the boxplot below.

Figure 5.31: Dispersion of /str/ across groups



Indeed, Group 2 is easily characterized by the lower mean z score, which is -0.012 for Group 1 and -0.765 for Group 2. The dispersion visualized in Figure 5.29 above further exemplifies this.

Interestingly, none of the other social factors further classify this group visually. They seem equally divided in terms of ethnicity, style and gender. The complexity of this division is outlined in Figures 5.32 and 5.33 below. While there seems to be a visible trend for Group 2 in terms of z-score lowering, no such trend exists for Group 1. Furthermore, we see an age split in Group 1. There are some African American speakers with higher overall z-scores in Group 1, but only before the birth year 1960. There are neither young Hispanic-Latinx nor African American speakers in Group 1.

Figure 5.32: Z-scored /str/ differentiated by group over time

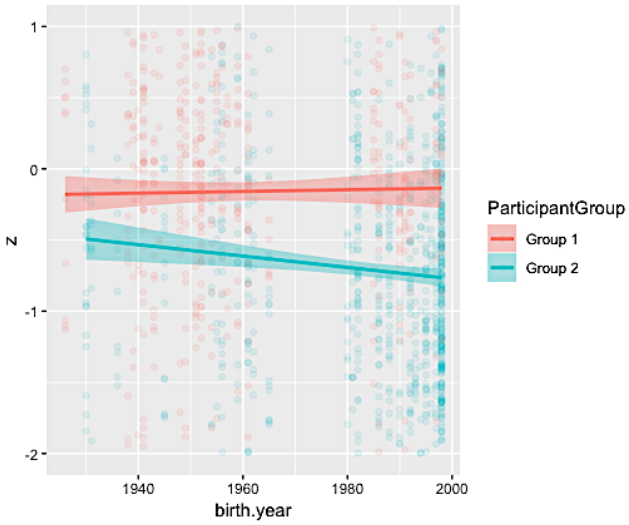
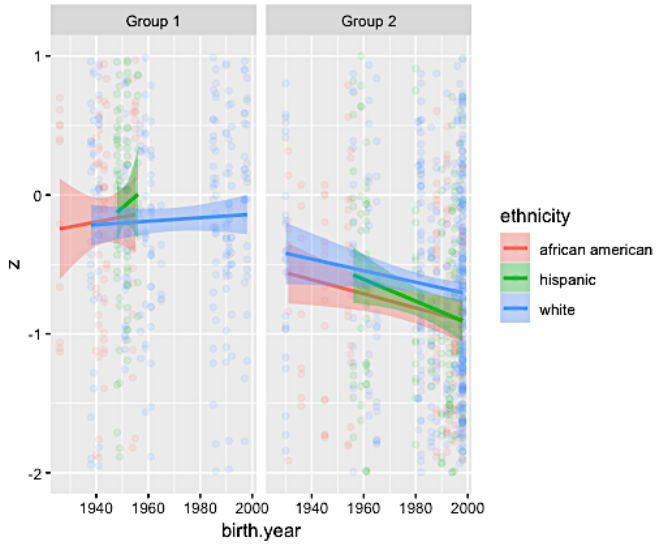


Figure 5.33: z-scored (str) differentiated by group and ethnicity over time



In a second step, I calculated the same mixed effects model on this subset of the data in order to see if any of the factors differently predict the z-score than in the overall model. If this is truly the group of innovators and adopters then all patterns from the models above should be stronger for this particular group.

Considering the fairly straight trend in Figure 5.32 above, it is no surprise that none of the predictors reach significance in the overall model when only Group 1 is included. Group 2 however is almost identical in its effect structure to the overall model.

Table 5.13: *Mixed Effects Regression Group 2*

Mixed Effects Regression Group 2

	Z ~
	year+ Lg10WF+ ethnicity+ Duration + style + (1|Participant) + (1|WordContext)
year	-0.005 (0.002) t = -2.769***
Ethnicity (Hispanic-Latinx)	0.101 (0.098) t = 1.023
Ethnicity (White)	0.226 (0.087) t = 2.588***
Duration	1.034 (0.575) t = 1.800*
Style (reading)	0.217 (0.107) t = 2.015**
Style (story interview)	0.135 (0.114) t = 1.191
(Lg10WF)	-0.162 (0.078) t = -2.090**
Constant	8.562 (3.292) t = 2.601***
Observations	849
Log Likelihood	-936.628
Akaike Inf. Crit.	1,895.257
Bayesian Inf. Crit.	1,947.441
Note:	*p<0.1; **p<0.05; ***p<0.01

The results show that higher birth years lower the z-score. They further indicate that White speakers are 0.226 higher in their estimated z-score than African American speakers, which reaches significance at p<0.01. Reading further increases the z-score estimate compared to naming by 0.217 at a significance of p<0.01. The last significant predictor is high word frequency which lowers the z-score estimate -0.162 at p<0.05.

Overall, the patterns strongly resemble the patterns described above. However, social factors are now slightly more significant than language internal factors.

A binomial model using group membership as dependent variable to further establish these patterns failed to converge, which is possibly due to the fact that a pattern is only present in group 2 while group 1 is mostly identified by not fitting the same pattern and thus great variance.

5.8 Comments on the Perception of /str/

Chapter 4 described how I elicited meta-commentary in the debriefing section of the interview. In the following, I use thematic analysis to summarize all /str/-related notions present in these comments. Although not experimental in nature, this qualitative insight will enhance the following discussion and help make sense of many of the patterns outlined above. I am aiming at a good depiction of the indexical field without underestimating the involvement of myself as a linguist biased toward certain theoretical expectations.

Based on the state of the art in retraction research, my interests in the interview were to understand how participants responded to the phenomenon. I wanted to gauge popular opinions of the phenomenon, see if people could acoustically detect it and learn if there was awareness of any social factors involved. I had however no sense of what the outcome

would be and had no influence on what else they might wish to talk about during the interview. This is to say some topics could be asked for through direct or indirect questions, but other themes were initiated by the participants. In the same interview, I could for example ask if a person could hear a difference between *street* and *shtreet* but had no influence on any statements of linguistic prescriptivism or orthography related justifications that might follow.

Due to the nature of the interview, the output created in these open-ended interviews is very heterogeneous and therefore needs to be analyzed in a careful and concise manner. There are many established techniques of qualitative analysis that were mainly developed in the field of psychology. Some of them are driven by an underlying theory such as discourse analysis or grounded theory while others are methods open to various theoretical and epistemological endeavors (Braun & Clarke 2006).

I chose to use thematic analysis to better understand the plethora of opinions and ideas put forth by my participants in the open interview. Braun and Clarke describe this qualitative analysis tool as "a method for identifying, analysing, and reporting patterns (themes) within data. It minimally organises and describes your data set in (rich) detail. However, it also often goes further than this, and interprets various aspects of the research topic" (Braun & Clarke 2006: 6).

Thematic analysis in the format used here is a process of detailed analysis and active structuring, which Braun and Clarke divide into 6 phases (Braun & Clarke 2006). I familiarized myself with everything participants said and transcribed the entire section of the interview, which Braun and Clarke consider Phase 1. I then identified all sections of the texts that relate to comments on /str/-retraction. Most of these codes are descriptive noun phrases or parts of sentences, some are adjectives. This is Phase 2 of thematic analysis. Phase 3 concerns the search for themes, which means I considered "how different codes may combine to form an overarching theme" (Braun & Clarke 2006: 19). Phases 4-6 involve reviewing the overarching themes, naming themes and reporting them (Braun & Clarke 2006). Which themes I identified and examples of utterances that are coded as part of these themes is described in detail below.

For a successful thematic analysis, Braun & Clarke underline the importance of making the unpreventable influence of theoretical positions and values transparent by discussing how they shape the analysis. A theme as such is defined as a topic that "captures something important about the data in relation to the research question, and represents some level of *patterned* response of meaning within the data set" (Braun & Clarke 2006: 10).

The thematic analysis I present below is inductive in the sense that its overall aim is an overview of the topics discussed in the interviews by organizing topics into overall themes of interest. However, in contrast to psychological studies, any part of a sociolinguistic interview will also be deductive in the sense that there are certain preconceptions about how participants respond when asked about language and in this particular case also what the social profile of the change could be.

Braun and Clarke caution against simply using thematic analysis as a way to group all information. According to their descriptions, a bad thematic analysis is one in which the

themes are congruent with the questions asked and there is no deeper analysis or attempt to structure the themes that follows (Braun & Clarke 2006).

Since I clearly provoked some answers through questions, I stress that the themes that I present below were developed further and represent not only the question content but the heterogeneity in responses I was confronted with. The coding was carried out in the software program MaxQDA. There are several themes present in many of the comments that I will discuss in the following. Some arise out of the questions asked, others randomly came up in the interview. Thematic analysis will allow me to pick up on any patterns related to these themes without overinterpreting the inherent mechanisms or the generalizability to cognitive theories of the speakers.

Concretely, many participants discussed their familiarity with the pattern, overwhelmingly as something they can detect when asked but do not notice in others. Awareness also includes self-judgment on the part of the participant. I divided between *auditory perception* and *awareness*, because speakers often stressed that these two are different. Many also expressed a *personal opinion*. The verbalization of opinions then often encouraged *folklinguistic explanations*. Some participants evaluated the *social characteristics* of the change and analyzed factors that may be at play. They were also asked to self-assess, many with surprising results. The content of these thematic groups will be described in the following.

Some participants were excluded from the analysis, because one could not be certain what they were discussing was actually /str/-retraction. One such participant kept referring to how people from England speak differently (P27), another participant used the opportunity to discuss the flaws of society and the education system with me (P80). Rather than making improper inferences, I excluded these metalinguistic interviews from the analysis. I was left with 66 metalinguistic interviews to code. Figure 5.32 below shows the different themes and the related codes. Codes, subthemes and themes will be presented in *italics* in the following sections. Subthemes are further categorizations of the codes shown in the figure. The quotes used to illustrate aspects of the themes are transcriptions of the utterances. In order to present these in the most natural format, I did not include any punctuation.

Figure 5.34: Themes and sub-themes in the metalinguistic interviews

5.8.1 Auditory Perception

Auditory perception was often the very first topic I discussed with participants, right after I explained the phenomenon in detail. If my rendition of retracted and un-retracted items did not yield an immediate reaction, I often asked "Can you hear the difference?". The first code I therefore used was *detection*, with a further sub-code for *replication/explanation*. 55 participants confirmed that they could hear a difference between STREET and SHTREET. 11 of them even confirmed their perception by repeating the sounds (P2, P4, P62, P67, P79) or describing the process in their own words: "yes the S[12] or S H" (P29). Other participants also made sense of the phenomenon by arguing on the basis of orthography (P39, P44, P81). One participant described his *auditory perception* as "one sounds deeper than second one" (P71).

Seven participants stressed their perception of the phenomenon as *subtle*, being surprised that this was a recognizable difference: "that's so subtle oh my gosh" (P4). Two participants acted with *surprise* after they heard the examples. Only one participant admitted to have *detection issues*.

Positive *detection* was often also confirmed *awareness* of the phenomenon. Therefore, many more participants confirmed the detection without differentiating between the *auditory perception* and an *awareness* of the phenomenon. I will describe the theme *awareness* in the following.

[12] Capital letters indicate an alphabetical description of the orthographic letters.

