UNIVERSIDAD DE SEVILLA

FACULTAD DE FILOLOGÍA

DEPARTAMENTO DE LENGUA INGLESA

THE EFFECT OF L1 DIALECT

ON THE PERCEPTION OF PHONETIC VARIATION

IN L2 MORPHOLOGICAL MARKERS

TESIS DOCTORAL

María Del Saz Caracuel

Dirigida por: Prof[a]. Dra. María Heliodora Cuenca Villarín

Programa de Doctorado: Lengua y Lingüística Inglesas

Sevilla, 2013

"Long is the way

And hard, that out of Hell leads up to Light"

(*Paradise Lost*, John Milton)

Abstract

Second language (L2) speech perception depends on a number of factors related to the characteristics of the listener and to the characteristics of the stimuli employed in the experimental task. This doctoral dissertation explores the role that the dialect of the listeners (African American English and General American English) plays in the perception of two dialect variants of the same L2 (Western Andalusian Spanish and Castilian Spanish) in the morphological marker *–s*. Both General American English and Castilian Spanish use *–s* to mark the plurality of nouns and the person of verbs. African American English makes other uses of this marker, whereas Western Andalusian Spanish aspirates it. Our initial hypothesis proposes that the Castilian variant will be better identified than the Andalusian variant in general, although to a greater extent by General American English listeners. For this purpose, an identification task was designed with random sentences in which second-person and third-person verbs, as well as plural and singular nouns were embedded. Results corroborate our hypothesis and indicate that i) L2 proficiency level influences the perception of Andalusian aspiration; ii) the listener's dialect influences the perception of Castilian sibilants, iii) the perception of both variants depends on the phonetic context of the stimuli. A subsequent acoustic analysis of the stimuli reveals that there are intrinsic characteristics in both L2 dialects that can explain these results, especially as far as fricatives and stops are concerned. As future investigation, attention to (inter)dental contexts is suggested, as they present the most acute results.

Resumen

La percepción del habla de una segunda lengua (L2) depende de un número de factores relacionados con las características del oyente y con las características de los estímulos empleados en la prueba experimental. Esta tesis doctoral explora el papel que juega el dialecto de los oyentes (inglés afroamericano e inglés americano general) en la percepción de dos variedades dialectales de una misma L2 (español andaluz occidental y español castellano) en el marcador morfológico *–s*. Tanto el inglés americano general como el español castellano utilizan *–s* para marcar la pluralidad en los sustantivos y la persona de los verbos. El inglés afroamericano hace otros usos de este marcador, mientras que el español andaluz occidental lo aspira. Nuestra hipótesis de partida propone que la variante castellana será mejor identificada que la andaluza en general, aunque en mayor medida por parte de los oyentes de inglés americano general. Para ello se diseñó una prueba de identificación con frases aleatorias en la que se encontraban verbos de segunda y tercera persona así como nombres en plural y singular. Los resultados corroboran nuestra hipótesis e indican que i) el nivel de competencia en L2 influye la percepción de la aspiración andaluza, ii) el dialecto del oyente influye en la percepción de los sibilantes castellanos, iii) la percepción de ambas variantes depende del contexto fonético de los estímulos. Un posterior análisis acústico de los estímulos revela que existen características intrínsecas en los dos dialectos de L2 que pueden explicar estos resultados, especialmente en cuanto a fricativas y oclusivas se refiere. Como investigación futura, se sugiere prestar atención a los contextos (inter)dentales, ya que presentan los resultados más acusados.

Acknowledgments

I would like to thank the many people that have definitely helped me through this journey. First and foremost, my gratitude goes to my doctoral advisor, Dr. María Heliodora Cuenca Villarín, for her invaluable guidance, encouragement, and support through this ordeal. I would also like to thank Dr. J. Tamayo, Dr. J. Comesaña and Dr. F. Garrudo for their help in making this process continue in the right direction.

I am also grateful to M. A. Lamprea, K. Goldsmith, J. Sánchez Torres, M. J. Daily, M. Buchner, Dr. J. A. Bravo-de-Rueda, Dr. G. Cowell, E. Cruz Jiménez, C. Carrasco Llopis, Dr. M. Tubino Blanco, Dr. A. Ascunce, Dr. T. Lorenz, Dr. C. Francom and Dr. I. Ortega-Santos for their help with participant recruitment, and to Dr. L.B. Schmidt for allowing me to know more about her work. Without them and without the participants and the speakers, this dissertation would have not been possible.

I cannot fail to acknowledge the efficient and prompt work of P. Anaya, the computer technician who mounted the experimental task, as well as the invaluable help and expertise of M. Barrio in all matters related to Acoustic Phonetics.

And last, but not least, I want to thank my family and friends for their patience, support, and understanding; especially to Dad and Nacho for those comments that gave me an outside and many times enlightening perspective on this project.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

When second-language (L2) learners encounter the sounds of the L2 in their initial stages of learning, they tend to perceive these sounds in terms of the specific characteristics of their first language (L1). For example, L1 Spanish listeners interpret L2 English vowels in terms of their limited five-vowel repertoire (Fox, Flege, & Munro, 1994), which means that English /æ, ɑː, ʌ/ tend to be identified and produced as Spanish /a/. Nevertheless, as listeners gain experience with the L2 and become more proficient in the language, they start discriminating L2 phonetic categories that are not contrastive or present in their L1.

An additional factor to take into account in L2 speech perception is related to dialectal or sociophonetic variation in the L2. An L2 phonetic category that has different phonetic realizations dependent on L2 dialect poses an additional difficulty for L2 learners. For instance, such is the case of British English intrusive /ɾ/ (Tuinman, Mitterer, & Cutler, 2007) or the aspiration of syllable- and word-final /s/ in several Spanish dialects (Schmidt, 2011). When stimuli are embedded in sentences and these variations are the result of connected speech processes, we can expect that "connected speech processes pose similar problems in learning a second language as new phonemes do. Processes that are unique to the L2 … lead to major perceptual problems" (Mitterer & Tuinman, 2012, p. 12).

Nevertheless, L2 speech perception is not only dependent on the characteristics of the L2 dialect in question, but also on the L1 dialect of the listeners involved (Best & Tyler, 2007). Any variation in the listeners' L1 dialects may render different perceptual patterns (Celata, 2007; Chladková & Escudero, 2012; Escudero & Williams, 2012).

Along these lines, this dissertation investigates the cross-dialectal perception of two L2 dialect variants of the same morphological marker by listeners of two L1 dialects. Specifically, it explores how African American English (AAE) and General American English (GAE) listeners, L2 learners of Spanish, perceive [h] as a dialectal variation of the morphological marker /s/, to which they have not been exposed and which is not a possible realization of implosive[1] /s/ in their L1 dialect.

Aspiration of implosive /s/ is found in southern Spain, the Canary Islands, the Caribbean and Pacific countries, but not in Central America or in central and northern areas of Spain, which generally retain sibilance. Given the characteristics of the two L1 dialects included in this study and, as no other study explored the perception of an L2 by AAE listeners before, this dissertation also aims to explore their perception of the mainstream Spanish variant [s] for comparison, which also functions as a plurality and agreement marker in GAE but may not function as such in AAE.

For this purpose, we devised an experiment in which participants were presented with randomized sentences that contained singular nouns, plural nouns, second-person verbs and third-person verbs from Western Andalusian Spanish (WAS), which uses aspiration, and Castilian Spanish[2] (CS), which retains sibilance. In Spanish, both plural nouns and second-person verbs carry word-final /s/, while singular nouns and third-person verbs end in a vowel sound [V]. Two forced-choice written options were given per sentence: the actual sentence spoken and the version with or without the final *–s* in the target word.

[1] Syllable-final or word-final position.
[2] The term Castilian Spanish is employed here to label the mainstream dialect in Spain.

We chose this type of task because it resembles what native speakers of a language do when decoding the acoustic signal they receive from continuous L1 speech, requiring the identification and categorization of phonetic segments according to their internalized language-specific categories to access meaning (Hawkins, 2011; Strange & Shafer, 2008). Studies have shown that experiments that address basic auditory capabilities and trigger language-general patterns of perception yield similar results for native, naïve, and experienced L2 listeners. As the cognitive demands of the task and the stimuli increase, language-specific patterns of perception are more likely to be reflected. We believe this may be especially relevant for elementary students, which have a more limited experience with the L2; thus, we also explored the proficiency level of the listeners when interpreting the results of the tests.

The organization of the rest of this dissertation is as follows: Chapter 1 provides a detailed description of the phonology of AAE and WAS, and a specific description of [h] and [s] in Spanish and English. Chapter 2 is devoted to acoustic phonetics and, in particular, to the description of English and Spanish sounds in terms of their acoustic properties. Chapter 3 reviews the background literature pertinent to this study and poses predictions for the two groups of listeners as well as our research questions. Chapter 4 provides an account of the methods employed in the experiments, a description of the stimuli used, and a report of the participants who took part in this study. Furthermore, descriptions of the experimental task as well as the statistical and acoustic analyses carried out are also provided. In Chapter 5, we report the main findings obtained from the experiments; first, the results of the pilot test and, second, the results of the current experiment in terms of overall performance, identification patterns by individual group, and identification patterns across groups of listeners. Additionally, we analyze the results according to the syntactic and the phonetic contexts in which stimuli are found, and we also provide an acoustic analysis of key stimuli to incorporate these findings to our discussion section. Finally, in Chapter 6, we discuss these findings, the implications of the results for L2 speech perception theory, and the limitations of this study with suggestions for future research.

CHAPTER 1

DESCRIPTION OF L1 AND L2 DIALECTS

In order to understand how AAE and WAS differ from mainstream GAE and CS, respectively, this chapter provides a detailed description of the phonological systems of both dialects. Additionally, complementary information concerning grammar use can be found in Appendices A and B.

1.1 L1 Dialect: African American English

The term AAE is generally employed to refer to the language varieties that African American people speak in the United States. However, as Baugh (2004b) points out, African Americans can fall into any of these three categories: a) GAE is their native language, b) GAE is not their native language, c) their native language is different than English. Most speakers of AAE belong to the second category, with the ability to switch from GAE to nonstandard English or AAE depending on the context and their interlocutors. We should also take into account that not all African Americans generally speak AAE nor are all speakers of AAE African Americans. As Green (2004a) describes it:

> African American English refers to a linguistic system of communication governed by well defined rules and used by some African Americans (though not all) across different geographical regions of the USA and across a full range of age groups. While AAE shares many features with mainstream varieties and other varieties of English, it also differs from them in systematic ways. (p. 77)

Ever since the early studies in the 1960s, AAE has seen a number of names applied to it, depending on the term employed to address African Americans at the time. “Negro dialect” and “Negro speech” were usual in the 1960s, “Black talk”, “Black dialect”, “Black English” and “Black Vernacular English” were popular in the 1970s, which then turned into “African American language” in the mid-1980s. “African American English”[3] became the preferred term in the 1990s, along with “African American Vernacular English” to design the nonstandard form of AAE. The term “Ebonics” (a blend of *ebony* and *phonics*) was initially employed to refer to the speech of those African Americans of West African descent but subsequently used as equivalent to those terms above, especially AAE. In any case, “the reality, however, is that most speakers of what is identified here as AAE do not have a name for their vernacular. Generally they say they speak English” (Mufwene, 2001, p.293).