5.8.2 Awareness

The theme *awareness* was often a rather spontaneous topic at the beginning of the debriefing section of the interview. Those participants that did not talk about whether they could hear the difference often jumped right into debating whether they noticed the retraction pattern before and where they may have noticed it. I treat *never heard* and *never noticed* as one code here, because the difference between the two is more a lack of self-confidence to make claims about the phenomenon than a semantic difference. Participants claiming to have never heard it are simply making a stronger statement about their familiarity with the phenomenon and their self-perception as a judge of phonetic phenomena.

With 105 coded sequences of 335 total codes, *awareness* is the most prominent theme in the corpus. 38 coded instances were descriptions of some form of having *never heard or noticed* retraction in their life.

Within this theme, *recognition* of the phenomenon amounts to 27 appearances of the code in the dataset, from 22 different participants. This code includes outbursts such as "yeah my daughter in law says things like that" (P18) indicating a direct connection to a specific person. However, most speakers make general statements such as: "I think it definitely happens around here because I can hear that the second form a lot more" (P60). This type of statement is often the answer to a direct question and not a freely uttered confirmation. Of the 22 participants, 17 offer this type of confirmation, while the others more strongly connect it to their own reality like P18.

Within the theme *recognition* one of the important and rather varied codes is that of *self-judgment*. Many participants made it a point to consider whether they themselves were retracting the sibilant. Of these coded instances, only 14 could be confirmed as *correct*, with P2, P82, P56, P57, P65, P32, P67 and P26 identifying as retractors and P53, P60, P62, P18, P71 and P25 identifying as not retracting. Some of these retractors are aware of the process but indicate that others had an influence on this awareness: "and I say shtraw[13] not straw […] people notice it in me" (P67). Another participant mentions that he has been aware of this difference as a regional pronunciation: "Like I remember growing up with shtreet but I was wondering I used to think like why is it pronounced like that but that's just where I grew up I don't know if it's also pronounced that way elsewhere" (P82).

Other participants make incorrect claims about their pattern, but this only occurs in 4 interviews. P18 first utters an incorrect judgment of his pattern and then corrects himself later while recognizing that he creates the retraction when speaking faster. P52 and P66 claim they are not retracting when they are in fact very strong retractors, even in the same utterance. On the other hand, P49 is overcorrecting, thinking he is a retractor in informal speech when no retraction was present in his interview: "I'd like to think I'd say it kind of more pronounced street but I know probably when I'm talking I don't think about it probably say the less pronounced version" (P49).

[13] Lower case spelling with h indicates a retracted production in this section. I chose to use orthographic identification in this section to improve the reading experience.

The majority of participants however describe their own pronunciation variant as *unknown*. Within this sub-code, many speakers ask what it is that they are doing (P6, P19, P28) or utter that they simply don't know (P42, P44). P4 elaborates on this lack of knowledge and states: "I don't even know like it wouldn't surprise me if I did both the other one seems like a more lazy tongue pattern so I imagine in a more casual setting I would be like I'd probably be a little less firm on how I pronounced that str sound that sounds like what it is like just a mild deformation of that". Another speaker states: "I think in realistically I might say shtreet more than shtreet but I'm not sure I feel like I say shtreet more but I feel like I said sh_street but I did not I have to think about it more to say street instead of what's this shtreet I don't know if that's the same I don't know if I'm saying the same thing" (P79). The acoustic production in this entire sequence, as indicated orthographically, is retracted. Even when trying to produce the alveolar sibilant, P79 is retracting to a point that indicates a categorical retraction pattern in her speech.

The description by P4 as a "mild deformation" further illustrates the connection between the theme *awareness* and the major theme *opinion*, which will be described in the following.

5.8.3 Opinion

After participants had talked about their *awareness* and *auditory perception* of the phenomenon, many of them included or added notions of what they thought about this variant. Again, some of these were freely uttered statements, while some were the result of me asking what they thought of the phenomenon and whether they had a preference. The codes included in this theme can be thought of as scalar from a positive perception of the phenomenon on the one end of the spectrum and a fully negative perception of the retracted version on the other side. In between, people uttered versions of a preference for the alveolar and an acceptance of the phenomenon. The majority of codes revolve around the acceptable reading of the retracted variant (17) with the second greatest number of codes at the opposite end of the spectrum, confirming a negative perception (10). There are relatively few positive evaluations of the phenomenon (4).

P65 confirms the *acceptable* status of both versions by stating: "I'm sure I have heard someone who uses the S at the front of their mouth and on the back I haven't noticed a difference and so I don't think one is better than the other". Many speakers address this acceptability along the same lines, saying that both versions are appropriate to them.

The confirmations of a *preference for alveolar* realizations are often similar in stating that both versions are functional, but these participants perceive a clear hierarchy, such as P49 who argues: "I guess it still counts or whatever the right way to say that is it's just I would respect street street more". More strongly voiced preferences were coded as negative perception, such as P49 stating "I'd like to think I'd say it kind of more pronounced street" or P52, who simply utters: "that's weird". Some participants even say that they are "scared" to say it the subjectively wrong way by not producing the alveolar sound (P74).

One participant however also utters a negative perception of the alveolar production, arguing: "probably the one with the sh sound because like shtrong when I try to say the s it just sounds weird" (P55). Similarly, few participants offer a positive evaluation of retraction. They mention that the retracted production "just kind of flows out

a little more natural so I guess that would be a preference is that I feel like I can do that subconsciously" (P59). Nonetheless, while it is generally found to be acceptable, the trend on the spectrum described above leans somewhat toward the negative side. Possible reasons for this emerge both within the theme *explanations* and *social characteristics* that are described in the following.

5.8.4 Explanations

In the course of the meta interview, discussions of auditory perception and awareness were often entwined in or followed by the search for an *explanation* of the process at hand. Many of these explanations connect to perceived norms of Standard English, relating to *prescriptive explanations* of how it should be (8). Some participants compare the pronunciation variant to an *accent*, trying to say that while being an acceptable pronunciation it is not the expected one (11). Another prominent code present in utterances relates to *attention paid to speech* (10), which is sometimes defined further as uses of *casual speech*. Closely related to attention, many participants voice notions of the alveolar pronunciation being more *effortful* and pronounced (9). A last attempt in explaining the variant is often to resort to defining it as *individual pronunciation problems* (6).

In the *accent*-related explanation mode, participants utter "they've got a dialect or something" (P77). Two participants also say that someone with a German accent would pronounce it this way (P58, P84).

Attention paid to speech is addressed in many different ways. P14 provides a lengthy explanation of how they understand how retraction comes about when one is not careful in their speech: "well if you're not if you're not thinking about what people are saying if you're really not looking just like you this something that you're doing you know you wouldn't paying much attention". Another participant explains: "I guess would attach it more to consciousness of just being aware of what you're saying not really like you purposely are avoiding saying it properly but more just like you don't think about how it sounds you just kind of say it and you have no reason to consciously think about it cuz people understand what you're saying so you're not going to change it unless it's pointed out as an issue you know" (P35). Both participants stress the unconsciousness of the process and the lack of attention to this feature. P35 also notes the importance of feedback when one produces the retracted variant. Since the retracted pronunciation is not an "issue", people are less inclined to produce the alveolar pronunciation. Other participants connect notions of attention to their own pattern and explain their perceived improper pronunciation: "my mind is a thousand places so I wasn't really trying" (P42). Some also say they don't recognize the phenomenon because it would require more attention (P9, P54).

These notions of paying attention both to detection and production of retracted sibilants are closely connected to the notions of a divide between the effortful and well enunciated alveolar in comparison to the less effortful retraction pattern, coded as *effort in enunciation*. Utterances included in this code are often similar to P56's statement: "just like takes more effort or something so when you're like speaking fast you just say shtreet". Another participant described retraction as: "like a more lazy tongue pattern so I imagine in a more casual setting I would be like I'd probably be a little less firm on how I pronunced

that str sound" (P4). One participant connects the lack of care eminent in the city life style to lack of effort in pronunciation: "maybe I think it's because how fast like the city life would be they might have something else in mind and maybe they get tongue tied to a point where the pronunciation comes out to sound like another you know another pronunciation" (P85).

A further sub-theme of *explanations* is *casual speech*, within which 5 participants identity and define retraction as a phenomenon of *casual speech*. When identifying as a retractor, P56 states that she retracts the sibilant "when I'm like in casual conversation". P60 describes retraction as stratified by speech style: "I probably would just be saying that if I wasn't like reading something".

Although it is mentioned by only one of the participants, another sub-theme uttered in the realms of explanations is the involvement of /r/, when P85 states: "like they roll off the r".

The last explanation that participants come up with, especially those who are unfamiliar with retraction and have a negative perception of it, is to say that someone using this pattern has *individual pronunciation problems*. Two participants (P3, P66) worry that a former lisp may have impacted their performance: "my brother and I both had to get to speech therapy for s's so I've always had a hard time with s's and and r's actually" (P3). Participants 6, 9 and 77 are worried that they have a problem even detecting retraction due to their hearing. Participant 13 says that a stutter may force them to produce the retracted variant.

Prescriptive notions of pronunciation and orthography's importance in it are another comparatively common sub-theme included under *explanations,* namely the *prescriptive explanation*. P33 argues "it should be pronounced correctly - straight". Three participants say that the alveolar sounds "more proper" (P39, P57 and P58). P58 further argues orthographically: "cause when you say it the other way say shtreet it sounds like you're adding an H in there and it's not there". Similarily, P82 states "well street seems more like reasonable because it is spelled like that". Two other explanations directly address taught pronunciation: "it just sounds like different from what I would say or what I was like taught to say I guess" P72 and "maybe they've not been corrected or maybe it's just something they picked up somewhere or maybe this they haven't learned to pronounce it correctly" (P81). The importance of education in pronunciation alludes to *social characteristics* of the pronunciation variant, which will be described in the following.

5.8.5 Social Characteristics

The theme *social characteristics* includes many subthemes that are not necessarily addressed by a number of speakers. This is due to the fact that questions concerning any social profile of /s/-retraction were very open-ended compared to other questions, but in contrast to *opinions* or *explanations*, it is also the theme that was the least likely to be addressed by participants at all. *Age* was mentioned by three participants. One of them said that "old people" (P75) were retractors while two other speakers connected retraction to adolescence (P65, P81).

Participant 72 is able to identify retractors in her family and concludes: "my aunt says like the second version like shtreet but she like she lives in Minnesota too but like my

mom doesn't do that so I some reason I feel like it's random". This utterance was coded as *individual*, because the participant alludes to the individual and seemingly random differences amongst her own family. Participant 66 mentions the sub-theme *education/income*, when he states: "I mean depends on when I was growing up in East Austin people wouldn't say shtreet it was shtreet yeah they would just say that I think it would be I think I would hear it more of where of people with a to me people with not as much education or a lower income would say it". Much like participant 72, he is the only one to describe this social aspect of the change. Another participant describes the pattern at hand as an *urban* phenomenon, when she states "I kinda imagine when people use that sh sound in like heavy Urban cities like New York" (P85). This is the same participant who argued that the fast lifestyle in the city encourages being "tongue tied". Although more participants have something to say about the *regional* profile of /s/-retraction, there is no conclusive picture emerging from these accounts. Participants say they have noticed in in speakers from Pittsburgh (P35), Canada (P51), US (P4), British (53), Chicago (P54), not Michigan (P67), Connecticut (P74), Michigan (P76) and the South (P50, P82 & 83). Only one participant identifies it as a *Texan phenomenon* (P51), while four others state that it does not exist in Texas (P14, P18, P29 & P54).