AAE has been the subject of much controversy, especially in education, since the Oakland School Board resolution in 1996 and subsequent resolution by the Linguistic Society of America in 1997, which recognized Ebonics as a systematic and rule-governed linguistic system, not related to English, and the primary language of African Americans. This resolution aimed at improving their proficiency in GAE and thus broadening their academic and professional future.

However, there is still no full agreement as to whether it is considered a dialect of English or a separate language. On the one hand, apart from having its own distinctive characteristics, AAE shares the vast majority of its features and patterns with GAE, which would support the first position. One of the authors who support the term *dialect* is Dillard (1993), stating that it is “the first clearly discernible and reportable dialect of American English” (p. 60). On the other hand, AAE involves

[3] There is still a difference of opinions among AAE speakers. Some prefer the term Black, “we have been here too long […] By now we have no African in us”. Others prefer to use African American to highlight “our origin and cultural identity.” (Smitherman, 1998, pp. 206-7)

sociological and ethnic connotations, as well as a unique background and development, which would account for the second position. This view is shared by authors such as Smitherman (1999), who states that "it is a language forged in the crucible of enslavement, US-style apartheid, and the struggle to survive and thrive in the face of domination" (p. 19). A third position invalidated by linguists and experts but still present in society, even among its own speakers, is that AAE is simply "bad English".

The origin of AAE is also a controversial issue, giving rise to three main views. The Anglicist hypothesis emerged in the mid-20th century and defends that AAE originated from the various dialects of English that white immigrants from the British Isles spoke at the time. Later, towards 1970s, the Creolist hypothesis appeared in defense of the view that AAE may have started as a creole language such as Gullah or Jamaican Creole, with which it shares features, influenced by the languages of the slaves brought from other colonies. Therefore, contact with other dialects in the USA would have originated a slow process of decreolization, by which AAE is converging with other varieties of English. Finally, the Africanist hypothesis defends that AAE is similar to West African languages in structure and regards any similarity to English as only superficial. Even when it may have incorporated English features, the substrate influence of West African languages is still preserved. Other than these major views, there is a new one named Neo-Anglicist hypothesis that also believes that AAE originated from British dialects but has undergone a unique evolution that has made it diverge from GAE. We may never know how AAE exactly originated, given the scarce recordings and data of which linguists dispose. As Wolfram (2006) states:

> Current evidence suggests more regional influence from English speakers than assumed under the Creolist Hypothesis and more durable effects from early language contact situations than assumed under the Anglicist positions, but the issue of regional accommodation and substrate influence continues to be debated. (p. 335)

The term African American Vernacular English (AAVE) must be distinguished as the vernacular or nonstandard form of AAE which carries more stigmatized aspects, used mainly for everyday communication among its speakers. It is generally attributed to the working class, although the middle class can also use it depending on the context, e.g., informal situations, adding emphasis, expressing ethnic solidarity, etc. While it is true that it shares features with creole languages and Southern White Vernacular English (SWVE), AAVE is still a systematic and rule-governed linguistic system with defined aspectual grammar, vocabulary of its own, distinctive phonology, and unique intonation which divert from GAE. At the same time, AAVE should not be regarded as mere slang, as slang refers to temporary vocabulary and expressions which grow out of fashion and are replaced by others with time. AAVE features are long-established and common throughout the country.

Nowadays, research shows a tendency for different trajectories in the development of AAVE according to geographical and other sociological factors. There are instances of assimilation of the regional variety of English and reduction of AAVE characteristics, as well as instances in which AAVE characteristics are reinforced and resistance to the regional variety of English takes place. "Original settlement history, community size, local and extra-local social networks, and racial ideologies in American society must all be considered in understanding the course of change in African American speech" (Wolfram, 2006, p. 340).

1.1.1 The Phonology of African American English

At first glance, the most distinctive features of AAE seem to lie in its morphology and syntax[4]. This has led to a great amount of research directed towards the origins of this variety and its implications for education. "In many ways phonology is the neglected stepchild of research on AAE. Even the most cursory review of the literature will show that morphology and syntax have long been the primary focus of work on AAE" (Bailey & Thomas, 1998, p. 85).

The phonology of AAE presents different types of variables, the majority of which are systematic and context-dependent. This does not imply that all African Americans always use all of them; there is variation among these variables that are the most salient patterns in AAE.

1.1.1.1 Consonant clusters

Consonant cluster reduction, especially when the second consonant is a stop, is well-known among AAE speakers. Even when this feature is common to other varieties of English, certain constraints in which it occurs seem to differ. The reduction is generally more likely to take place when the following word begins with a consonant sound (*fast car*) than when it begins with a vowel sound (*cold air*) or when the second consonant in the cluster is a morpheme, such as past tense *–ed* (*talked*). These phonological and grammatical constraints are found in all varieties of English.

What is interesting is that phonological constraints seem to dictate cluster reduction in GAE, while AAE is more driven towards respecting grammatical constraints. In other words, AAE speakers are less likely to simplify the cluster when

[4] For a description of the most salient grammatical features, see Appendix A.

it represents a morpheme, whether followed by a consonant or a vowel sound. However, there seems to be an exception to this rule and this is the case of irregular part tense verbs (*kept*), which are more likely to suffer reduction than regular past tense verbs, probably because the tense is additionally marked by a change in the vowel sound. Since utterances do not occur in isolation but within a context, it is possible for speakers to employ cluster reduction in past tense verbs when there are other clues of time in the sentence (*Yesterday, she call me three times).*

The number one rule for consonant cluster reduction is that both consonant sounds must share voicing (*fast, kind*), but in spite of this rule we can also find an exception, and this is negative auxiliary verbs. It is common to hear *can't* realized as [ˈkeɪn] and *don't* uttered as [ˈdoʊn].

Wolfram and Thomas (2002, pp. 133-4) enumerate a list of constraints that affect the frequency of this type of reduction. First, simplification is less likely when both consonants are stops (*pact*) than when the first one is a sibilant (*past*). Second, it is less likely when the first one is a sibilant (*past*) than when the first one is /l/ (*bold*). Third, it is less likely when the first one is /l/ (*bold*) than when the first one is nasal (*kind*). And fourth, consonant cluster reduction is more commonly found in unstressed syllables than in stressed syllables.

There seems to be opposing views about the origins of consonant cluster reduction. On the one hand, the reduction is believed to be a process that occurs according to the phonological context in which the cluster is given, as is the case in other varieties of English such as nonstandard British accents. On the other hand, this feature is attributed to the influence of West African languages, which do not allow final consonant clusters (Green, 2002b). In fact, there are speakers who actually do not seem to have a cognitive representation of the cluster; therefore, it is possible to find the plural form of these reduced words as if indeed the cluster never existed, giving way to *test* > [ˈtes], the plural form of which would be [ˈtesəz], as in *buses* (Green, 2002a, 2002b; Mufwene, 2001). Furthermore, a more common plural form would be realized by lengthening the continuant as in *tests* > [ˈtes:] (Thomas, 2007).

Additional features involving consonant clusters are metathesis and the backing of */str/* clusters. Metathesis consists of switching the position of the consonants in the cluster, whose main representative example is *ask* > [ˈæks]. Backing of /str/ cluster means that the cluster is realized as [skr], especially before high front vowels, as in *street* > [ˈskri:t].

1.1.1.2 Fricatives

A second feature attributed to AAE that is one of its most representative characteristics involves the absence of interdental fricatives /θ/ and /ð/, which are either labialized or stopped. The former phoneme is usually replaced by [t] in initial position (*think* > [ˈtɪŋk]) and final position (*month* > [ˈmʌnt]) or by [f] in final position (*both* > [ˈboʊf]), while the latter phoneme is replaced by [d] in initial position (*this* > [ˈdɪs]) and by [v] in medial position (*mother* > [ˈmʌvə]) and final position (*bathe* > [ˈbeɪv]). This feature is also found in other nonstandard varieties of English; nevertheless, it is much more commonly found in AAE and inversely correlated with social class and formality of speaking style.

The constraints on this feature are somewhat unclear, as we can find the word *with* uttered in all four different ways: [ˈwɪt], [ˈwɪd], [ˈwɪf], and [ˈwɪv], mostly depending on the voicing of the following sound (Bailey & Thomas, 1998). One point should be made here: AAE speakers know how to realize interdental fricatives, as in *thing* > [ˈθɪŋ]. The alternative realizations are seen by Africanists as a West African influence, whose languages do not include /θ/ or /ð/; by contrast, Anglicists claim that nonstandard dialects of British English also included initial [d] and final [f] for /θ/. Likewise, Creolists state that /θ/ realized as [t] or [d] is also a feature in creole and pidgin languages (Rickford & Rickford, 2000).

Another case of fricative stopping takes place especially in medial position before nasal sounds. In this instance, it is the substitution of [b] for /v/, as in *seven* > [ˈsebm], and the substitution of [d] for /s/, as we can see in *isn't* > [ˈidnt] (Rickford, 1999; Rickford & Rickford, 2000; Bailey, 2001). The first characteristic finds a similar phenomenon in creole languages, which realize /v/ as a bilabial continuant [β] (Lerer, 2007).

1.1.1.3 R-lessness and l-lessness

The following features used to be shared by both AAE and SWVE at the beginning of the 20th century; however, it is reversing for the latter while it seems to persevere with the former. It is the case of what is known as 'r-lessness' or non-rhoticity, i.e., the deletion or vocalization of constricted /r/ in any of the following phonetic environments: (1) postvocalic position (*four*), (2) word-medial position (*carry*), (3) unstressed syllable (*mother*), (4) and stressed syllable (*work*) –although deletion in this last case is mostly restricted to Southern AAE.

(1) *four* > [ˈfoʊ], [ˈfoə], [ˈfo:]

(2) *carry* > [ˈkæi]

(3) *mother* > [ˈmʌvə]

(4) *work* > [ˈwɜ:ⁱk], [ˈwɜ:rk]

Mufwene (2001) explains the frequency of occurrence of /r/ deletion or non-rhotic /r/ according to its position. The most frequent cases of deletion take place in word final and pre-consonantal positions, next in frequency we find word final position followed by vowel sound, and the least frequent deletion occurs word-medially between vowel sounds. In this last case, deletion is prohibited if the preceding vowel sound belongs to a prefix (Green, 2002b).

The absence of linking *r* is also related to this phenomenon, but it is not the only linking element that can be omitted. Linking glides /j/ and /w/ found in GAE connected speech do not occur in AAE. This phonological phenomenon is so embedded that it transfers to the realization of the definite article *the* as [ˈdə] before vowels sounds, rather than [ˈdi]. Also before words starting with a vowel sound, the indefinite article *a* is often preferred to *an* (Mufwene, 2001).

Deletion of /r/ also occurs after the voiceless interdental fricative /θ/, as in *throw*, and other consonants in unstressed position, as in *prefer*. Likewise, vocalization of /r/ can occur in the cluster /ʃr/, resulting in [ʃw] or even [sw], as we can see in *shrimp* > [ˈʃwɪmp] or [ˈswɪmp]. In any case, this feature, as many others, seems to decrease in frequency as formality and social level increase (Thomas, 2007).

A similar phenomenon, known as ‘l-lessness’, involves the deletion or vocalization of /l/ in syllable-final position. While vocalization to [ʊ] is found in both GAE and AAE, vocalization to [ə] is strongly attributed to the latter, as in *feel* > [ˈfiːə]. Deletion of /l/ is much more common in AAE than in GAE, and it tends to occur after rounded vowels; however, in the South, it is also given before labial consonants (*twelve* > [ˈtwev]) or in *–self* compounds (*myself* > [maːˈsef]). This rule is extended beyond word boundaries, an example of which is the reduction of the contracted form of *will*, as in *she*[ə] *be here in a minute*.

While both ‘r-lessness’ and ‘l-lessness’ can occur in several varieties of English, with either vocalization or deletion of /r/ and /l/, in the case of deletion we can find a lengthening of the previous vowel in AAE, as in *cold* > [ˈkoʊ:] (Green, 2002b). Notice that, in this case, consonant cluster reduction seems to take place first. Rickford & Rickford (2000) mention that these two phenomena may come from African influence, although it is unclear. This could be related to the affirmation made by Dillard (1993) that the syllable structure in West African languages tends to be CVCV.

1.1.1.4 Consonant deletion

Modifications to other consonant sounds tend to occur especially in, but not limited to, final positions. It is common to delete a consonant in final position when it is preceded by a vowel sound and followed by a consonant, as is the case of *cat* > [ˈkæ]. For this reason, if the consonant eliminated is a nasal sound, nasality is transferred to the preceding vowel, as we can see in *man* > [ˈmæ̃]. It is also very common, and apparently unique to AAE, to devoice final voiced stops, as in *pig* > [ˈpɪk], especially before vowels and pauses, and to insert a glottal stop whether it is deletion or devoicing that occurs, example of which is *bad* > [ˈbæʔt] or [ˈbæʔ].

Further consonant modification occurs in the case of auxiliary verbs and gerunds. AAE speakers usually delete initial /d/ and /g/ in auxiliary verbs, giving way to reduced versions of these, as is the case of *didn't* > *ain't*. This characteristic seems to be unique to AAE and it may indicate an influence from creole languages (Rickford & Rickford, 2000). Final /ŋ/ sound in gerunds is commonly develarized to [n], as in *talking* > *talkin'*, and it also happens in words such as *something* and *nothing*. However, other than [ˈsʌmθɪn] and [ˈnʌθɪn], usual realizations of these two words are [ˈsʌmpm] and [ˈnʌʔn]. Green (2002b) points out that develarization is restricted to words with more than one syllable; therefore, *sing* could not be realized as *sin*. This phenomenon is common to other nonstandard varieties of English and GAE in unstressed syllables.

1.1.1.5 Vowels and diphthongs

The AAE vowel system shares the same vowel sounds as GAE and other varieties of English; therefore, at first glance there is nothing remarkable in this regard. The main difference lies in what is known as the organization of vowel space. Bailey & Thomas (1998) explain that there are three major patterns to which American dialects adhere: the Northern Cities Chain Shift, the Southern Shift, and the Low Back Vowel Merger, none of which AAE pertains to.

These two authors describe the development of the AAE vowel system in relation to creole languages and, as it progressed, to SWVE. While some of the early features that linked AAE to creole languages have disappeared, a number of them still prevail to this day. Additionally, from 1875 to 1940 AAE underwent a series of innovations, some of which are common to SWVE, while others remain unique for AAE. Recent innovations that appeared in SWVE, such as the Southern Shift, are not shared by AAE, which deepens the differences between both varieties and raises the current question of divergence. "The AAE vowel system, then, suggests a history marked by unique origins, shared history, and independent development – the same kinds of things that characterize the histories of most languages" (Bailey & Thomas, 1998, pp. 106-7).

As mentioned at the beginning of this section, in the AAE vowel system the vast majority of the variants are systematic, with a few cases of lexical variants. First of all, we will focus on the first group.

Before 1860, AAE was characterized by a set of vowel phonemes that did not occur in SWVE but have parallels with creole languages and West African languages. These were: (1) the realization of /eɪ/ and /oʊ/ as monophthongs, (2) fully back vowels /u:/ and /ʊ/, (3) non-fronted onsets of /aʊ/, and (4) full diphthong /aɪ/ before voiced obstruents. By the end of the century, the monophthongal realizations of /eɪ/ and /oʊ/ were replaced by diphthongs; however, the non-fronted onsets of /aʊ/ and the fully back vowels are maintained. At this time, we start to see the shortening of the glide /aɪ/ before voiced obstruents, which is common to SWVE as well. Nevertheless, a new development at this time that only pertains to AAE was the raising of /æ/ to mid-front position. Even when this last feature is also part of the Northern Cities Chain Shift, we should deem it unrelated to AAE since both varieties had no connection or interaction during this period.

Also towards the last years of the century, SWVE started to develop fronted realizations of /u:/ and /ʊ/ to central position and fronted onsets of /aɪ/, phenomena which AAE resisted. If anything, what we find is /aɪ/ realized as monophthongized [a:] in Southern AAE.

Around the last 25 years of the century, we find the beginning of a set of conditioned mergers which are common to both AAE and SWVE: the merger of /e/ and /ɪ/ before nasal consonants, which leads to the realization of *pen* = *pin* > [ˈpɪn], the merger of tense and lax vowels before /l/, as in *feel* = *fill* > [ˈfi:l], *bale* = *bell* > [ˈbel], *pool* = *pull* > [ˈpu:l], and the merger of /ɔ:/ and /oʊ/ before /r/, giving way to *horse* = *hoarse* > [hɔ:(r)s]. Labov, Ash, and Boberg (2006) point out that there seems to be individuals unable to make a clear distinction in both production and perception of some of these pairs of phonemes.

Nevertheless, SWVE continued to develop with the Southern Shift as the reorganization of its vowel space progressed rapidly and away from AAE. We find features such as the realization of the onset in /eɪ/ as low as [æ], the realization of /u:/ and /ʊ/ as front rounded vowels, and the centralization and lowering of the onset in /oʊ/. Additionally, the transfer of the glide-shortened /aɪ/ to voiceless environments and the merger of /ɔ:/ and /ɑ:/ set SWVE further apart from AAE.

These differences between AAE and SWVE reflect the unique origins of the former, as well as the innovations in the latter that did not transfer to AAE and vice versa. After having looked at the path that AAE followed and at some of its features, we now proceed to list further features that characterize this variety at the present time. Some of these features will be shared with other varieties of English, although still much more frequently found in AAE, and some others will be considered unique to this variety.

We have seen that monophthongal pronunciation of /aɪ/ is common before voiced obstruents, and this also extends to /ɔɪ/ before /l/ and /r/ in SWVE. While AAE does not embrace this second phenomenon, it does lower the glide to [ɔə], as in *boil* > [ˈbɔəl] (Thomas, 2007). There seems to be some important disagreement about this feature, as Mufwene (2001) gathers that, not only do AAE speakers frequently monophthongize /aɪ/ and /ɔɪ/, but also /aʊ/ and /oʊ/, or at least they realize a very weak glide.

Other than the raising of /æ/ to [e] in isolation, as in *bad* > [ˈbed], we can also find instances of a slight diphthongization of /æ/ to [æ^{j}], exemplified by *hand* > [ˈhæ^{j}n] (Mufwene, 2001). Additionally, it is also very common for AAE speakers to raise /e/ and /ɪ/ as well, as we can see in *get* > [ˈgɪt] and *did* > [ˈdi:d], as part of what Thomas (2007) defines as the African American Shift.

We have also seen the merger of /e/ and /ɪ/ before nasal sounds; moreover, we can also find the lowering of /ɪ/ to [æ] before the velar nasal sound, as is the case of *thing* > [ˈθæŋ] and *stink* > [ˈstæŋk] (Mufwene, 2001; Rickford, 1999).

Also, we have learned that AAE resists to merge /ɔ:/ and /ɑ:/; nevertheless, there exists the case of an unrounded pronunciation of /ɔ:/, giving way to [ɑɒ], as in *thought* > [ˈθɑɒʔt]. Additionally, there are instances of backing and rounding /ɑ:(r)/, a phenomenon that is common to both varieties of English, giving way to [ɒ:(r)] as we can see in *start* > [ˈstɒ:(r)ʔt] (Thomas, 2007).

Finally, a relatively new development in AAE mentioned by Pollock (2001) is the centralization of /er/ and /ɪr/ to [ɜ:r], as in *here* > [ˈhɜ:r], with the possibility of additional insertion of schwa off-glides, as in *chair* > [ˈtʃɜ:ə].

Apart from the systematic variables described so far, there are certain lexical-specific vowel variants which are worth mentioning. First of all, the realization of the vowel in *can't* is usually [eɪ], as in *can't* > [ˈkeɪn] or [ˈkæʲn]. Also noticeable is the pronunciation of *aunt* generally as [ˈɑ:nt] rather than [ˈænt], which would be much more common in GAE. This probably began as prestige pronunciation used by plantation owners and then transferred to slaves. Finally, we find the realization of /ɪ/ as /ʊ/ before an unstressed syllable (*sister* > [ˈsʊstə]), which began with the retraction of the former sound and eventually developed into the latter.

The following characteristics apply to syllables rather than single phonemes. It is common for AAE speakers to delete reduplicated syllables, a phenomenon called haplology, as in the case of *Mississippi*. They also tend to delete initial and medial unstressed syllables, with much more incidence if the syllable is composed of a vowel sound only, as in *again* > *'gin*. Finally, it is also usual to hear stress on the first rather than the second syllable of a set of given words, such as *police* and *hotel*.

We have previously seen how the pronunciation of certain auxiliary verbs are affected in AAE and now we will pay attention to the pronunciation of two aspectual markers[5] of AAE, which share their forms with two GAE words but which differ in pronunciation (and, of course, in meaning). These are (1) *remote been* (written *BIN* due to its stressed pronunciation and not to be confused with GAE *been*) and (2) *completive done* (written *dən* due to its unstressed pronunciation and not to be confused with GAE *done*). The former is employed to indicate that something occurred a long time ago or that something has been happening for a long time to this day. The latter refers to the completion of an action with present results.

(1) *She BIN ate all the candy.*

GAE: She ate all the candy a long time ago.

(2) *I dən did my homework.*

GAE: I have done my homework.

Rickford (1975) conducted an experiment in which one half of the informants were AAE speakers and the other half were GAE speakers. He presented them with different sentences in which these two aspectual markers were present and tested their understanding of their meanings. All AAE speakers obtained correct answers while only one GAE speaker answered all questions correctly. This example is to give an understanding of how this variety is ruled-governed and forms a well-designed system in some aspects different than GAE. As Rickford & Rickford (2000) put it, “these processes are highly systematic, and not the careless or haphazard pronunciations that observers often mistake them for” (p. 104).

[5] See Appendix A

1.2 L2 Dialect: Andalusian Spanish

Andalusian Spanish (AS) is the variety of the Spanish language spoken by people from and in the province of Andalusia, the southern region of Spain. It appeared as the result of the changes in the Medieval Castilian taken to the region with the Reconquest by Ferdinand III The Saint in the 13th century. Record of these changes dates back to the 15th and 16th centuries and indicates that the variety was consolidated in the 18th century. Whether it is a dialect or a variety of speech still remains undetermined; diachronically, it is a dialect which evolved from the historical Castilian brought to the region by settlers and colonizers around the 13th century; synchronically, it is a linguistic variety of Spanish as other regional varieties are, which took form from elements of other dialects in the Iberian Peninsula and the influence of foreign languages. Additionally, there is a minority of researchers who defend that the origin of AS is not entirely Castilian but a pidgin language with a Castilian-based lexicon and morphosyntax combined Mozarabic features, a view to which Narbona, Cano, and Morillo (1998), Jiménez Fernández (1999), and Cano Aguilar and González Cantos (2000) oppose. Finally, there are a number of claims that AS is simply "bad Spanish", even among its own speakers. In this study, we will abide by Alvar's (2006) definition:

> *Precisamente, diferencias e historia me hacen ver el andaluz como un dialecto y no aceptar que me digan que la «manera de hablar» una lengua es –así, sin más- "el sentido vulgar del término [dialecto], no el técnico" ... pues buen cuidado he tenido siempre en no confundir «la comprensión de un habla y el metalenguaje de una ciencia.»* (p. 13)
>
> [Precisely, differences and history make me see Andalusian as a dialect and not accept to be told that "the way of speaking" a language is –just like that- "the vulgar sense of the term [dialect], not the technical" … since I have always been very careful not to mistake «the understanding of speech and the metalanguage of a science.»]