While most people see no *ethnic group* as characteristically using this feature, one participant, who identifies as African American, recalls: "where I grew up at and that's what I'm talking about the neighborhood the people I'm basically around all the time and a lot of my friends the crowd of people that I'm running with basically speak the pretty much the same you know what I mean" (P2). While he didn't directly say that he grew up in an African American neighborhood in Austin, he seemed to imply this meaning in the interview. Another participant mentions that Hispanic-Latinx and African American people retracted /s/ when he grew up in East Austin, but he is not hearing as much of them anymore due to gentrification: "I don't hear it as much as I used to I did because neighbors kind of gentrified and some of them a lot of the neighbors here are not the same" (P66).

This freest section of the metalinguistic interviews is by far the least mentioned and conclusive theme discussed in all interviews.

5.8.6 Analysis

The meta interviews have overall indicated that /str/-retraction, while being an acceptable pronunciation variant to many speakers, shows many characteristics of a change below the level of consciousness. Participants agree that they can hear it, but less than half say they have recognized it before. This recognition has to be understood carefully, because it was specifically asked for. As indicated in section 5.8.2, only 5 participants connect this recognition to their own reality. Even though many participants confirmed recognition, I conclude that less than a tenth of the participants have consciously recognized the pattern in their own life and within their speech community.

A widely discussed aspect is linguistic correctness. In contrast to other themes discussed by participants, this aspect has to be understood through the lens of the interview process. Both the observer's paradox and the perceived hierarchy between the interviewer as expert linguist and the interviewee as layman promote corrections and worries about doing it improperly. It further encourages participants to use available linguistic knowledge

(orthography, pronunciation) to create an image of competence. From this point of view, every participant uttering negative perceptions and worries can also be seen as a participant who is truly unfamiliar with the process and does not consciously encounter it in peers. The four participants that do offer a positive evaluation are therefore all the more interesting, not only because they offer this positive opinion but rather because it is based on usage and naturalness.

Although the acoustic analysis has indicated that the dataset includes many retractors, only two participants confirm that they have been corrected on their pronunciation. Furthermore, only 8 participants correctly identify as retractors.

The combined results throughout the themes *auditory perception*, *awareness* and *opinion* thus indicate little conscious awareness of the variable. Further signs of this lack of conscious awareness are visible in the lack of a social profile of the variable. This not only shows a change below the level of consciousness however. It also implies a low level of social salience and indexicality.

Concretely, this lack is visible in the fact that even among self-identified retractors, there is no agreement on a social profile of /str/-retraction. Neither the region nor any other social aspect delivers frequent or conclusive results. As a consequence, there is no notion of an ongoing change. While two participants did say they can see adolescents using this pronunciation variant, what they meant was adolescents are generally careless enough to use it, not that they had increasingly noticed it in younger speakers. Furthermore, only one participant, the one who grew up in East Austin and connected the pronunciation to these speakers from the past, mentions any type of social stratification.

In a socially salient variable, we would expect outbursts such as “This is what African Americans do all the time!” or “More and more young people are doing this”. And even in first order indexicality, for which the social profile is not salient to the speech community itself, the expectation would be that speakers who recently moved to Austin or, in line with a change in apparent time, older speakers would recognize differences. That this is not the case is surprising considering the findings in the acoustic analysis.

Although Resnik & Phillips saw some indication that the triple consonant clusters /spr/ and /skr/ may index masculine toughness, neither that nor any related notion of social salience came up in these metalinguistic interviews (Phillips & Resnick 2019).

6. Discussion

The mixed methods approach to /s/-retraction in /str/ clusters paints a varied picture of how the change progresses through the speech community. In the following I discuss the implications of the results presented thus far by addressing hypotheses put forth in previous research and the hypotheses I investigated in this particular project. To recoup, these hypotheses concern both phonological and sociolinguistic aspects. I then connect all findings and provide a broader understanding of how and why they influence each other. The hypotheses underlying the current research project were:

Phonological Aspects:

1.) /s/ is produced with /ʃ/ -like qualities in the triple consonant cluster /str/ (Labov 1984; Shapiro 1995; Lawrence 2000; Labov 2001; Durian 2007; Baker, Archangeli & Mielke 2011; Rutter 2011; Gylfadottir 2015; Stevens & Harrington 2016; Wilbanks 2017).
2.) /s/ is most retracted in /str/ as a result of coarticulation (Baker, Archangeli & Mielke 2011; Stevens & Harrington 2016; Phillips 2018)
3.) The production of /str/ is changing toward a more retracted pronunciation in apparent time.
4.) The change will present as phonetically and lexically gradual, with frequent words significantly affecting Centre of Gravity measurements (Rutter 2011; Wilbanks 2017).
5.) Some individuals have become retractors with full phonologization of the post-alveolar sibilant in the /str/-environment (Rutter 2011).
6.) Sibilants with longer duration are more likely to be retracted.

Sociolinguistic Aspects and Variation Patterns

7.) Females are not leading the change (Gylfadottir 2015).

The hypotheses that none of the previous studies have sufficiently addressed before were:

8.) Activity-based speech style will show differing levels of retraction.
9.) Ethnic identity will have an effect on retraction such that White speakers are less likely to retract.
10.) For retractors and non-retractors alike, the change in progress shows little social evaluation.

6.1. Phonological Aspects

The founding observation of the present thesis was that a growing number of English speakers in various English-speaking countries produce a version of the sibilant in e.g. *Streetlife* that is much closer to their canonical post-alveolar sibilant production. Both the analysis of the overall dispersion of /str/ values as well as the comparison of means and the overall difference between the mean for /ʃ/ and the mean for /str/ have confirmed this observation and thus H1.

Chapter 2 on phonological concepts has shown not only how complex it is to pinpoint where this change happens in terms of the understanding of what a phoneme is, but also indicated the intricacies of this variability in multiple ways. For one, the specific phonotactics of English do allow greater variability in the sibilant in the specific cluster, because the post-alveolar does not appear in consonant clusters and is thus not parsed as a theoretical possibility in the respective position. This is to say an alternation between the two does not violate any meaning-bearing differential expectations in terms of a minimal pair. Furthermore, the specific acoustic profile of sibilants allows for analytic methods that successfully differentiate between post-alveolar and alveolar sibilants through CoG

measurements. Yet, the nature of complex and aperiodic sounds makes this acoustic distinction a gradual one that shows great variability.

The comparison of various ways to model the sibilant groups in section 51. delivered further evidence that direct assimilation to affricated /t/ models the patterns very poorly. While this approach is not as fine-grained as the temporal measurements employed in the phonetic analyses by Stevens and Harrington, it serves as additional evidence for H2 (Stevens & Harrington 2016).

Further theoretical issues in developing an understanding of the change in /str/ include the problematic concept of gradualness and the distinction between phonetic and lexical gradualness. Within the analysis on gradualness for this specific dataset, the meaningfulness of the apparent time concept for grasping lexical gradualness became evident through the interaction term. The close connection between word frequency and birth year shows that frequency is a complex measure in language change, confirming this aspect of H4. Rather than predicting variability, which it would do as a non-interacting fixed effect, lexical frequency only becomes meaningful when it models a change in time (through the interaction). This is to say, for a change in apparent time, frequency effects are highly dependent on the stage of the ongoing change and where within the change continuum specific speakers are located.

Phonetic gradualness was explored by binning all per-speaker according to their individual sibilant dispersion. The great number of items that fall between the 3rd quartile of /ʃ/ and the 1st quartile of /s/ per participant suggest that the production of /str/ is most often not a categorical choice but rather phonetically gradual with a larger number of items leaning toward retraction, which confirms H4.

Both the groups observed in the conditional tree as well as the analysis of median differences per speaker in Figure 5.25 have shown that some speakers show a very clear pattern in producing retracted forms. Since Rutter himself judges production instead of using the acoustic measures employed in this study, it is much easier for him to confirm phonologization. Nonetheless, the small differences between these measures found here, as well as anecdotal evidence and some of the comments in the debriefing section (“Am I saying shtreet or shtreet?” P59), are strong indicators that some speakers are nearing phonologization of the pattern (H5).

Both total duration in the variable importance of the forest model as well as the significance of the effect of log. duration in the regression model indicate a language internal effect that relates to articulation (H6). The following sections will elaborate on these findings.

6.1.1 Retraction and coarticulation of /s/ in /str/

The acoustic analysis of the data at hand provides a strong indication that the CoG in /str/ is likely to be located at much lower frequencies than all other phonologically alveolar sibilants. The complex analysis in section 5.1 has illustrated not only that /str/ is significantly different from all other phonologically alveolar sibilants, but also that the variability present in the data can be best accounted for by a model that groups /str/ as a variable. Neither the presence of /r/ in the second position, nor the presence of /t/ reach a better fit.

Since the study focused on the /str/ environment, sufficient data for the other triple consonant clusters is lacking. Nonetheless, even with small numbers of /spr/ and /skr/, everything indicates that this particular cluster is the big bang of this change, confirming Janda and Joseph's theory. They have argued that sound change is always initiated phonetically and in a very limited context, such that /str/ can be seen as the "point of origination" for retraction as a sound change (Janda & Joseph 2003a: 11). Model comparison has shown however that the big bang is not the only factor in the present dataset. The binary classification of the second phoneme as either /r/ or not /r/ improved model fit, even when numbers are comparatively small for sCr environments other than /str/. This is further indication that consonant harmony is at play here, where the planned production of /r/ interferes with the pronunciation of the previous sibilant. While this certainly works to confirm theories in previous research, it also makes the claim stronger in many respects. I indicated above that this dataset was one of the first to be created entirely for the purpose of evaluating retraction in /str/. This is in contrast to corpus studies in which the threshold for a speaker to be included was relatively low and we can only assume the mean number of /str/ items uttered per speaker. Therefore, if this study confirms /str/-retraction it means not only that it does exist but it also means that it exists even in variable speech activities and amongst non-categorical speakers.

6.1.2 Gradualness in Language Change

Section 5.2 dealt with two different notions of gradualness in language change that have been widely debated in research as early as neogrammarian hypotheses of change (Labov 1994; Bybee 2002; Hale 2008). The question of gradualness has been addressed both through including various notions of lexical frequency and by binning the production of speakers based on interquartile dispersion.

Both analytical tools used in this analysis indicate that /s/-retraction can be characterized as gradual. Very few speakers produce a categorically retracted variant of /s/ in all styles and environments. Grouping all sibilants shows that the highest number fall in the gap between the 75th percentile of /ʃ/ and the 25th percentile of /s/, which amounts to almost half of the produced /str/-sibilants. However, there are fewer alveolar or fronted productions of the sibilant in the data set than there are retracted and very retracted productions. For the phonetic gradualness of this change it can thus be concluded that while the process is gradual, the trend goes in the direction of retraction.

Lexical frequency was also shown to have an effect on the retraction of the sibilant. In the analysis above, it became apparent that lexical diffusion is a problematic theoretical concept. This is the case because frequency effects are not linear in their impact on language users. Furthermore, frequency effects are much easier to evaluate in more common structures such as vowels rather than the small fraction of the lexicon that is occupied by the specific cluster analyzed here. The fact that the complex interplay of birth year and Lg10WF successfully improves model fit is therefore a strong indicator of lexical diffusion in this change. The analysis of a subsample of speakers in whom the pattern is prevalent adds a further aspect to this analysis, because it illustrates that lexical gradualness cannot yield stronger results due to the great differences between individuals in terms of where they currently are in the sound change. In apparent time, there is a coexistence of

fronted speakers, lexically gradual participants in the sound change, and fully phonologized retractors.

This is to say that word frequency is an especially important factor to those older speakers that began the change by retracting some words but not others. If the expectation is that the post-alveolar production of the sibilant in consonant clusters is the phonotactic reality of young and future speakers, as anecdotal evidence as well as some almost categorical retractors in the dataset seem to indicate, the resulting understanding must be that frequency is not a good predictor overall (as age would be), but rather works well to explain the pattern in some speakers, while others have already surpassed that state where frequency matters.

These confirmations of gradualness in both the phonetic and lexical nature of this change also connect the patterns at hand to exemplar models of language change. Speakers show inconclusive results in all aspects where categorization would be possible, whether that is the activity being performed or the acoustic production of the sibilant. The vast dispersion of /str/-items in the sibilant space indicates the great number of possible input and output that language users are confronted with for this variation. If this change were simply a matter of ease, more consistent patterns would be expected. The idea of an exemplar cloud that stores examples of the cluster and the words representative of the entire range of variability seems the model that best captures the present sound change.

Furthermore, in contrast to vowel analyses with large numbers of items per speaker, the specific and rare environment /str/ is of limited suitability to frequency analyses. Many of the words used in the corpus are relatively infrequent and almost all of them (except for FRUSTRATION, CONSTRUCTION and RESTAURANT) are instances of word initial /str/. A larger corpus could certainly provide a more stable and trustworthy depiction of the pattern, with more nuanced hierarchies for the individual words.

Yet, this makes finding a pattern at all even more valuable. Furthermore, as Erker and Guy have argued, there may be a threshold between frequent and infrequent words that creates a categorical rather than continuous difference (Erker & Guy 2012). Even just differences in mean underline these differences between the words, showing that there is a notion of frequency present in the data, which is also underlined by the node in the conditional tree for CONSTRUCTION and the variable importance of word context. The expectation would therefore be that more word examples would simply strengthen this notion on both ends, still visualizing the threshold between frequent and infrequent words.

6.1.3. Individuals and categorical retractors

Identifying retractors within the dataset has proven a very demanding task to perform. While there is anecdotal evidence from oblivious retractors such as the initially mentioned X-Factor contestant and the participant in my study, who struggles to correctly voice the difference in asking “am I saying shtreet or shtreet?”, there is still variability in their acoustic production that makes it impossible to truly define a threshold for categorical retractors. Whether grouped by the binning system as retracted users or by their mean difference between /ʃ/ and /str/, there seems to be no person in the entire dataset that produces all /str/ sibilants in a post-alveolar position. However, there are several speakers

whose z-scored /ʃ/ and /str/ difference remains much below 0.5 and thus less than half of the average absolute distance in the sibilant space overall.

6.1.4. Other language internal factors

The analysis of other language internal factors focuses mainly on duration and to some extent following sound. Duration is highly significant in the models. While the random effect for speaker and the factor style do account for individual differences in duration due to speech rate, a measure of speech rate would be a slightly more appropriate measure to operationalize this effect and exact duration was chosen for simplicity's sake. Nonetheless, both the significant effect of duration, the complex interplay of sibilant production with following sound and following /r/ as explained in section 4.1. as well as the non-significant differences between speech styles are evidence for a subconscious process. It has been argued in the literature that this is coarticulation rather than assimilation, because assimilation is perceived as a predictive change on the level of articulation planning related to the cognitive representation of a phoneme, while coarticulation occurs at the motor level of speech articulation (Baker, Archangeli & Mielke 2011; Phillips 2018). If faster, non-read, less planned speech is indicative of a higher retraction frequency, we can conclude that coarticulation is at play in the process.

To what extent assimilation is ever-present at the cognitive level in the sense that I allow myself to retract more because I am aware that /r/ is following appears to be splitting hairs to some extent. Whether we call it long-distance assimilation or coarticulation, consonant harmony or variability in the exemplar, there is significance to both frequency and duration. While the significance of frequency speaks for a somewhat cognitively entrenched system, the significance of duration indicates motor-control effects. I therefore do not believe that an absolute categorization is the solution here. If the system is allowed to be both phonetically and lexically gradual, exemplar theory can best account for the fact that we see both an effect of frequency and pressures of use. The concept of evolutionary phonology describes how frequency increases the probabilistic chance of motor control issues (faster pronunciation means tongue approaches /r/ more quickly and gestures overlap to a greater extent), which in turn may create change (Blevins 2006a). If incrementation is then understood as a reordering of the number of exemplars in a certain way, changes in lexical frequency are expected, at least until a change is so prevalent that the phonotactics of English are restructured. Due to the special phonotactic status of /s/ in /str/, a categorical change seems to offer no additional discriminatory benefit.

6.2 Social Aspects

The involvement of social factors in this change has thus far not been clear. While the factor age was established as meaningful and has strengthened the argument for a change in apparent time, the only other frequently studied factor in the change that has been mentioned is the factor gender (Gylfadottir 2015; Wilbanks 2017). This is partially the case because it is coded for in the corpora used for the analysis thus far.

The regression model as well as the random forest variable importance confirmed that birthyear influences retracted production in /str/-retraction to a great extent. This is the most important variable to confirm that the pattern observed is not just a common

production variant but rather indicates a sound change in apparent time, which confirms H3.

An influence of activity style on the variability of sibilant production could not be confirmed in this study (H8). There was a slight trend in the data exploration that showed that reading may be the activity with the smallest amount of retracted sibilant productions, but this pattern was not strong enough to have an effect in the regression model.

In line with previous findings, the close analysis of the patterns has shown that gender is only a very weak predictor of the change discussed here, especially in a more diverse dataset (H7). This is a surprising finding considering that this counters Labov's understanding of how females behave in language change. In line with Eckert's assumptions about how gender is meaningful, the data exploration has shown that ethnic identity and the indexicality of being a White male may have more to do with the pattern than a simple binary gender distinction.

Ethnicity is an important predictor of retraction in all models based on the present dataset, which confirms H9. Hispanic-Latinx patterned similarly to African American speakers but the model only predicted a significant difference between African American and White speakers due to the small number of speakers.

The social evaluation of this change did not show any conclusive patterns outside of prescriptive notions of how pronunciation should relate to orthography (H10).

The following sections illuminate the patterns observed in greater detail.

6.2.1 Age

Gylfadottir, Wilbanks and Stuart-Smith et al. have shown how age predicts /str/-retraction (Gylfadottir 2015; Wilbanks 2017; Stuart-Smith, Sonderegger, Macdonald, et al. 2018). The present data shows a similar trend and the mixed effects model reveals age as the best predictor, better even than the language internal factor duration. However, a close look at this specific dataset also shows that there is an increase in variability in the younger generation. This means that when we discuss this change in apparent time, there is an increasing number of speakers who use retraction. To phrase this more clearly, at this point, this is not a community wide regular change in a certain direction where everyone retracts every once in a while, but rather that the youngest age group shows the entire range from a fully alveolar production of /s/ in /str/ clusters to an almost fully retracted rendition of the sound. There is also added complexity represented in the fact that the change seems to occur much earlier in some African American speakers. If we just look at the White population, we can clearly state that a growing number of speakers is retracting in the younger generations, but the community-wide sibilant space for /str/ still shows all sibilant productions. This finding is different from Wilbanks' age trend, which indicates that males are decreasing in retraction rates compared to females in apparent time (Wilbanks 2017).

6.2.2 Gender

Some studies have investigated the possibility of a female-led change in apparent time. Gender was investigated both by Wilbanks and Gylfadottir (Gylfadottir 2015; Wilbanks 2017). The small numbers of participants in other studies did not lend themselves to such evaluations.

The extent to which the three-way-interaction reported in Wilbanks' mixed effects model of the /str/-subset truly shows a significant gender effect cannot be thoroughly evaluated without knowing how he decided on the best model. I have addressed the substantial differences in approaches to when to include parameters (e.g. keeping it maximal) and I have shown how both R2 as well as variable importance can be employed to test effect strength. Wilbanks however neither reports steps taken to establish the significance of a three-way-interaction nor the model outcome itself apart from a visual display of estimates (Wilbanks 2017).

The present investigation found no significant results for gender, much like Gylfadottir (Gylfadottir 2015). Once again, this seems to be representative of the pattern for White speakers, but not for everyone else. To the best of my knowledge, the Labovian understanding of the female role in change from below was largely based on studies of White males and females. This may explain the pattern at hand in the White community. The meta commentary seems to support the lack of awareness of the phenomenon at hand and a focus on just the White community shows exactly what we would expect from such a change.

The notion of Labov's concept provides agency to females. Yet, a lack of movement in the White male community seems to be the anomaly here when all ethnic groups are taken into consideration. They seem to overlap with the notion of the non-mobile older rural male (NORM) that early dialectologists considered their best subjects (Trudgill 1986). While White males show a relatively stable pattern in this dataset, everybody else is changing. If we take current discussions of gender and masculinity in western societies into consideration, a change in perspective may be necessary. So instead of saying that females are the ones stepping up and doing something different from a sociolinguistic point of view, a more appropriate interpretation of the pattern presented above would be that White males are just the least likely to change their patterns until everybody else is doing it. So rather than calling this a female-led change from below, it is a change from below that everyone but the White males in this dataset participates in.

6.2.3 Style

Style is a complicated concept to measure. The present study, as indicated above, is truly only investigating differences in activity types without making any claims about levels of attention paid to speech or the vernacularity of interview speech. It is due to the nature of the /str/-cluster that even the most detailed and well-planned study cannot induce /str/-clusters in casual conversation. The mere quantitative difference between items in the three activity groups can therefore explain why style did not significantly influence the modeled change in the mixed effects regression. It is all the more important therefore to look at the trends with these flaws in mind to understand how and why style may relate to /str/-retraction. Naming, the activity with the most words, is slightly more variable than reading as presented above, but significantly more retracted and the median is similar to the median in the story interview. Due to the low level of conscious awareness, this shows that the change is present in all but the most cognitively complex and ideologically prescriptive speech activity. We can therefore say that individuals do differentiate between speech activities in retracting /str/.

6.2.4 Ethnicity

In section 6.2.2, I have started to explain the importance of the factor ethnicity on /str/-retraction. What would simply look like a female-led change in apparent time when we focus solely on self-identified White speakers, becomes much more complex when Hispanic-Latinx and African American speakers are analyzed along with them. Not only is the retraction rate greater amongst these two groups, but both their median CoG measurement and their standardized CoG measurements are continuously lower than the ones produced by White speakers. It would be necessary to study whether the lower acoustic production has any effect on it being perceived as retracted.

Without any such knowledge however, the pattern both visible in the graphs and modeled in the overall model of the change depicts how much more likely a value below the overall regression line and slope is when the speaker is African American. This is the case even with z-scored measurements that reflect a position within the sibilant space rather than an acoustic measurement. These two groups are also the ones that present with a very slight difference between the production of /str/ and the production of /ʃ/ in apparent time, while the White speakers still produce this difference to a greater extent.

With none of these speakers, we can be certain to what extent they speak with a sociolect that we would consider the type of distinct variety described above as Chicano or African American English. It could certainly be the case that what is perceived here are not differences between speakers of the same variety but rather differences between three distinct varieties. There are some reasons to consider the change a change amongst different ethnic groups rather than a change in specific varieties. First of all, the change is present in all three groups. Secondly, the White individuals are possibly just as affected by regional varieties as the other two ethnicities are affected by ethnolects. In the sample it appeared that participants spoke with a Texan coloring while others could less easily be identified as regional speakers. Yet all of the groups show retraction in some speakers. Furthermore, no perception study has delivered conclusive results pinpointing a connection to a certain ethnicity. All this supports the assumption that /str/-retraction is not an aspect of a certain sociolect or dialect but rather a variable feature being used to varying degrees by all speakers of American English.