In Europe, unlike in America, it is usual to find dialects which are contemporary and even more ancient than the standard language of a country; therefore, its divergence from the norm must not be seen as simplifications of the standard language. In this case, the variation which presents more prestige and is used as the norm is northern Spanish, pushing southern Spanish to the background with a different social acceptance. The Castilian spoken in the Reign of Toledo rose as the standard variety of the language; thus termed Spanish, which was spread to Europe. The variety spoken in the Reign of Seville (Seville, Huelva, Cadiz), subsequently named Andalusian, was the norm transferred to the Canary Islands and South America.

1.2.1 The Phonology of Andalusian Spanish

As we will see, none of the phonetic and phonological characteristics of AS is common to all speakers in the area nor are they all exclusive of Andalusia. AS also presents a faster and more varied rhythm than CS and some of its phonemes are realized in a more lax way, while others are uttered in a more tense way, than CS. In this section, we will see some characteristics which are spread all over the region, other characteristics that are less spread but still present, and some characteristics that are also found in other Spanish varieties but very commonly used in AS.

1.2.1.1 Andalusian /s/

One of the best-known features of AS is its /s/ realizations and the linguistic phenomena concerning this phoneme. Spanish /s/ is realized by placing the tip of the tongue in the alveolar region of the mouth with the tongue in a concave position. Andalusian /s/ has several realizations, usually with the tongue in a flat position –as the [s̄] in Cordova- or in a convex position –as the [ṣ] in Seville. The manner of articulation is dental in these two cases, with the actual blade of the tongue and not the tip touching the teeth.

In Andalusia, more than a third of the speakers make a distinction between /s/ and /θ/, about the same number merges these two sounds into a dental [s] (phenomenon called *seseo*), while less than a third merge both sounds into an interdental [θ] (phenomenon called *ceceo*), as in *poso* = *pozo*, *casa* = *caza*. The distinction between these two sounds is seeing a widespread tendency nowadays, particularly among young and educated speakers, partly favored by the media and the more accepted peninsular norm. It is mostly given in northern and eastern regions of Andalusia while, in the rest of the regions, there tends to be a coexistence of *seseo* and *ceceo*. *Ceceo* is considered as low status and is a minority phenomenon due to the presence of the distinction of sounds and *seseo* in urban areas and in the media.

In speakers of low socioeconomic status, *ceceo* can become *heheo*, especially in rural areas. This means /s/ and /θ/ are uttered with a retracted position of the tongue in a relaxed and aspirated manner, such as [h]. We can also find *ceseo* and *seceo*, especially among those speakers who do not make a distinction between /s/ and /θ/. This means speakers realize those sounds as one or the other without a clear pattern, as in *cerveza* > [θerβésa] or [serβéθa].

1.2.1.2 Aspiration

In relation to the Andalusian /s/ realizations, the most characteristic feature of AS and the most widespread to other varieties of southern Spanish is the realizations of the syllable-final and word-final /s/ (implosive /s/). Being uttered with less articulatory force, as is the case of all Spanish final consonants, it can either be maintained or derive into aspiration, assimilation to the following consonant, gemination, or deletion. This depends on the context, whether /s/ is placed before consonant, vowel or pause, and on geography and socioeconomic status.

Aspiration of implosive /s/ occurs before any consonant sound in all geographical areas and social levels, and it is also characteristic of the varieties in the Canary Islands and South America. Reduplication or gemination of following consonants is the general tendency in informal and spontaneous situations.

Before voiced stops /b, d, g/, educated speakers aspirate implosive */s/* without modification of the following consonant sound; however, in vernacular speech aspiration can be transferred to those consonant sounds, turning them into fricatives [f], [v], [θ], [ð] and [x]. Alvar (1996, p. 243) gives examples of the possible realizations and allophones derived from each phoneme:

I) *–s + b: lah brujah, lab bragah, lav viñah, lo brimbe, muncho fohqueh* (= 'las brujas, las bragas, las viñas, los brimbes, muchos bosques')[6].

II) *–s + d: loh dienteh, buenoð ðía, uno θeoh* (= 'los dientes, buenos días, unos dedos,').

III) *–s + g: lah gatah, log güebo, loj jabilane, la jraná* (= 'las gatas, los huevos, los gavilanes, las granadas').[7]

Before voiceless stops /p, t, k/, aspiration of implosive /s/ also occurs without modification to the following consonants in careful speech. However, in vernacular speech aspiration derives into reduplication or gemination of following consonants. Speakers in Cordova and certain areas of Granada and Seville can infuse aspiration to /p, t, k/ (Gerfen, 2002; Torreira, 2007a, 2007b, 2012), and the latter sound can also be

[6] Jiménez Fernández (1999) also adds labio-dentalization (resbalar > refvalá) to this list.
[7] Jiménez Fernández (1999) also indicates complete assimilation (rasgo > ráho) as possible.

uttered with aspiration in the whole region, especially before *–ié*. This feature is omitted in formal situations. Alvar (1996, p. 243) exemplifies:

I) *loh pieh, doh toroh, lah casah;*

IIa) *lo*^{hp} *pieh, do*^{ht} *toroh, la*^{hk} *casah;*

IIb) *lo*^{p} *pieh, do*^{t} *toroh, la*^{k} *casah;*

III) *lo pieh, do toroh, la casah;*

A more recent variant, which is the affricate palatal pronunciation of /t/, occurs when aspirated /s/ before this dental phoneme influences its articulation into [t͡s]. This phenomenon is given in all areas where final /s/ is aspirated or dropped and it is more frequent among the young population and mid-class to high-class speakers (Moya Corral, 2007; Ruch, 2008, 2010, 2012).

Jiménez Fernández (1999) mentions further contexts in which the aspiration of implosive /s/ is involved. Aspiration of /s/ before fricative consonants /f/, /s/ and /θ/ causes the gemination of these sounds with almost complete loss of aspiration. Before *ch, ll, y*, it is hardly maintained, with complete assimilation to these consonants. In the case of *ch*, aspiration can lead to fricatization.

los llevo > [loɟéβo]

más chico > [máʃĩko]

Before *r* and *rr* aspiration is lost, giving way to a complete assimilation, as in *las ratas* > [lar̄áta]. Furthermore, in the case of *l*, there can be two solutions: aspiration + consonant gemination or no aspiration + consonant reduplication, as in *muslo* > [múl.lo] or [múhl.lo]. Implosive /s/ is aspirated before *m, n, ñ*, such as *las niñas* > [lahníɲa]. The nasal consonant can also be geminated in the presence of aspiration or, to a lesser extent, without it, as in *mismo* > [míhm.mo] or [mím.mo].

The aspiration of /x/ is found throughout the whole region and all social classes except in Jaen and some areas of Granada and Almeria, where we can find the full realization of velar /x/. We must distinguish between the voiceless [h], given in western Andalusia among educated speakers, and the voiced [ɦ], which is more relaxed and generally found in less educated speakers. A very weak aspiration is also possible but it is considered fairly vulgar. Additionally, there exists a halfway sound between velar /x/ and aspirated [h] which is found in eastern areas neighboring western Andalusia and educated speakers who intend to approach the standard. These two allophones can be represented as [h^x] or [x^h], depending on their approach to either the velar or the aspirated sound.

The aspiration of Castilian /x/ can be traced back to the 16th century and the evolution of sibilants during that period as a transition from Medieval to Modern language. The minimal pair /ʃ/-/ʒ/ merged into /ʃ/ after a process of devoicing, the result of which was forced to retract its place of articulation due to its similarity to /s/, giving way to the velar voiceless fricative /x/ that we know today. However, it was not until 1815 that the orthography reforms began and finally changed the spelling of /x/ from *x* to *j*. Nevertheless, in areas where aspiration was kept in words derived from *f*-initial Latin words, /ʃ/ did not result in /x/ but was confused with the aspiration of initial *f*, as in the case of Andalusia.

The aspiration of *h* occurs when *h* is in initial position in words derived from Latin terms beginning with *f*, as *facere* > *hacer*. It is an archaic feature present in the whole western part of Andalusia, virtually in rural areas and uneducated speakers, which lacks in prestige among experts because it only applies to certain words. It is also linked to expressive and informal situations, and has been fixed in specific words and expressions, such as *cante jondo* (flamenco type of singing). The evolution of *f*- >

[h] > /ø/ is a much debated issue which has not seen a consensus to this day, although some historians link its geographical distribution in Andalusia to the Reconquest of this region. On the one hand, the reconquest and repopulation of Jaen was carried out by the Reign of Castile, which had lost the aspiration of f-initial words; thus this region does not preserve aspiration. On the other hand, the reconquest of Seville and Cordova took place after the unification of the Reigns of Castile and Leon, the latter of which still preserved the aspiration of f-initial words in the 16th century. This feature spread throughout the whole western region of Andalusia and eastern Granada, the reconquest of which initiated in Seville.

1.2.1.3 Mergers

Among the features that can be found in the region to a lesser extent is the merger of the alveolar vibrant /r/ and the alveolar lateral /l/ in syllable-final and word-final positions. In the western side of the region, both sounds tend to converge towards [r] and tend to be lost in word-final position. In the eastern part of the province, the sounds tend to converge towards [l], although it is following a decreasing tendency. In isolated areas /r/ can be aspirated as [h] and even dropped giving way to the gemination of the following consonants, usually /n/ and /l/.

[r] *soldado* > [sordáo]

[l] *cuerpo* > [kwélpo]

[h] *carne* > [káhne]

The consonant cluster *rl* can find diverse realizations: a) standard pronunciation, b) aspiration of /r/, c) gemination of /l/, d) complete assimilation, e) palatalization into [ʎ], given in areas where *ll* and *y* merge. The cluster *rn* usually undergoes aspiration of /r/ and germination of /n/, with or without aspiration, as in *carne* > [káhne] > [káhn.ne] > [kán.ne]. In word-final position, as all final Spanish consonants, they tend to be lax and lose their phonemic opposition following the tendency to keep open syllables CVCV. In those cases where total deletion occurs, it can cause the opening of the previous vowel, as in *ver* > [bé] or [bę́].

The first examples of this merger date back to the 12th and 13th centuries in Toledo and frequently given in Andalusian texts in 14th-17th centuries, although it seems it did not spread until the 16th century, testimony of which is also found in America from this century onwards.

Another type of merger is called *yeísmo.* This term refers to the merging of lateral palatal /ʎ/ and fricative palatal /ɟ̄/ (pronunciation of *ll* and *y*). It is spread throughout the whole Spanish community except in some areas in Huelva and Seville, much rare in Cadiz, Malaga, and Almeria. However, in Jaen we can find an affricate pronunciation also given in Toledo and South America.