Not all of the speakers in this study grew up in Austin. Both in Chapter 4 and in section 5.8, I have depicted some of the ethnographic reasons why the behavior by Hispanic-Latinx and African American speakers here is important. Especially historically, East Austin is both home to African American and Hispanic-Latinx communities. This does not mean that they live on the same block or share a communal space, as described by Participant 66 in the excerpt below:

> P66: I don't hear it as much as I used to I did because neighbors kind of gentrified and some of them a lot of the neighbors here are not the same
> I: okay so and are you saying then that that's a feature that's prominent amongst African American English
> P66: umh yeah and I would that and some Hispanic-Latinxs
> I: I'm just because of East Austin I'm just wondering

P66: right yeah I didn't really grow up around a lot of blacks we stayed in they stayed in theirs and we stayed in ours but you did have some cross over went to high school
I: so you grew up mainly with Hispanic-Latinxs then
right and then I mean I didn't really see a lot of blacks until I went to Middle School together catholic school together but yeah I would say then mostly around Hispanic-Latinxs when I was young

In the communal spaces that I visited, there is currently a varying degree of mixing. Two of the community centers were only frequented by African American seniors, while events for younger generations attracted a more mixed crowd. This mixing is rather complex, because for the environments that I worked in it often depended on income. As mentioned above, the relative number of African American residents in Austin is decreasing due to the strong gentrification of formerly black neighborhoods in South East Austin. Within an environment where the perception of "non-white" is still felt by members of the community, while the community itself is getting smaller and smaller, it comes as no surprise that features are shared amongst the "non-whites" in Austin. Another factor that may play into this is the fact that many of the young Hispanic-Latinxs in my study are second generation immigrants who are developing a US Hispanic-Latinx identity with the plethora of features available for these formation processes in the linguistic marketplace.

As a result of these circumstances, it is not surprising that features seem to develop in the African American community first and the Hispanic-Latinxs share them long before they are picked up on by White speakers. Since duration and other language internal factors were less predictive in the regression model than ethnicity, it seems to be more than just a more relaxed speaking style or awareness/attention of linguistic correctness that informs this change. There is something beyond coarticulation that started this change in Austin. Without an analysis of covarying features and a matched-guise study, "ethnic community" as a factor seems to make this change more likely to happen. I am not arguing that there is any overt indexical meaning to this change yet.

6.2.5 Individuals and retraction

Individual variation was investigated at three different points within the data set. I first looked at specific individual differences by comparing the absolute distance between their average post-alveolar sibilant and their production of the sibilant in the /str/-cluster. The findings here showed that there are some speakers that produce almost all cluster-items similar to their post-alveolar and thus strongly retract the sound. These individuals are mainly younger speakers. However, other than their age, the differences are individual and no conclusive social profile of the stereotypical retractor could be gathered from the data.

6.3.6 Perception and indexicality of /str/-retraction

The thematic analysis of comments made by participants revealed both surprising and expected results. While the strong opinions indicated in section 1.1 would lead anyone to believe that /str/-retraction is a widely recognized and problematic sound change in the English language, the expert expectation for a change from below would not be in line with such strong negative emotions.

The debriefing section brought forth several interesting aspects that add to the understanding of the sound change presented here. One of them is the prescriptivism involved that rests on orthography. Many participants voiced the opinion that /s/ should be produced alveolar, because they directly connect the post-alveolar production to being spelled with H. And even if one considers that some post-alveolar productions are instead spelled with SS such as in the word SESSION, in contrast to many other phonemes in the English language, the post-alveolar sibilant is very consistently marked orthographically. This results in a salience of this phonetic feature in orthography that further explains the discrepancy between read speech and free speech.

A component of the opinion that follows from these notions of correctness is the hope to say it "right" and the use of awareness and stress as excuses for doing it "wrong". The use of phrases such as "lazy tongue pattern" or "tongue-tied" illustrates this understanding. Furthermore, one participant mentions being directly called out for retraction, which shows that there may be communities in which the feature is initially salient.

The indexical field of the variable was not easily determined in the interviews. Some participants describe the feature as individual or random. There seems to be some significance to metropolitan areas in the North East, Southernness and African Americanness as factors in retraction. The only meaningful way to connect these regions would be through the great migration and thus African American history (Farrington 2019). Based on the varied responses, this interpretation is slightly too far-reaching in the current study.

7. Conclusion

The present study has both underlined some of the hypotheses connected to /str/-retraction and broadened the perspective on its social profile. New methods were applied to investigate the status of /str/ retraction both as a sound change and a sociolinguistic change. The sound change investigation combined the theory of language change, exemplar models of cognitive processing and the involvement of the lexicon in language change.

Especially the analysis of frequency effects in /str/-retraction is new to the investigation of this process and a promising approach to further understand the nuances of this change. A more varied dataset with even greater numbers of /str/-tokens per speaker would certainly help to understand this mechanism better. Through this analysis I have shown that even if there is an effect of lexical diffusion, the fact that individuals in the change are at different stages within the change spanning from no retraction to certain words to almost categorical and regular retraction not only explains why these effects are so rarely significant in our models of change, but it also questions the entire theoretical set up of ever describing a sound change as regular/categorical or diffuse/gradual. A study with greater numbers of tokens per speakers would also benefit from more items with word-medial /str/, both because of the effect shown by Wilbanks and because of the special status of the word CONSTRUCTION that became apparent in the conditional tree (Wilbanks 2017).

The greatest contribution to the state of the art on /str/-retraction is however that I have shown that an ethnically homogenous corpus underestimates the complex interplay of social factors involved in the change. The emergence of the phenomenon is so widespread that it seems as though the arguments for coarticulation render it uninteresting from a sociolinguistic point of view. Both initial and sociolinguistic salience are so low for the phenomenon that any language external explanation seems to become obsolete. Both my own qualitative and experimental approaches were thus far unsuccessful in pinpointing any such salience. It is questionable whether even first-order indexicality exists for the phenomenon.

And yet, the documented effect of ethnicity shows that this change may have originated in African American English. If African Americans are modeled as more likely to retract, then we are missing an interesting puzzle piece in this investigation. This is not to say that this factor can account for /str/-retraction in English globally, but it is undeniably a factor that warrants further research.

This further research should then include a more diverse sample of speakers. The lower socio-economic status that united many of my Hispanic-Latinx and African American participants needs further investigation. Many of the corpora we use to ask questions about a sound change currently include a questionable level of diversity.

I think that ease of articulation gets a whole new meaning when we consider it on a larger scale in connection to this social profile. If the pressures for identity formation are greater in denser networks such as the community centers I visited, the tipping point of the scale between communication and ease, and similarly communication and stance practices may change. Expecting the same tipping point for all speakers negates variability where it may crucially influence production. The resulting assumption would be that coarticulation

as a motor-control process is present to a greater degree in certain groups of speakers than in others. Ethnicity may be a factor in this analysis that simply masks this type of differentiation, where the attraction of ease is simply more likely in this specific sample for speakers of certain ethnicities. The widespread productions of White young speakers may also point in this direction. The reason for a change would then not be the sociolinguistic salience of the feature itself but the sociolinguistic salience of ease. In order to evaluate such patterns, a multi-variable approach is necessary.

Furthermore, research aiming at indexicality and salience of the feature was thus far limited to Phillips and Resniks investigation of masculinity and attraction (Phillips & Resnick 2019). Considering that White males are the least likely to retract, it is not surprising that this connection is not meaningful in social classification. Whether the feature carries any ethnic indexicality has thus far not been tested, although my open interviews do not offer support for any overt indexicality of the phenomenon. The assumption is thus that /str/-retraction carries first order indexicality, if any. It is questionable whether any experimental design would find sociolinguistic salience, but the differences in categorization of /str/ and other clusters in Phillip and Resniks study as well as the strong trend in apparent time would certainly speak for some level of indexicality.

A further aspect to be considered in an evaluation of the indexicality and salience of the variable is the possibility of multiple meanings both per region and per community of practice as well as the complexity of its indexical field. The feature may include meanings such as power, masculinity and ethnicity without a one-way interpretation in any of them. As Campbell-Kippler cautions: "even […] key meanings are found only in restricted areas of the data, overcome in other instances by factors such as regional accent or perceived class. This malleability as a result of other features is predictable from work in impression formation which has documented the degree to which individuals are perceived as gestalt wholes, with more central attributes strongly influencing the social contribution of less central attributes" (Campbell-Kibler 2010).

I assume that these key meanings do not exist for /str/-retraction as of yet and its progression as a sound change is currently motivated mostly by language internal factors, which explains the parallels in global spread. Baker et al. have explained this in probabilistic terms (Baker, Archangeli & Mielke 2011). The strong effect of ethnicity in my dataset however shows that more than the language internal factors are at play in Austin. This doesn't negate the probabilistic likelihood of retraction, but it gives the change a social profile. Not everyone is equally likely to retract and the social profile of this pronunciation has changed across generations so that it is a pronunciation variant among the entire community in apparent time where older White speakers did not show any indication of retraction. I have made sense of this pattern above. The point here is to say that just because at the community level, the described malleability of meanings prohibits uni-directional indexicality, the resulting interpretation should not be that a change is purely influenced by language internal factors. Furthermore, just because a feature is probabilistically common in a language that spans the globe, the interpretation should not be that it cannot be a sociolinguistic change within the community that uses it. First results of comparative analyses from Trinidad English and the data presented here show that /str/-retraction is

connected to a notion of "Americanness", which fails to explain any of the observed patterns in Austin (Ahlers & Meer 2019). For British English, Nichols and Bailey draw a connection between /stj/ and /str/ that also is not as functional in explaining the Austinite perspective (Nichols & Bailey 2018).

The argument is therefore that more research is necessary to understand the sociolinguistic part of the change. The global spread of a phenomenon does not negate local meaning and it seems that the overwhelming amount of acoustic studies with few social factors accounted for in the design speaks for a reluctance to dive any deeper. Furthermore, absence of categorical initial salience is taken to negate sociolinguistic salience, which is theoretically problematic as Hay and Foulkes have shown that social context can be negotiated in an exemplar way so that the connection between social meaning and retraction may be more salient/strongly connected in some contexts than in others (Hay & Foulkes 2016).

This thesis has provided just as many perspectives on a complex sociophonetic phenomenon as it has identified pathways for future research in /str/-retraction that exceed any of the research that was done so far. Some of these pathways include new types of data with more diverse social and linguistic context, others lead to evaluations of new aspects such as indexicality, sociolinguistic salience and covariation with other variables. While the amount of research on /str/-retraction has greatly increased since I first collected the data presented in this study, I am certain that a more nuanced understanding of this change is yet much further down the shtreet.

Works Cited

Agha, Asif. 2005. Voice, Footing, Enregisterment. *Journal of Linguistic Anthropology* 15(1), 38–59.

Ahlers, Wiebke & Philipp Meer. 2019. Sibilant variation in New Englishes : A comparative sociophonetic study of Trinidadian and American English /s(tr)/-retraction. *INTERSPEECH*, 291–295.

Auer, Peter, Birgit Barden & Beate Grosskopf. 1998. Subjective and Objective Parameters Determining 'Salience' in Long-term Dialect Accommodation. *Journal of Sociolinguistics* 2(2), 163–187.

Baayen, R. H., D. J. Davidson & D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4), 390-412.

Bailey, Guy. 2008. Real and Apparent Time. In J. K. Chambers & Natalie Schilling (eds.), *The Handbook of Language Variation and Change*. Malden: Blackwell, 312–332.

Baker, Adam, Diana Archangeli & Jeff Mielke. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23(03), 347–374.

Bakhtin, Mikhael M. 1981. *The Diologic Imagination*. (Ed.) C. Emerson & M. Holquist. Austin: University of Texas Press.

Barr, Dale, Roger Levy, Christoph Scheepers & Harry J Tily. 2014. Keep it maximal. *Journal of Memory and Language* 68(3), 1–43.

Barth, Danielle & Vsevolod Kapatsinski. 2018. Evaluation Logistic Mixed Effects Models of Corpus Lingusitic Data in Light of Lexical Diffusion. In Dirk Speelman, Kris Heylen & Dirk Geeraerts (eds.), *Mixed Effects Regression Models in Linguistics*. Cham: Springer International Publishing, 254–303.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2018. Parsimonious Mixed Models. Available from arXiv:1506.04967

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* 67(1).

Bauer, Laurie. 2014. English phonotactics. *English Language and Linguistics* 19(3), 437–475.

Baxter, Gareth & William Croft. 2016. Modeling language change across the lifespan: Individual trajectories in community change. *Language Variation and Change* 28(2), 129–173.

Bayley, Robert. 1994. Consonant cluster reduction in Tejano English. *Language Variation and Change* 6, 303–326.

Bayley, Robert. 2018. The Quantitative Paradigm. In J. K. Chambers & Natalie Schilling (eds.), *The Handbook of Language Variation and Change*.2nd ed. Hoboken, 85–107..

Becker, Kara. 2009. /r/ and the construction of place identity on New York City's Lower East Side. *Journal of Sociolinguistics* 13(5), 634–658

Becker, Kara & Elizabeth L. Coggshall. 2009. The sociolinguistics of ethnicity in New York city. *Linguistics and Language Compass* 3(3), 751–766.

Beebe, Leslie M & Howard Giles. 1984. Speech-accomodation theories: a discussion in terms of second-language acqusition. *International Journal of the Sociology of Language* 46, 5–32.

Bell, Allan. 2001. Back in Style. In Penelope Eckert & John R. Rickford (eds.), *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(02), 145.

Benor, Sarah Bunin. 2010. Ethnolinguistic repertoire: Shifting the analytic focus in language and ethnicity. *Journal of Sociolinguistics* 14(2), 159–183.

Bernstein, Basil. 1962. Social Class, Linguistic Codes and Grammatical Elements. Language and Speech 5(4), 221–240.

Blevins, Juliette. 2006a. *Evolutionary Phonology*. Cambridge: Cambridge University Press.

Blevins, Juliette. 2006b. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32(2), 117–166.

Bloomfield, Leonard. 1984. *Language*. 3rd ed. Chicago: The University of Chicago Press.

Blumstein, S.E. 1989. Theoretical implications of the quantal nature of speech: a commentary. *Journal of Phonetics* 17(1–2).

Boersma, Paul & Silke Hamann. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25(2), 217–270

Bois, John W. Du. 2007. The stance triangle. In Robert Englebretson (ed.), *Stancetaking in Discourse*. Amsterdam: John Benjamins Publishing Company, 139–182.

Bourdieu, Pierre. 1977. The economics of linguistic exchanges. *Social Science Information* 16(6), 645–668.

Braun, Virginia & Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3(2), 77–101.

Browman, Catherine P & Louis Goldstein. 1992. Articulatory Phonology: An Overview. Haskins Laboratories Status Report.

Brown, Penelope & Stephen Levinson. 1994. *Politeness: Some universals in language usage.* Cambridge: Cambridge University Press.

Brysbaert, Marc & Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41(4), 977–990.

Burnett, Heather, Hilda Koopman & Sali A. Tagliamonte. 2018. Structural explanations in syntactic variation: The evolution of English negative and polarity indefinites. *Language Variation and Change* 30(1), 83–107.

Burnham, Kenneth P., David R. Anderson & Kathryn P. Huyvaert. 2011. AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. Behavioral Ecology and Sociobiology 65(1), 23–35.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14(May), 261–290.

Bybee, Joan. 2017. Grammatical and lexical factors in sound change: A usage-based approach. *Language Variation and Change* 29(03), 273–300.

Callahan, Erin. 2017. Interlanguage and Cross-Generational Assimilation: Past Tense Unmarking in Hispanicized English. *Journal of English Linguistics* 45(2), 103–129.

Campbell-Kibler, Kathryn. 2010. Sociolinguistics and Perception. *Linguistics and Language Compass* 4(6), 377–389.

Census. 2017. City of Austin, Texas.

Chambers, J. K. 1996. *Sociolinguistic Theory. Linguistic Variation and its Social Significance*. Cambridge: Blackwell.

Chambers, J K. 2002. Studying Language Variation: An Informal Epistemology. In J K Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 3–14. Blackwell.

Chen, Matthew Y. & William S-Y. Wang. 1975. Sound Change: Actuation and Implementation. *Language* 51(2), 255–281.

Cheshire, Jenny. 2008. Sex and Gender in Variationist Research. In J K Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 423–443.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge: The MIT Press.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Clements,George N. 1975. *Vowel Harmony in Nonlinear Generative Phonology: An Autosegmental Model*. Indiana University Linguistics Club.

Clyne, Michael. 2000. Lingua Franca and Ethnolects in Europe and Beyond. *Sociolinguistica* 14(1), 83–89.

Coates, Jennifer. 1986. *Women, men, and language: a sociolinguistic account of sex differences in language*. London: Longman.

Coleman, John. 1998. Cognitive reality and the phonological lexicon: A review. *Journal of Neurolinguistics* 11(3), 295–320.

Coupland, Nikolas. 1980. Style-Shifting in a Cardiff Work-Setting. *Language in Society* 9(1), 1–12.

Coupland, Nikolas. 2001. Age in social and sociolinguistic theory. *Sociolinguistics and Social Theory,* 185–211. Edinburgh: Pearson Education Limited.

Coupland, Nikolas. 2007. *Style: Language Variation and Identity*. Cambridge: Cambridge University Press.

Coupland, Nikolas. 2014. Language change, social change, sociolinguistic change: A meta commentary. *Journal of Sociolinguistics* 18(2), 277–286.

Currie, Haver C. 1952. A projection of socio-linguistics: The relationship of speech to social status. *The Southern Speech Journal* 18(1), 23–37.

Dart, Sarah N. 1991. Articulatory and Acoustic Properties of Laminal Articulations. UCLA Working Papers in Phonetics. Los Angeles.

Davidsen-Nielsen, Niels. 1975. A Phonological Analysis of English sp, st, sk with Special Reference to Speech Error Evidence. *Journal of the International Phonetic Association* 5(1), 3–25.

Deuchar, Margaret. 1990. *A pragmatic account of women's use of standard speech. Women in Their Speech Communities. New Perspectives on Language and Sex.* London: Longman.

Dillard, Joey L. 1972. *Black English: Its history and usage in the United States*. New York: Random House.

Dresher, B. Elan. 2011. The Phoneme. *The Blackwell Companion to Phonology*, 241–266.

Durian, David. 2007. Getting [(esh)]tronger Every Day?: More on Urbanization and the Socio-Geographic Diffusion of (str) in Columbus, OH. *University of Pennsylvania Working Papers in Linguistics* 13(2), 64–79.

Eckert, Penelope. 1997. Age as a Sociolinguistic Variable. In Florian Coulmas (ed.), *The Handbook of Sociolinguistics*, 151–169. Oxford: Blackwell.

Eckert, Penelope. 2001. Style and social meaning. In Penelope Eckert (ed.), *Style and Sociolinguistic Variation*, 119–128. Cambridge: Cambridge University Press.

Eckert, Penelope. 2008a. Where do ethnolects stop? *International Journal of Bilingualism* 12(1–2), 25–42.

Eckert, Penelope. 2008b. Variation and the indexical field. *Journal of Sociolinguistics* 12(4), 453–476.

Eckert, Penelope. 2012. Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology* 41(1), 87–100.

Eckert, Penelope. 2013. Research Ethics in Linguistics. In R. J. Podesva & Devyani Sharma (eds.), *Research Methods in Linguistics*, 11–26. Cambridge: Cambridge University Press.

Eckert, Penelope & McConnell-Ginet. 2003. *Language and Gender*. Cambridge: Cambridge University Press.

Eckert, Penelope & Sally McConnell-Ginet. 1992. Think Practically and Look Locally : Language and Gender as Community- Based Practice. *Annual Review of Anthropology* 21, 461–490.

Edgell, Stephen. 1993. *Class*. New York: Routledge.

Erker, Daniel & Gregory R. Guy. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88(3), 526–557.

Fant, Gunnar. 1971. *Acoustic Theory of Speech Production*. The Hague: Mouton.

Farnetani, Edda & Daniel Recasens. 2010. Coarticulation and Connected Speech Processes. In William Hardcastle, John Laver & Fiona E Gibbon (eds.), *The handbook of Phonetic Sciences*. 2nd ed.

Farrington, Charles Ryan. 2019. *Language Variation and the Great Migration: Regionality and African American Language*. Dissertation: University of Oregon.

Fasold, Ralph W. 1990a. *The Sociolinguistics of Language*. Oxford: Blackwell.

Fasold, Ralph W. 1990b. Tense Marking in Black English. *WORD* 41(2), 203–218.

Fasold, Ralph W, William Labov, Fay Boyd Vaughn-Cooke, Guy Bailey, Walt Wolfram, Arthur K Spears & John Rickford. 1987. The American Dialect Society ARE BLACK AND WHITE VERNACULARS DIVERGING? American Speech 62(1), 3–80.

Fought, Carmen. 1997. *The English and Spanish of Young Adult Chicanos*. Dissertation. University of Pennsylvania.

Fought, Carmen. 2007. Ethnicity. In J K Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 444–472.

Fought, Carmen. 2010. Language as a representation of Mexican American identity. English Today 26(3).

Fridland, Valerie, Kathryn Bartlett & Roger Kreuz. 2004. Do you hear what I hear? Experimental measurement of the perceptual salience of acoustically manipulated vowel variants by Southern speakers in Memphis, TN. *Language Variation and Change* 16(1), 1–16.

Fruehwald, Josef. 2016. The early influence of phonology on a phonetic change. *Language* 92(2), 376–410.

Fruehwald, Josef. 2017. Generations, lifespans, and the zeitgeist. *Language Variation and Change* 29(1), 1–27.

Garrett, Andrew. 2015. Sound Change. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, 227–248.

Goldinger, S D. 1996. Words and voices: episodic traces in spoken word identification and recognition memory. *J Exp.Psychol.Learn.Mem.Cogn*. 22(5), 1166–1183.

Gordon, Elizabeth. 1997. Sex , Speech , and Stereotypes : Why Women Use Prestige Speech Forms More than Men. *Language in Society* 26(1), 47–63.

Guy, Gregory R. 2011. Variability. The Blackwell Companion to Phonology, 1–24.

Gylfadottir, Duna. 2015. Shtreets of Philadelphia : An Acoustic Study of / str / -retraction in a Naturalistic Speech Corpus Shtreets of Philadelphia : An Acoustic Study of / str / -retraction in a. *University of Pennsylvania Working Papers in Linguistics* 21(2), 89–97.