1.2.1.4 Fricatization of ch

While Spanish *ch* is affricate /t͡ʃ/, Andalusian *ch* can also be fricative /ʃ/ in expressive situations, with a variety of realizations that range from interdental or dental to palatal, being the pre-palatal version the most frequent. Although it is a feature decreasing in frequency, which speakers who use it tend to avoid in formal situations, it is still identified as a stereotyped characteristic outside the area. This feature is closely linked to the merger we have just described above, *yeísmo*, as they exemplify the Andalusian tendency to merge phonemes: [t͡ʃ] > [ʃ] and [ʎ]-[ɟ̄] > [ʝ], giving way to the minimal pair of voiceless and voiced pre-palatal fricatives [ʃ]-[ʝ].

For Alvar (1990), these processes are irreversible and will lead to the establishment of the opposition mentioned above. What is clear for him is that *ll* will never be lateral again and *ch* will never be affricate after fricatization is established. The second assumption in this statement seems overambitious, especially if we take into account Villena Ponsoda's (2002) claims. Speakers in areas of *seseo* are relatively conservative and avoid the lenition of [t͡ʃ], while speakers in areas of *ceceo* are more innovative, allowing the lenition of [t͡ʃ] and [ɟ]. As we have seen before, *ceceo* is a minority phenomenon and the least prestigious solution in the presence of *seseo* and distinction.

Table 1

Phonetic inventory of seseo, ceceo and distinction. Based on Villena Ponsoda (2002, p. 199)

seseo				ceceo				distinction			
labia	denta	palata	vela	labia	denta	palata	vela	labia	denta	palata	Vela
l	l	l	r	l	l	l	r	l	l	l	r
p	t	t͡ʃ	k	p	t		k	p	t	t͡ʃ	k
b	d	ɟ	g	b	t		g	b	d	ɟ	g
f	s		h	f	θ	ʃ-ɟ	h	f	θ-s		h

1.2.1.5 Consonant deletion

Other than the aforementioned /s/, /r/ and /l/, the rest of consonant phonemes tend to be uttered with a lax pronunciation in the whole southern region of Spain, with the possible appearance of total deletion of the final consonant.

In syllable-final position within a word and followed by another consonant phoneme, the first consonant tends to disappear and the second one is geminated, as in *obturar* → [otturár]. Before [h], /n/ can be dropped with the possible nasalization of the previous vowel sound, as in *naranja* → [narãha]. In the case of word-final /n/, we can find two situations: a) stressed syllable and b) unstressed syllable, with the following solutions:

a) velar nasal consonant [ŋ] is the most frequent
   alveolar nasal consonant [n] is random
   nasal vowel [ã] is random
b) velar nasal consonant [ŋ] especially in western Andalusia
   nasal vowel [ã] especially in eastern Andalusia

(Jiménez Fernández, 1999, p. 72)

When the consonant cluster *–ns-* is followed by another consonant sound, either /n/ is dropped, as in CS, or the /s/ is, as in some regions of Andalusia.

Deletion of intervocalic consonants is characteristic of vernacular Spanish in general, with higher incidence in suffixes and verbal endings. Deletion of *d* between vowels of the same nature, as is the case of feminine past participles, gives way to the assimilation of both vowels into one: *cansada* > [kansá]. This is also the case of *todo* and *nada*, which are reduced to *to* and *na* in these same contexts. Nonetheless, the most common case of deletion between different vowels is the past participle *–ado*, even in high-class speakers, with restitution of [ð] in careful speech. This feature is extending even among higher spheres of peninsular Spanish, so much so that the Real Academia de la Lengua Española accepts it. In past participle *–ido*, [ð] is deleted in informal situations and in less educated speech, being less socially accepted. Deletion of [ð] in words ending in *–dor* is considered vulgar, except with established terms such as *cantaor* and *bailaor*. Deletion of intervocalic *g* [ɣ] and, to a lesser extent *b* [β], are also observed, as exemplified in *migajita* > [mihíta] and *tobillo* > [toíɟo].

Other occasional consonant loss phenomena occur with intervocalic /r/, which can frequently disappear from certain verbal forms such as *mirar, parecer,* and *querer*:

mira tú > [míatú]

me parece > [mepaése]

quieres tú >[kjétú]

and also in words such as *padre* and *madre*, giving way to [páe] and [máe].

1.2.1.6 Vowels and diphthongs

The aspiration and loss of implosive consonants, especially /s/, can imply the opening of the previous vowel, which is pronounced in a more open manner and is also lengthened, with the appearance of certain aspiration depending on the speakers. This opening is omitted in western Andalusia when there is no such aspiration; however, in eastern Andalusia, the opening of the vowel is maintained even in the absence of aspiration. Therefore, this characteristic has an impact in the opposition singular-plural nouns and second-third person singular verb tenses[8], which is solved in the following manner (Jiménez Fernández, 1999, p. 18):

a) In eastern Andalusia, speakers tend to close vowels in singular forms, as in *poco* > [pọ́kọ], which also applies to third personal singular verbs. In the case of plural forms and also second person singular verbs, speakers tend to open vowels, as is the case of *pocos* > [pǫ́kǫ]. Notice the assimilation of this opening by all vowels within the same word, a phenomenon found mainly in Cordova and Granada called umlaut, which especially takes place when vowel sounds share pitch. Likewise, stressed vowels can be uttered with greater articulatory force, leading to lengthening: *pocos* > [pǫ́:kǫ].

[8] See Appendix B.

b) In western Andalusia, the neutralization of the opposition is solved by means of the aspiration of /s/. In cases where this aspiration disappears, speakers resort to determiners, subject pronouns, and linguistic context in general to eliminate ambiguity.

There is indeed a great variety of realizations derived from the aspiration or the loss of final /s/. From opening to lengthening, to umlaut, to total loss with no influence on the previous vowel sound. Geographically, we can find that in western Andalusia vowel sounds are not modified when final /s/ is dropped while in eastern Andalusia these vowels are more open, with umlaut present in Granada and Cordova. All in all, the aspiration and/or loss of final /s/ are characteristic features of southern Spanish. Narbona et al. (1998, p. 142) enumerate a list of possible solutions to these phenomena:

a) opening and systematic umlaut
b) opening and random umlaut
c) opening without umlaut
d) opening, lengthening, and umlaut
e) opening, lengthening, and random umlaut
f) opening, aspiration, and umlaut
g) opening, aspiration, and random umlaut
h) opening, lengthening, and systematic aspiration
i) opening, lengthening, and random aspiration
j) alternation between opening and leveling, predominance of the former
k) alternation between opening and leveling, predominance of the latter

In this line of vowel loss, there can also exist an occasional loss of /o/ and /e/ in specific contexts.[9] /o/ tends to be dropped in *estoy* and *voy* when they function as auxiliary verbs within a verb tense (present continuous and periphrastic future, respectively):

Estoy pensando que… > Est'y pensando que…

Voy a tener que... > V'ya tené que ...

/e/ is usually omitted in the definite article *el* when the following word begins with a vowel sound, linking the remaining /l/ to that vowel sound. It also tends to disappear from personal pronouns such as *me, te, se* and auxiliary verbs such as *he* in compound verbs tenses (present perfect). It can also be dropped in the preposition *en* when preceded by a vowel sound.

El abuelo > l'abuelo

Se me ha roto > Se m'a roto

Mira en el cuarto > Mira'n er cuarto

Indeed, /e/ tends to be dropped in combination with another vowel sound, especially /a/, which is the strongest vowel phoneme and often makes neighboring vowels become lax or even disappear. This shows the tendency that exists in Andalusia towards elision of phonemes and merging of words in speech.

1.2.1.7 Other phenomena

Apart from the characteristics described so far, there exist a further number of linguistic phenomena present in AS, mentioned in Gutier (2010). Metathesis consists of the pronunciation or writing of a word in which one or more of its phonemes or letters switch positions, as in *nadie* > *naide* or *pobre* > *probe*. Contraction is the

[9] For further information and examples of how pronunciation affects morphosyntax, refer to Appendix B.

linking of two words, the second usually beginning with a vowel phoneme, without an apostrophe. This derives into a new word, as in *para adelante* > *palante.* Dissimilation is the transformation of a phoneme from the influence of a neighboring and similar phoneme. /b/ can be velarized to /g/ in everyday words such as *bueno* > [gwéno] and *abuelo* > [aɣwélo]. The opposite is also possible, as in *aguja* > [abúha]. Finally, epenthesis is the addition of an extra phoneme inside a word. It is common to add /n/ after initial vowel followed by /r/, as in *irritación* > *inritación.*

In 1995, Alvar, Llorente, and Salvador concluded a ten-year field study of the speech features of AS. They traveled all over the region of Andalusia recording subjects reading a series of texts, which they then phonetically transcribed in detail, to subsequently determine the areas where the different features described above were given.

Furthermore, Alvar (1990, 2004) also conducted a thorough study to determine the characteristics of the speech in Seville. The research comprised the recording of the informants in spontaneous conversation, followed by an indirect question for the uneducated informants, and the reading of a passage for the educated ones. The study was then implemented by complementary questionnaires administered to informants from diverse occupations.

He observed that most of the characteristics of WAS were present in the speech of the less educated informants, while the educated ones had only some of them. Both groups employed *seseo,* i.e., substituted [ş] for /θ/, with occasional *ceseo* in uneducated informants and occasional distinction in educated speakers. Both groups also employed *yeísmo*, i.e., realized both *ll* and *y* as [ɟ̄], being the uneducated speakers the ones who presented other realizations, such as affricate or vibrated. The production of *ch* was divided between a palatalized version and the standard for both groups. The aspiration of f-initial Latin words was absent from the speech of all speakers, and both groups aspirated [x], the voiceless allophone [h] for initial position, and the voiced allophone [ɦ] for intervocalic position and before voiced consonant. The loss of intevocalic *b,d,g* was mainly restricted to the case of *d,* particularly for the past participle, and almost exclusively given in uneducated speakers. In the case of implosive /n/, there is a tendency in both groups to velarize this phoneme and nasalize the previous vowel; when it is preceded by another consonant sound, aspiration of this sound and gemination of *n* tends to occur. Finally, aspiration of implosive /s/ before consonant was constant among all speakers, with a tendency towards the gemination of the consonant, more frequently observed in uneducated speakers; nevertheless, before a vowel sound, speakers generally linked /s/ to the following sound.

1.3 The Morphological Marker –*s*

In the two previous sections, we have seen the main phonetic characteristics of AAE and WAS. One aspect of WAS, the aspiration of /s/, is linked to the morphology of this dialect, giving way to the phonetic realization of the morphological marker –*s* of verb agreement and plurality in nouns.

Second-person verbs and third-person verbs in Spanish are distinguished by final /s/, as in *Él viene mañana* (He comes tomorrow) and *Tú vienes mañana* (You come tomorrow), just as English verbs are. In the presence of subject or personal pronouns, distinction is not problematic for WAS speakers. When these elements are omitted, speakers recover the meaning from context and from resorting to the cue of aspiration. In the case of singular nouns and plural nouns, just as in English, these are generally distinguished by the final /s/. The presence of masculine articles makes distinction easy for WAS speakers, as in *El perro/Un perro* (The dog/A dog) and *Los perros/Unos perros* (The dogs/Some dogs), because the form of the article changes. In the case of feminine nouns, speakers resort to aspiration, as in *La mesa/Una mesa* (The table/A table) and *Las mesas/Unas mesas* (The tables/Some tables).