Hale, Mark. 2008. Neogrammarian Sound Change. In Brian Joseph & Richard Janda(eds.) *The Handbook of Historical Linguistics*, 343–368.

Halle, Morris, Bert Vaux & Andrew Wolf. 2000. On Feature Spreading and the Representation of Place of Articulation. *Linguistic Inquiry* 31(3), 387–444.

Hansson, Gunnar Ólafur. 2001. Theoretical and Typological Issues in Consonant Harmony. Genes & Development. Dissertation. University of California Berkeley.

Hardcastle, W. J. 1985. Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication* 4(1), 247–263.

Harrington, Jonathan. 2010. 3. Acoustic Phonetics. In William J Harcastle, John Laver & Fiona E Gibbon (eds.) *The Handbook of Phonetic Sciences*. Malden: Blackwell.

Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America* 123(5), 2825–2835.

Hay, Jennifer. 2018. Sociophonetics: The Role of Words, the Role of Context, and the Role of Words in Context. *Topics in Cognitive Science* 10, 696–706.

Hay, Jennifer & Paul Foulkes. 2016. The evolution of medial /t/ over real and remembered time. *Language* 92(2), 298–330.

Herk, Gerard Van. 2015. The English Origins Hypothesis. In Sonja Lanehart (ed.), *The Oxford Handbook of African American Language*, 23–34. New York: Oxford University Press.

Hickey, Raymond. 2000. Salience, stigma and standard. In Laura Wright (ed.), *The Development of Standard English 1300–1800. Theories, descriptions, conflicts*, 57–72. Cambridge: Cambridge University Press.

Hinrichs, Lars, Alexander Bergs, Axel Bohmann, Erika Brozovsky, Brian Hodge, Kirsten Meemann & Patrick Schultz. 2015. Sibilants and ethnic diversity: A Sociophonetic Study of Palatalized /s/ in STR-Clusters among Hispanic, White, and African-American Speakers of Texas English. NWAV 44.

Hlavac, Marek. 2018. *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2.1. R.

Holliday, Nicole R. 2016. Identity Performance among Black / Biracial Men through Intonation : Examining Pitch Accents and Peak Delay. *University of Pennsylvania Working Papers in Linguistics* 22 (Select(2), 71–80.

Holmes-Elliott, Sophie & Erez Levon. 2017. The substance of style : Gender , social class and interactional stance in / s / -fronting in southeast England. *Journal of Sociolinguistics* 55(5), 1045–1072..

Holmes, Janet. 1998. Response to Koenraad Kuiper. *Journal of Sociolinguistics*, 104–106.

Hothorn, T., K. Hornik, C. Strobl & A. Zeileis. 2017. *party: A laboratory for recursive partytioning*. R package version 1.2-3. (1994).

Jaeger, T. Florian & Kodi Weatherholtz. 2016. What the heck is salience? How predictive language processing contributes to sociolinguistic perception. *Frontiers in Psychology* 7(AUG), 1–5.

Jaffe, Alexandra. 2009. *Stance. Sociolinguistic Perspectives*. New York: Oxford University Press.

Janda, Richard D., and Brian D. Joseph. 2001. Reconsidering the canons of sound-change: Towards a "big bang" theory. In B. J. Blake, K. Burridge, and J. Taylor(eds.) *Historical Linguistics 2001: Selected Papers from the 15th International Conference on Historical Linguistics*.

Kraljic, Johnson, Keith. 1997. Speech Perception without Speaker Normalization: An Exemplar Model. In Keith Johnson & John W. Mullennix (eds.), *Talker Variability in Speech Processing*, 145–65. San Diego: Academic.

Johnson, Keith. 2003. *Acoustic & Aditory Phonetics*. Second Ed.. Malden: Blackwell.

Johnson, Keith. 2005. *Decisions and mechanisms in exemplar-based phonology*. UC Berkeley Phonology Lab Annual Report.

Johnstone, Barbara. 2016. Enregisterment: How linguistic items become linked with ways of speaking. *Linguistics and Language Compass* 10(11), 632–643.

Johnstone, Barbara, Jennifer Andrus & Andrew E. Danielson. 2006. Mobility, Indexicality, and the Enregisterment of "Pittsburghese." *Journal of English Linguistics* 34(2), 77–104.

Jongman, Allard, Ratree Wayland & Serena Wong. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America* 108(3), 1252.

Kerr, Beverly. 2018. High Tech Industry. https://www.austinchamber.com/blog/06-05-2018-high-tech-industry.

Kerswill, Paul. 1996. Children, adolescents, and language change. *Language Variation and Change* 8, 177–202.

Kerswill, Paul & Ann Williams. 2002. "Salience" as an explanatory factor in language change: Evidence from dialect levelling in urban England. In Mari Jones & Edith Esch (eds.) *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*, 81–110. Berlin: De Gruyter.

Kiesling, Scott. 2009. Style as Stance. In Alexandra Jaffe (ed.) *Stance: Sociolinguistic Perspectives*, 45–66.

Kohn, Mary Elizabeth & Hannah Askin Franz. 2009. Localized patterns for global variants: The case of quotative systems of African American and latino speakers. *American Speech* 84(3), 259–297.

Kuha, Jouni. 2004. AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods and Research* 33(2), 188–229.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13). doi:10.18637/jss.v082.i13.

Labov, William. 1963. The Social Motivation of a Sound Change. *WORD* 19(3), 273–309.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 1981a. *Field Methods of the Project on Linguistic Change and Variation*. Austin.

Labov, William. 1981b. Resolving the Neogrammarian Controversy. *Language* 57(2), 267–308.

Labov, William. 1984. Field Methods of the Project on Linguistic Change and Variation. In John Baugh (ed.), *Language in Use*, 28–53. Englewood Cliffs: Prentice-Hall.

Labov, William. 1990. The intersection of sex and social factors in the course of language change. *Language Variation and Change* 2, 205–254.

Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Malden: Blackwell.

Labov, William. 2001. *Principles of Linguistic Change. Social Factors*. Malden: Blackwell.

Labov, William. 2010. *Principles of Linguistic Change: Cognitive and Cultural Factors*. Malden: Wiley Blackwell.

Labov, William, Sharon Ash & Charles Boberg. 2006. 18. The South. *The Atlas of North American English*, 240–262. New York: Mouton de Gruyter.

Labov, William, Ingrid Rosenfelder & Josef Fruehwald. 2013. One Hundred Years of Sound Change in Philadelphia: Linear Incrementation, Reversal, and Reanalysis. *Language* 89(1), 30–65.

Ladd, D. R. 2014. Phonetics in Phonology. In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The Handbook of Phonological Theory*, 348–373. Second. Malden: Wiley Blackwell.

Ladefoged, Peter. 1993. *A course in Phonetics*. Fort Worth: Harcourt Brace Jovanovich College Publishing.

Ladefoged, Peter. 2006. *Vowels and Consonants: an introduction to the sounds of languages*. Malden: Blackwell.

Ladefoged, Peter & Ian Maddieson. 1996. *The Sounds of the World's Languages*. 9th ed. Malden: Blackwell.

Lavie, Rene-Joseph. 2007. *Exemplar theory in linguistics: a perspective for the cognitive subject.* Communication to the 11th Congress of Cognitive Linguistics, 1–21.

Lawrence, Wayne P. 2000. /str/ -> /Str/: Assimilation at a distance? *American Speech* 75(1), 82–86.

Liljencrants, Johan & Björn Lindblom. 2011. Numerical Simulation of Vowel Quality Systems : The Role of Perceptual Contrast. *Linguistic Society of America* 48(4), 839–862.

Lindblom, B. 1990. Explaining Phonetic Variation: A sketch of the H&H theory. In William J. Hardcastle & Alain Marchal (eds.), *Speech Production and Speech Modelling*, 403–440. Dordrecht: Kluwer Academic Publishers.

Lindblom, Björn E F. 1986. Phonetic universals in vowel systems. In Ohala & Jaeger (eds.), *Experimental Phonology*, 13–44.

Lippi-Green, Rosina. 2012. *English with an Accent: language, ideology and discimination in the United States*. 3rd ed. London: Routledge.

Luce, Paul A. & David Pisoni. 1998. Recognizing Spoken Words: The Neighborhood Activation Model. *Ear & Hearing* 19(1), 1–36.

Luke, Steven G. 2017. Evaluating significance in linear mixed-effects models in R. B*ehavior Research Method*s 49(4), 1494–1502..

Marx, Karl. 1887. *Capital: A Critique of Political Economy.*

McCarthy, John J. 2007. Consonant Harmony via Correspondence: Evidence from Chumash. *Papers in Optimality Theory* 3(January), 223–237.

Mielke, J., A. Baker & D. Archangeli. 2010. Variability and homogeneity in American English /r/ allophony and /s/ retraction. *Laboratory phonology* 10, 699–730.

Mielke, Jeff. 2008. *The emergence of distinctive features*. New York: Oxford University Press.

Mielke, Jeff & Elizabeth Hume. 2006. Distinctive Features. In Frank Keil & Robert Wilson (eds.) *The MIT Encyclopedia of the Cognitive Sciences*, 723–731.

Milroy, James & Lesley Milroy. 1993. Mechanisms of change in urban dialects: the role of class, social network and gender. *International Journal of Applied Linguistics* 3(1), 57–77.

Mufwene, Salikoko. 2015. The Emergence of African American English: Monogenetic or Polygenetic? With or Without "Decreolization"? Under how much Substrate Influence? In Sonja Lanehart (ed.), *The Oxford Handbook of African American Language*, 57–84. New York: Oxford University Press.

Müller, Samuel, J. L. Scealy & A. H. Welsh. 2013. Model selection in linear mixed models. *Statistical Science* 28(2), 135–167.

Nakagawa, Shinichi, Paul C.D. Johnson & Holger Schielzeth. 2017. The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface* 14(134).

Nakagawa, Shinichi & Holger Schielzeth. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4(2), 133–142.

Nichols, Stephen & George Bailey. 2018. *Varying sibilant articulation and s-tratraction in Manchester English*. Invited Talk at University of Manchester Linguistics and English Language Research Seminar.

Nolan, Francis, Tara Holst & Barbara Kühnert. 1996. Modelling [s] to [ʃ] accommodation in English. *Journal of Phonetics* 24(1), 113–137.
Ohala, John J. 1993a. Coarticulation and Phonology. *Language and Speech* 36(2), 155–170.
Ohala, John J. 1993b. The phonetics of sound change. In Charles Jones (ed.) *Historical Linguistics: Problems & Perspectives.*London: Longman, 237-278.
Ohala, John J. & Maria-Josep Solé. 2008. *Turbulence and Phonology*. UC Berkeley annual lab report.
Olive, Joseph, Alice Greenwood & John Coleman. 1993. *Acoustics of American English Speech*. New York: Springer-Verlag.
Page, R. B. Le & Andrée Tabouret-Keller. 1985. *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.
Paolo, Marianna Di & Malcah Yaeger-Dror. 2011. *Sociophonetics: a student's guide*. London: Routledge.
Phillips, Jacob, B. 2016. Phonological and prosodic conditioning of /s/-retraction in American English. Paper presented at Higher order structure in speech variability, LabPhon15 satellite workshop, 1–2.
Phillips, Jacob B. & Paige Resnick. 2019. Masculine toughness and the categorical perception of onset sibilant clusters. *The Journal of the Acoustical Society of America* 145(6).
Phillips, Jacob B. 2018. Temporal dynamics of / s / -retraction in American English. *Preprint submitted to Journal of Phonetics*.
Pierrehumbert, J. 2002. Word-specific phonetics. *Laboratory Phonology VII*, 101–140.
Poplack, Shana. 2006. How English Became African American English. In A. van Kemenade & B. Los (eds.), *The Handbook of the History of English*, 452–476. Oxford: Blackwell.
Pouplier, Marianne. 2012. The gaits of speech: Re-examining the role of articulatory effort in spoken language. In Maria-Josep Solé & Daniel Recasens (eds.), *The Initiation of Sound Change: Perception, Production and Social Factors*. Amsterdam: John Benjamins Publishing Company.
Purnell, Thomas C. 2010. Phonetic Detail in the Perception of Ethnic Varieties of US English. In Dennis Prestion & Nancy Niedzielski (eds.) *A Reader in Sociophonetics*, 289–326.
Purnell, Thomas, W. Idsardi & John Baugh. 1999. Perceptual and Phonetic Experiments on American English Dialect Identification. *Journal of Language and Social Psychology* 18(10), 10–30.
R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
Rácz, Péter. 2012. Operationalising salience: Definite article reduction in the North of England. *English Language and Linguistics* 16(1), 57–79.
Rácz, Péter. 2013. *Salience in Sociolinguistics. A Quantitative Approach*. Berlin: De Guyter.
Rankin, Robert L. 2008. The Comparative Method. Handbook of Historical Linguistics, 181–212.
Reetz, Henning & Allard Jongman. 2009. *Phonetics. Transcription, Production, Acoustics, and Perception*. Malden: Blackwell.
Rickford, John & MacKenzie Price. 2013. Girlz II women: Age-grading, language change and stylistic variation. *Journal of Sociolinguistics* 17(2), 143–179.
Rickford, John R. 2015. The Creole Origins Hypothesis. In Sonja Lanehart (ed.), *The Oxford Handbook of African American Language*, 35–56. New York: Oxford University Press.
Riley, M. W. 1971. Social gerontology and the age stratification of society. *The Gerontologist* 11(1), 79–87.
Robinson, Ryan. 2016. Top Ten Demographic Trends in Austin, Texas. https://austintexas.gov/page/top-ten-demographic-trends-austin-texas.
Romaine, Suzanne. 1995. *Bilingualism*. 2nd ed. Oxford: Blackwell.
Rose, Sharon. 2011. Long-distance Assimilation of Consonants. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, 1811–1837. Oxford: Wiley Blackwell.
Rose, Sharon & Rachel Walker. 2011. Harmony Systems. In Jason Riggle, John A. Goldsmith & Alan C. L. Yu (eds.) *The Handbook of Phonological Theory*. Second Edition, 240–290.

Rutter, Ben. 2011. Acoustic analysis of a sound change in progress: The consonant cluster /stɹ/ in English. *Journal of the International Phonetic Association* 41(01), 27–40.

Rutter, Ben. 2014. The acquisition of newly emerging sociophonetic variation: /str-/ in American English. *Journal of Child Language* 41(5), 1166–1178.

Sankoff, David & Suzanne Laberge. 1978. The linguistic market and the statistical explanation of variability. In David Sankoff (ed.), *Linguistic Variation: models and methods*. 239-250.

Sankoff, Gillian. 2006. Age: Apparent time and real time. In Brown, K. (ed.), *Encyclopedia of language and linguistics*. 2nd ed. Oxford: Elsevier. 110–116.

Sankoff, Gillian. 2006. Linguistic Outcomes of Language Contact. In Peter Trudgill, J. Chambers & N. Schilling-Estes, eds. *The Handbook of Language Variation and Change*, Malden: Blackwell, 638–668.

Santa Ana, Otto. 1993. Chicano English and the Nature of the Chicano Language Setting. Hispanic *Journal of Behavioral Sciences* 15(1), 3–35.

Santa Ana, Otto. 1996. Sonority and syllable structure in Chicano English. *Language Variation and Change* 8, 63–69.

Sapir, Edward. 1925. Sound Patterns in Language. *Language* 1(2), 37–51.

Saussure, Ferdinand de. 1978. *Course in general linguistics*. Charles Bally & Wade Baskin (eds.). Glasgow: Fontana.

Scudieri, Kirsten. 2012. On the Perception of /s/ and /ʃ/: Considering the Effects of Phonotactics. Masterthesis. The Ohio State University.

Semon, Richard. 1923. *Mnemic Psychology*. Bella Duffy (ed.). London: Allen and Unwin.

Shadle, Christine H., Michael I. Proctor & Khalil Iskarous. 2009. Revisiting the role of the sublingual cavity in the /s/-/ʃ/ distinction. *The Journal of the Acoustical Society of America* 125(4), 2569–2569.

Shadle, Christine H. & Celia Scully. 1995. An articulatory-acoustic-aerodynamic analysisof [s] in VCV sequences. *Journal of Phonetics* 23(1–2), 53–66.

Shapiro, Michael. 1995. A Case of Distant Assimilation: /str/ → /ʃtr/. American Speech 70(1), 101–107.

Shaw, Patricia. 1991. Consonant Harmony Systems: the Special Status of Coronal Harmony. In Carole Paradis & Jean-Francois Prunet (eds.), *The Special Status of Coronals: Internal and External Evidence*. Amsterdam: Elsevier. 125–157.

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language and Communication* 23(3–4), 193–229.

Solé, Maria Josep. 1992. Phonetic and Phonological Processes: The Case of Nasalization. *Language and Speech* 35(1), 29–43.

Stauffer, Howard B. 2007. *Contemporary Baysian and Frequentist Statistical Research Methods for Natural Research Scientists*. Hoboken: John Wiley & Sons.

Stevens, Kenneth & Helen M. Hanson. 2010. Articulatory – Acoustic Relations as the Basis of Distinctive Contrasts. In William J. Hardcastle John Laver & Fiona E. Gibbon (eds.) The Handbook of Phonetic Sciences: Second Edition. Malden: Blackwell. 424-453.

Stevens, Kenneth N. 1998. *Acoustic Phonetics*. Cambridge: The MIT Press.

Stevens, Mary & Jonathan Harrington. 2016. The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics* 58.118–134.

Stewart, William A. 1967. Sociolinguistic Factors in the History of American Negro Dialects. *The Florida FL Reporter*. Vol. 5.

Strand, Elizabeth A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18(1), 86–100.

Stuart-Smith, Jane, Morgan Sonderegger, Rachel Macdonald & Michael Mcauliffe. 2018. Cross-dialectal variation in /s/-retraction in English. BAAP.

Stuart-Smith, Jane, Morgan Sonderegger, Michael Mcauliffe & Rachel Macdonald. 2018. Dialectal and social factors affect the phonetic bases of English /s/-retraction. NWAV 47.

Tabain, Marija. 2001. Variability in Fricative Production and Spectra : Implications for the Hyper- and Hypo-. Language and Speech 44(1), 54–94.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice . *Language Variation and Change* 24. 135-178.

Tagliamonte, Sali & Alexandra D'Arcy. 2009. Peaks Beyond Phonology: Adolescence, Incrementation, and Language Change. Language 85(1), 58–108.

Tamminga, Meredith. 2014. Sound Change without Frequency Effects : Ramifications for Phonological Theory. *Proceedings of the 31st West Coast Conference on Formal Linguistics*, 457–465.

The X Factor UK. *Maria Ellinas sings Randy Crawford's Street Life* - Audition Week 2 - The X Factor UK 2014. 9/6/2014. https://youtu.be/NnbggCljAhk.

Thomas, Erik R. 2016. *Sociophonetics of Consonantal Variation*. Annual Review of Linguistics 2.

Thomas, Erik R & Holly Ericson. 2007. Intonational Distinctiveness of Mexican American English. *University of Pennsylvania Working Papers in Linguistics* 13(2), 1–15.

Thomas, Erik R, Norman J Lass & Jeannine Carpenter. 2010. Identification of African American Speech. In Dennis Preston & Nancy Niedielski(eds) *A Reader in Sociophonetics*, 265–285.

Thomason, Sarah & Terrence Kaufman. 1991. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.

Trudgill, Peter. 1972. Sex, covert prestige and linguistic change in the urban British English of Norwich. Language in Society 1(02). University Osnabrueck, 179.

Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.

Trudgill, Peter. 2006. Linguistic and Social Typology. In Peter Trudgill, J. Chambers & N. Schilling-Estes, eds. *The Handbook of Language Variation and Change*, Malden: Blackwell.707-728.

Twaddell, William Freeman. 1935. On Defining the Phoneme. *Language* 11(1), 5–62.

Välimaa-Blum, Riitta. 2009. Phonotactic constraints in cognitive phonology. *Corela* 7(1), 1–15.

Wang, William & Chin-Chuan Cheng. 1977. Implementation of phonological change: The Shuāng-fēng Chinese case. *Chicago Linguistic Society* 6, 552–559.

Weinreich, Uriel, William Labov & Marvin I. Herzog. 1968. Empirical Foundations for a Theory of Language Change. In W. P. Lehmann & Yakow Malkiel (eds.), *Directions for Historical Linguistics*. Austin: University of Texas Press. 95–188.

Whinnom, Keith. 1971. Linguistic Hybridization and the "Special Case" of Pidgins and Creoles. In Dell Hymes (ed.), Pidginization and Creolization of languages Cambridge: Cambridge University Press. 91–116.

Wickham, Hadley. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wilbanks, Eric. 2017. Social and Structural Constraints on a Phonetically-Motivated Change in Progress : (str) Retraction in Raleigh , NC. *University of Pennsylvania Working Papers in Linguistics* 23(1), 301-310.

Williams, Robert L. 1997. The Ebonics Controversy. Journal of Black Psychology 23(3), 208–214.

Winford, Donald. 2015. The Origins of African American Vernacular English: Beginnings. In Sonja Lanehart (ed.), *The Oxford Handbook of African American Language*. New York: Oxford University Press, 85–104.

Wolfram, Walt. 1974. The Relationship of White Southern Speech to Vernacular Black English. *Linguistic Society of America* 50(3), 498–527.

Wolfram, Walt. 1989. Structural variability in phonological development. Final nasals in Vernacular Black English. In Ralph W Fasold & Deborah Schiffrin (eds.), *Language Variation and Change*. Amsterdam: John Benjamins Publishing Company.

Wolfram, Walt. 2013. The contact dynamics of socioethnic varieties in North America. In Daniel Schreier & Marianne Hundt (eds.), *English as a Contact Language*. Cambridge: Cambridge University Press, 106–30.

Wolfram, Walt & Natalie Schilling. 2016. *American English. Dialects and Variation*. 3rd ed. Malden: Blackwell.

Wolfram, Walt & Erik R. Thomas. 2002. *The development of African American English*. Malden: Blackwell.

Wood, Sidney A.J. 1996. Assimilation or coarticulation? Evidence from the temporal co-ordination of tongue gestures for the palatalization of Bulgarian alveolar stops. *Journal of Phonetics* 24(1), 139–164.

Yu, Alan C L. 2013. *Origins of Sound Change: Approaches to Phonologization*. Oxford: Oxford University Press.

Zelinsky, Willbur. 2001. *The enigma of ethnicity: another American dilemma*. Iowa City: University of Iowa Press.

Zipf, George K. 1936. *The Psychobiology of Language*. Oxon: Routledge.

1999. *Handbook of the International Phonetic Association: a guide to the use of the international phonetic alphabet*. Cambridge: Cambridge University Press.

2019. *Demographics*. City of Austin. https://austintexas.gov/demographic