Connected to this aspect, O'Neill (2005) conducted a study on the production and perception of final /s/ by native speakers of WAS in second-person singular verbs and plural nouns. As the target words were in sentence-final position, /s/ tended to be uttered as a very weak aspiration. However, when compared to the production of third-person singular verbs and singular nouns, the author observed that there also seemed to be a very slight aspiration in these cases, giving way to the phonetic neutralization of this phonological contrast. These results were also reflected in his perception experiment, i.e., "in final position, in normal speech, there is no distinction between the sequence VS and V and therefore, the morphological distinctions which rely on this final sibilant element are lost in this position" (p. 159). In our current experiment, target words are embedded in initial or medial position precisely to avoid this neutralization.

In AAE, the use of the morphological marker *–s* is different than in GAE, in the sense that it is usually omitted from third-person verbs and, to a lesser extent, from plural nouns in the presence of quantity markers, and also from genitive constructions. Nevertheless, it is employed to function as a narrative indicator or as indicative of habitual behavior in first-person verbs (Green, 2004a; Smitherman, 1999). Thus, the morphological marker *–s* exists in AAE, although its use and function differ from those in GAE.

In relation to this phenomenon, particularly interesting is the work by Johnson (2005) and de Villiers and Johnson (2007), who studied the comprehension of third-person singular /s/; in the first case, by AAE-speaking children and, in the second case, across dialects of American English (including AAE). Results from these studies indicate that AAE-speaking children do not understand /s/ as a number agreement marker. If it were part of their underlying system, in spite of its infrequent realization phonetically speaking, they would be sensitive to its perception: "If the speaker has available two competing grammars, then in comprehension, third /s/ would be understood as an agreement marker of singular subjects, but zero marking would be ambiguous between the two grammars" (Jonhson, 2005, p. 117). These findings have been recently supported by Beyer and Hudson Kam (2012). However, from these results we cannot determine if speakers of AAE may use third-person /s/ as a subject marker at a later age. We will revisit these studies when stating our objectives in Section 2.3.1.

1.3.1 Description of [h] and [s]

Fricatives are sounds produced with an obstruction in the vocal tract that generates noise. "Frication noise is generated in two ways, either by blowing air against an object … or moving air through a narrow channel into a relatively more open space" (Hagiwara, 2009[10]). The first description is pertinent to [s] while the second one refers to [h], our two sounds under study. The obstacle in the production of [s] is the teeth, at the front of the oral cavity, while [h] is produced at the back of the oral cavity without an obstacle against which air blows.

[10] http://home.cc.umanitoba.ca/~robh/howto.html

Aspiration in Spanish can take place two ways: derived from the phoneme /x/ or from implosive /s/. In both cases, it has been traditionally transcribed as [h], without taking into account its position and its different realizations according to the surrounding context. As reported by Marrero (1990) in her study of the Spanish spoken in the Canary Islands, aspiration derives from /x/ tends to be pharyngeal [h], while aspiration of implosive /s/ is laryngeal and can be voiced [ɦ] in intervocalic position, similar to breathy speech, but velar [x][11] before velar consonants. The laryngeal aspiration, she argues, is similar to the English /h/, with which Widdison (1993) agrees: *"según Ladefoged, el sonido del murmullo corresponde, a groso modo* [sic], *a la pronunciación de la [h] intervocálica de las palabras inglesas* <u>*ahead*</u> *y* <u>*behind*</u>*"* (p. 47) ["According to Ladefoged, the sound of the murmur corresponds, broadly speaking, to the pronunciation of intervocalic [h] of the English words <u>ahead</u> and <u>behind</u>"]. An example of WAS intervocalic [h] can be seen in Figure 1 below:

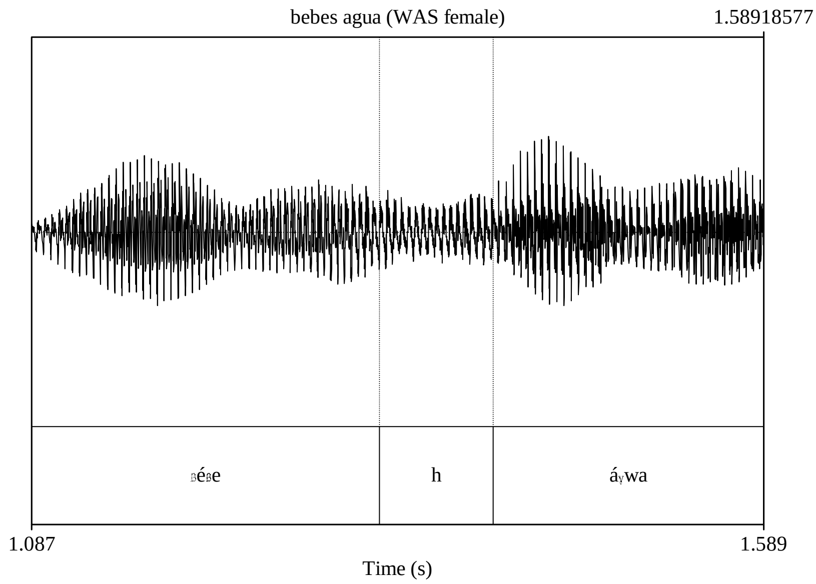


Figure 1. *Intervocalic [h] in "bebes agua" by a WAS female speaker*

[11] Although /x/ is used to label the CS phoneme, [x] is employed here to account for the velar place of articulation of aspiration.

In English, /h/ is defined as a voiceless glottal fricative sound by some authors (Collins & Mees, 2003; Ogden, 2009, Roach, 2010; among others), while other authors do not classify it as a consonant but rather as part of the vowel (Hagiwara, 2009; Jongman, Wayland, & Wong, 2000; Ladefoged, 1982). As Johnson (2012) points out, this sound is fricative "if we define the class as sounds produced with turbulent airflow; but, unlike other fricatives, they are nonconsonantal in the sense that they have … vowel-like spacing between formants" (p. 160) with higher amplitude in their higher formants than vowels. Lorenz (2012) explains that "the IPA chart lists it as glottal, but the constriction is rather somewhere in pharynx or larynx" (p. 30). In any case, in intervocalic position, /h/ also becomes voiced [ɦ].

The characteristics of [s] and [h] make [s] the strongest fricative sound while [h] the weakest fricative sound. In fact, the spectral peak in [s], with the highest frequency concentration of all fricative sounds, is near 8 kHz, with a minor peak around 4 kHz, whereas [h] has a much lower frequency. The pharyngeal fricative [h] peaks at 1.5 kHz, while the laryngeal fricative [ɦ] peaks at 2.56 kHz and the velar fricative [x] peaks at 3.45 kHz (Martínez Celdrán & Fernández Planas, 2007).

In her study, Barreiro Bilbao (1994) conducted an acoustic cross-analysis of RP English and Spanish fricatives. As an isolated sound, she found that both English and Spanish [s] have a smaller range of frequencies than [f]. English [s] showed a concentration of energy around 13 349 Hz, while Spanish [s] had a concentration of energy around 10 915 Hz. In citation form, English [s] had a duration of 205.8 ms in initial position, 209.1 ms in medial position, and 299.7 in final position, while Spanish [s] presented a duration of 192.5 ms in initial position, 156.4 ms in medial position, and 197.5 ms in final position. The energy of English [h] stretched up to

8703 Hz and showed a duration of 91.9 ms in intial position and 156.8 ms in final position. Nevertheless, she did not analyze Spanish aspiration [h].

However, as seen in Section 1.2.1, aspiration of implosive /s/ is not only a matter of /s/ → [h]. It is generally deleted in absolute position, while it is commonly realized as [h] or even [s] in word-final position followed by vowel. When followed by voiceless stops, we can observe pre- and/or post-aspiration; when followed by voiced stops, these tend to become fricatives (as we will see in the following chapter) and, when followed by nasals, lateral, and other fricatives, gemination is the most common solution (Romero Gallego, 1995).

In the case of /s/, it is a voiceless alveolar fricative sound in CS, as well as in English, particularly in syllable-initial position and in syllable-final position when followed by a voiceless consonant (except /θ/ or /t/) or a pause, and in intervocalic position. When in syllable-final position and followed by a voiced consonant (except /d/), it becomes a voiced alveolar fricative sound. Additionally, when followed by /θ/ or /t/, it is a voiceless dental fricative sound, but if followed by /d/, it becomes a voiced dental fricative sound (Garrido Almiñana, Machuca Ayuso, & de la Mota Gorriz, 1998). Additionally, the following consonant not only affects the place of articulation and voice of /s/ but it can also affect its intensity, frequency, and duration. Nevertheless, implosive /s/ also has an effect on its surrounding context, e.g. it lengthens the preceding vowel (Widdison, 1993).

As an example, Figure 2 below shows CS intervocalic [s]. Observe that, as opposed to Figure 1, the sibilant is clearly delimited between the vowels.

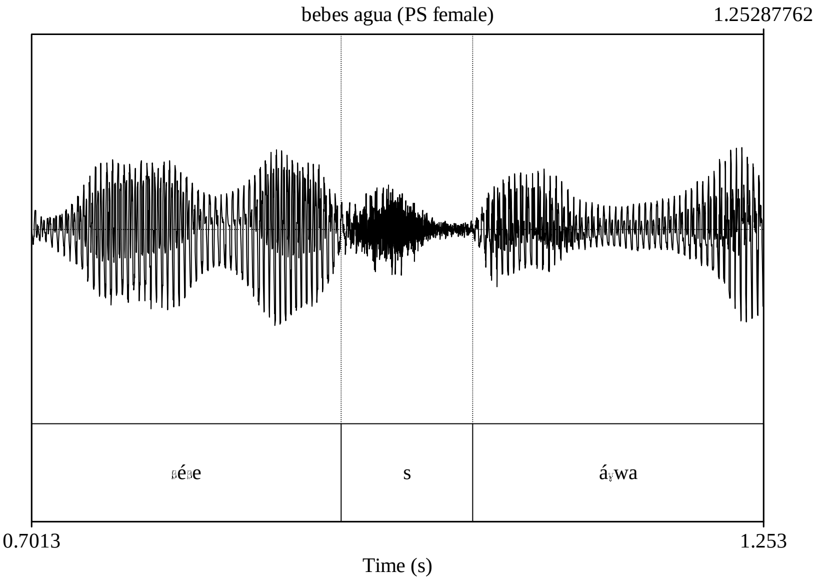


Figure 2. *Intervocalic [s] in "bebes agua" by a CS female speaker*

In relation to this, Widdison (1993) conducted an experiment to explain the possible origins of Spanish aspiration in syllable- and word-final position. A native Spanish speaker recorded words with and without implosive /s/, such as *pasta* and *pata*. The author then separated the vowel preceding /s/ and inserted it in the word without /s/, replacing its actual vowel. Upon doing this, he conducted an identification task with native speakers of Spanish, a great number of which identified the new word as containing /s/, even though it was not physically present. His conclusions were that the vowel alone already indicates the presence of the sibilant that the listeners associate with /s/ at a lexical level, whether it is /s/ that actually follows or aspiration, i.e. *"[h] siempre está presente en la señal acústica de la vocal, pero sólo se percibe cuando los rasgos esenciales de [s] se reducen a un mínimo"* (p. 55) ["[h] is always present in the acoustic signal of the vowel, but it is only perceived when the essential features of [s] are reduced to a minimum"].

1.4 Summary

In this chapter, we have seen how both AAE and WAS are dialects of English and Spanish that carry certain stigmatization and the origins of which are not completely defined. The two dialects have a set of phonetic features that makes them unique and sets them apart from the mainstream characteristics of English and Spanish. AAE is represented by an absence of interdental fricatives in syllable-initial or syllable-final position. Instead, we find alveolar stops and labiodental fricatives. It is also characterized by the absence of /r/ and /l/ in syllable-final position. WAS is mostly characterized by the aspiration of the sibilant /s/, which affects the following sounds: aspiration and post-aspiration in voiceless stops, fricatization of voiced stops, gemination of nasals and other consonants. Additionally, it displays a set of mergers and the fricatization of /tʃ/.

What these two dialects have in common is the deletion of consonants in medial or final position and how their morphological marker for verb agreement and plurality is affected. AAE absence of the morphological marker *–s* from third-person verbs and plural nouns seems to come from internal grammatical rules while WAS aspiration of this marker in second-person verbs and plural nouns is derived from its phonetic characteristics.

In the following chapter, we review the acoustic characteristics of the English and Spanish sounds that concern us in this study.

CHAPTER 2

ACOUSTIC PHONETICS

When we talk about phonetics, we can do so from the point of view of production, transmission, and perception of speech sounds. Articulatory phonetics describes how sounds are formed by the vocal tract of the speaker; acoustic phonetics describes the characteristics of the sounds that reach our ears; and perceptual phonetics studies how these sounds are understood by the listener. The listener needs to actively participate in the process, extracting information from the signal in terms of intrinsic characteristics and context characteristics. The listener also uses information that is independent of the signal and in relation to their linguistics experience stored in memory.

To decode a linguistic signal, listeners go through three stages (Marrero, 2001): i) audition, it is a passive and automatic mechanism by which the signal activates the fibers in the auditory nerve that allow us to distinguish sounds, ii) perception, when the nerve system converts the signal into linguistic units, and then segments, classifies and categorizes them, and iii) comprehension, which is the interpretation of the message in terms of grammatical and semantic meaning. Given that our articulatory system tends to produce sounds as similar as possible, and that our perceptive system needs sounds to be as distinguishable as possible, it seems that our perceptive system has played a key role in the evolution of language.

In discrimination tasks, where listeners have to determine whether two sounds are similar or different, the mechanism activated is auditory, i.e., the characteristics of the sounds are essential. In identification or categorization tasks, when listeners have to identify and label stimuli, they resort to their mental models that they have of such sounds to make a decision. One of the main differences between discriminating and identifying is that we can potentially detect minimal differences between sounds but our capacity to categorize and store them in memory is limited. Here, two processes of perception are at play, as we mentioned above. Auditory perception is a bottom-up process based on the physical characteristics of the sounds, while categorical perception is a top-down process that interprets sounds in terms of the pre-existing categories in memory. When we use this last process, we label sounds that share certain characteristics within the same category. Two sounds can considerable differ in parameters such as duration and frequency but still be assigned to the same category. This encompasses the variability that can be found in the signal, such as coarticulation and dialectal variation. On the contrary, other sounds may minimally differ in one property that is important enough to be categorized as two distinct sounds. As Martínez Celdrán and Fernández Planas (2007, p. 113) state "*diferencias articulatorias pueden producir cambios acústicos muy destacables o, por el contrario, cambios mínimos*" ["articulatory differences can produce very remarkable acoustic changes or, on the contrary, minimal changes"]. Therefore, categorical perception maintains the characteristics that distinguish sounds and minimizes irrelevant differences to compensate for the imperfect one-to-one correspondence between acoustic cues and phonetic features.

2.1 Acoustic Cues

We make speech sounds audible when the air is pushed out of our lungs while producing a noise in our throat or mouth. By means of the actions of the tongue and the lips (articulators), we make changes in these basic noises. The speech sound is a spectrum of acoustic energy produced by the vibration of our vocal folds and then filtered by the articulators in our vocal tract.

The mechanism of speech production involves four processes (Ladefoged & Johnson, 2010): i) the airstream process, that is, the ways in which we push air out of our lungs; ii) the phonation process, which involves the actions of our vocal folds. When they vibrate, they produce voiced sounds; when they do not vibrate, they produce voiceless sounds; iii) the oro-nasal process, by which we produce oral sounds when the air escapes through the oral cavity and nasal sounds when the air escapes through the nasal cavity); and iv) the articulatory process, by which our tongue and our lips interact with the roof of the mouth and the pharynx to articulate the sounds.

Speech sounds can be divided into three categories (Hagiwara, 2009; Ladefoged & Johnson, 2010): i) periodic voicing, which is produced when the vocal folds vibrate; ii) devoicing, sounds produced without vibration of the vocal folds; and iii) aperiodic noise, which is when a turbulent airflow is produced in a random way.

In the production of vowels, on the one hand, the vocal tract is relatively open and the air escapes without obstruction, which gives these sounds great loudness. The vocal folds vibrate and, thus, vowels are voiced. “The primary acoustic characteristics of vowels is the location of the formant frequencies, specifically, the first three formants (F1-F3)” (Reetz & Jognman, p. 182), which provide information about vowel quality. The rest of the formant frequencies above F3 provide more information about the identity of the speaker than about the type of vowel. The case of consonants, on the other hand, is more complex, as Ladefoged and Johnson (2010) point out:

> The acoustic structure of consonants is usually more complicated than that of vowels. In many cases, a consonant can be said to be a particular way of beginning or ending a vowel, and during the consonant articulation itself, there is no distinguishing feature. (p. 198)

According to the movements and position of the articulators, consonants are classified in terms of place of articulation and manner of articulation. For example, /p/ is a bilabial consonant because the lips close to form the sound. It is also a stop because the air is stopped from escaping the mouth when the lips close. Consonants can be voiced or voiceless, i.e., they are voiced when the vocal folds vibrate when producing the sounds, and voiceless when the vocal folds do not vibrate in such production. Stops and fricatives are the only English consonants that can be either voiced or voiceless. In Spanish, only stops can be voiced and voiceless. Nevertheless, if we take into consideration the allophones produced after the aspiration of /s/ in WAS, we can say that this dialect can additionally have voiced fricatives.

In general, stops and fricatives behave in a similar way according to certain aspects. Vowels before voiceless stops and voiceless fricatives are generally shorter than vowels before their voiced counterparts. Additionally, voiceless stops and voiceless fricatives are longer in word-final position than their voiced counterparts. Both types of consonants produce an obstruction to the passage of the air, and that is why they are called obstruents. However, fricatives are produced by the close approximation of two articulators so that friction can be heard. In turn, stops are produced by a total closure of the airstream (Ladefoged & Johnson, 2010). The primary difference between voiced and voiceless stops in English, rather than voicing itself, is that voiceless stops display aspiration. In Spanish, the main source for distinction between both sets of sounds is Voice Onset Time (Abramson & Lisker, 1973). English fricatives, on the one hand, are traditionally divided into sibilants /s, z, ʃ, ʒ/ and non-sibilants /f, v, θ, ð/. On the other hand, Spanish fricatives are the sibilant /s/ and the non-sibilants /f, θ, x/. In all case, the sibilants display greater loudness than the second set of fricatives. As we explained in Section 1.3.1, the classification of [h] is controversial. Some authors describe it as a glottal fricative (Collins & Mees, 2003; Ogden, 2009, Roach, 2010; among others), while others view it as part of the vowel because it presents vowel-like formants (Hagiwara, 2009; Jongman, Wayland, & Wong, 2000; Ladefoged, 1982). We will consider [h] a fricative sound for the purposes of our study.

An acoustic cue "consists of one or more acoustic properties that are considered to provide unique information about the identity of a particular segment" (Reetz & Jognman, 2009, p. 185). When reading spectrograms, we should be aware of the existence of three basic types of sounds (Ladefoged & Ferrari Disner, 2012): i) a stop sound is characterized by a white gap representing a period of silence (closure), followed by a thin vertical stripe that is darker (the release burst); ii) a fricative sound is characterized as dark areas close to the top of the spectrogram, iii) vowel, approximant, and nasal sounds are characterized by parallel horizontal bands (formants), which can be from two to five in number, generally with one of these bands below 1000 Hz and another band between 2000 Hz and 3000 Hz. In experimental tasks, it is common to employ citation-form words as stimuli. These forms are isolated words presented one at a time. However, when stimuli are sentences or recorded conversations, "the range of phonetic variability found in connected speech is a good deal greater and more subtle than the variability found in citation forms" (Ladefoged & Johnson, 2012, p. 108), which makes it difficult to describe the sound patterns in terms of phonetic symbols. For this reason, quantitative measurements of acoustic cues such as duration, amplitude, frequency, voice onset time, center of gravity, and formant transitions is often a more adequate way to carry out these descriptions.

2.1.1 Stops

The study of consonants has focused on the perception of stop consonants from the very beginning. Their articulation consists of three periods: i) shutting, the period where the articulatory organs move towards the place of articulation of the stop sound, ii) closure, the period of total closure that prevents the passing of the air and increases its pressure, and iii) release, the period where a burst of air is released (Johnson, 2012, p. 169). When we produce stops, the air is first blocked by a complete constriction in the oral cavity that causes the closure interval between the previous vowel and the release burst of the stop. This is a primary characteristic of stops represented by a white gap in the spectrogram that contains no energy (voiceless stops) or very low-frequency energy called voiced bar (voiced stops). After the release period, and before the voicing of the following vowel begins, we find what is called Voice Onset Time (VOT). This is defined as “the time interval between the burst that marks the release and the onset of periodicity that reflects laryngeal vibration” (Lisker and Abramson, 1964, p. 422), i.e., the delay between the release of air of a stop sound and the beginning of the vocal cord vibration.

Lisker and Abramson (1964) conducted a well-known cross-language study of voicing in initial stops according to acoustical parameters. They sought to determine the best acoustic characteristic that listeners use to distinguish English voiced stops /b, d, g/ and voiceless stops /p, t, k/. Their data provided enough information that VOT was a good parameter to distinguish the different stops. Across the eleven languages that they studied, they came to the conclusion that their place of articulation plays a key role in the duration of VOT. Labial stops /p, b/ are shorter than alveolar stops /t, d/, and these, in turn, tend to be shorter than velar stops /k, g/ (Whalen, Levitt, & Goldstein, 2007; Cole, Kim, Choi, & Hasegawa-Johnson, 2007, reported in Rojczyk, 2011). Additionally, voiced stops are shorter than voiceless stops. As Martínez Celdrán and Fernández Planas (2007, p. 89) point out, "*a medida que se atrasa el punto de articulación en dirección del exterior al interior, el VOT aumenta*" ["The further back the place of articulation is, the higher the VOT is"]. Figure 3 shows the mean duration values for the six English stops (based on Zue, 1976).

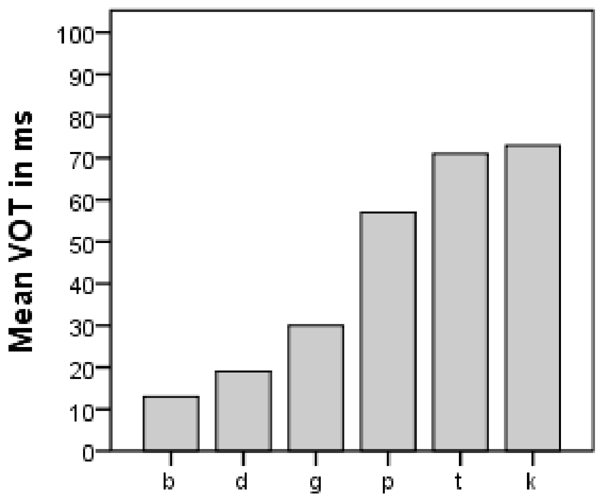


Figure 3. *Mean VOT of English voiced and voiceless stops*

More recent studies by Chen and Alwan (2001, 2006) investigated the effect of VOT and first formant (F1) transition in the perception of pairs of voiced and voiceless stops in noise. Although acoustic cues for voicing can be found in a higher fundamental frequency, the absence of aspiration, and the presence of a voice bar, VOT proved to be the best characteristic for the classification of stops. F1 was relevant when stops were followed by /a/ and when the stimuli were presented in noise. In this respect, Lisker's (1975) perception study found that varying the duration of the VOT lead to significant changes in the perception of stops, and although F1 had some effects on perception, it was not enough to differentiate stops alone.

Even more interestingly, Lisker and Abramson (1964) also determined that VOT in word-initial position can be classified into three groups: i) negative VOT, voicing starts before the release of the stop (-30 ms or more), characteristic of voiced stops; ii) zero VOT, voicing starts approximately at the same time or shortly after the release of the stop (0 to + 30 ms), characteristic of unaspirated voiceless stops; iii) positive VOT, voicing starts after the release of the stop (around +50 ms or more), characteristic of aspirated voiceless stops. On the one hand, as the negative value of it increases, so does the voicing of the consonant. On the other hand, as its positive value increases, so does the aspiration of the consonant. Spanish has negative VOTs for initial /b, d, g/ and zero VOTs for /p, t, k/. In contrast, English initial /b, d, g/ show zero VOTs and /p, t, k/ show positive VOTs (Table 2).

Table 2

Classification of English and Spanish stops by VOT category

	negative VOT	zero VOT	positive VOT
English		b, d, g	p, t, k
Spanish	b, d, g	p, t, k	

Cho and Ladefoged's (1999) later focused on a cross-language study of VOT in 18 languages in which they corroborated the existence of the three categories, claiming that "the strongest evidence in favor of there being only three values is that no languages have more than three contrasts" (p. 226). Furthermore, they state that it is the language-specific phonetic rules of a given language that dictates the timing between the beginning of the articulatory gesture and the beginning of the laryngeal gesture, i. e., “the grammar of the language would be supplying context restricted values for features” (p. 227).

Apparently, our natural psycho-acoustic boundary is found at around +35 ms, which makes zero – positive VOT distinction easier than negative – zero VOT distinction. According to Rojczyk (2001, p. 42), “VOT is perceived categorically, that is the discrimination performance is discontinuous” and categorical boundaries are dependent on the place of articulation of the sounds. Approximately, categorical boundaries for English labial stops are about 25 ms, for alveolar stops they are about 35 ms, and for velar stops they are around 42 ms. The Spanish boundary for /t/, according to Lisker and Abramson (1965) is 10 ms to the left of their English counterpart. The same authors (1973) later established the boundaries at 14 ms. In CS, bilabial stops are also reported to have a boundary of about 10 ms (López-Bascuas, Fahey, García-Albea & Rosner, (1998). According to Martínez Celdrán and Fernández Planas (2007), the mean duration of the VOT for /k/ is 35 ms, for /t/ it is 20 ms, and for /p/, 14 ms.

Rosner, López-Bascuas, García-Albea, and Fahey (2000) studied VOT in CS initial stops for comparison with the results in Williams’s (2007) study of stops in Venezuelan, Peruvian, and Guatemalan Spanish. Both studies reported significant effects of voicing and place of articulation on stops VOT, with negative VOT and longer zero VOT for velar sounds than for bilabial and dental sounds. Nevertheless, The VOT for CS /p, t/ were longer than the VOT for their Guatemalan counterparts, while Venezuelan and Peruvian VOTs for /t, k/ were longer than their CS counterparts. In general, the negative VOT in Guatemalan and Peruvian voiced stops were longer than those in CS stops, while it was shorter for Venezuelan /d/ than for CS /d/. Therefore, dialect differences appear for Spanish voiced and voiceless stops.

One study that deviates from the previous studies in terms of the type of stimuli used is the one conducted by Yao (2007) on the closure and VOT of voiceless stops in English connected speech. It seems that the factors taken into consideration proved to only account for 26% of the variability in closure and VOT values. Age and gender could only explain 1% of the variability (considering that their age range was from under 30 to over 40), speaking rate only accounted for 13% of variability in VOT and 4.5% in closure duration, place of articulation could only explain 2.2% of variability in VOT but a higher percentage of variability in closure duration (8.1%). Word frequency, on the contrary, was found to have an effect on both VOT a closure duration, i. e., "If some words occur extremely often, it is possible that they become the target of certain changes in production, for instance, acceleration, phone reduction and coarticulation" (p. 218).

2.1.1.1 Stops after WAS aspiration

In Section 1.2.1 (Alvar, 1996; Gerfen, 2002; Torreira 2007a, 2007b, 2012), we saw a general description of how aspiration of implosive /s/ can affect the following sounds; more in particular, how it affects voiceless stops. As Parrell (2012) argues:

> the productions of /s/ in Western Andalusian Spanish are reported to be somewhat variable, ranging from a full sibilant to preaspiration to a breathy period at the end of the preceding vowel to post-aspiration … with the last being the most common. (p. 37)

Observe Figures 4 and 5, where a clear difference in VOT duration can be detected between CS [t] and WAS [t^h]:

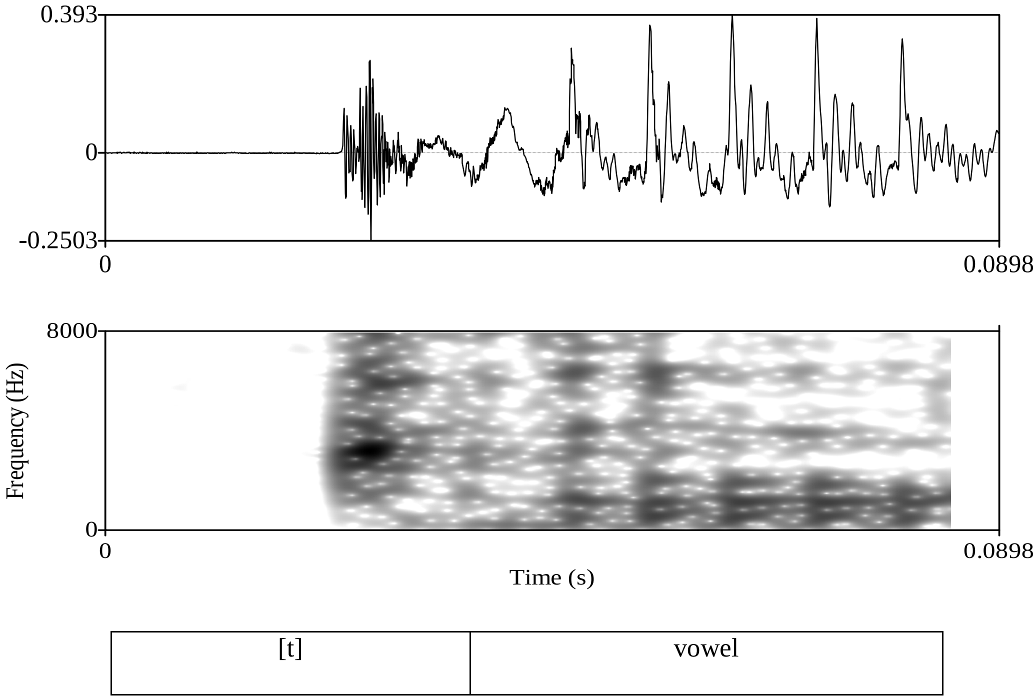


Figure 4. *Word-initial [t] after sibilance*

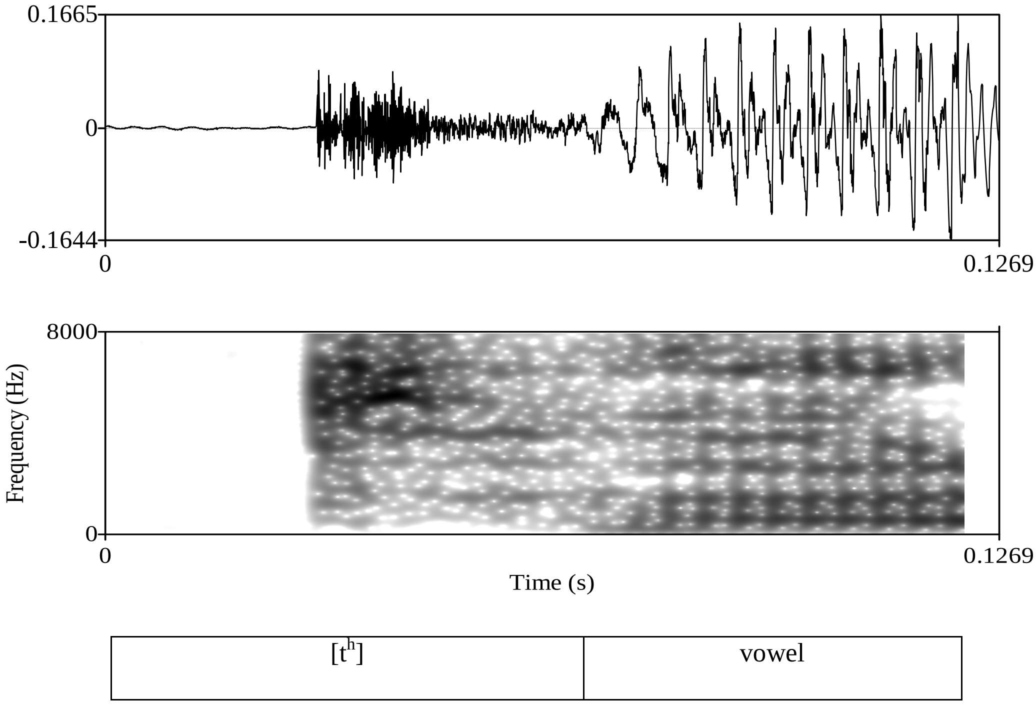


Figure 5. *Word-initial [t^h] after aspiration*

Torreira (2007a) was a pioneer in describing this phenomenon, although he acknowledges that two previous studies had already pointed in this direction (Maza, 1999; Vaux, 1998). His study first analyzed word-internal /st/ in laboratory-recorded speech of CS and WAS speakers, and subsequently in spontaneous speech of WAS and Eastern Andalusian Spanish (EAS) speakers. In both cases, he observed higher VOTs for the Andalusian stop after aspiration in comparison to the CS stop after sibilance. Despite variability found in the recordings, and factors such as speech rate, prosodic context, syllable stress, his findings were consistent with the premise that Andalusian aspiration induces longer VOTs. Another factor that was derived from aspiration is that both stop closure and the previous vowel were lengthened, as long as we consider the aspiration of /s/ as part of the vowel. Otherwise, as is the case with vowels before /s/, they were actually shorter.

Subsequently, Torreira (2007b) compared the production of word-internal /st/ of WAS with the production of the same sequence by speakers of Porteño (Buenos Aires, Argentina) and Puerto Rican Spanish. What he found is that WAS displays shorter pre-aspiration and longer stop closure and post-aspiration period than the other two Spanish dialects. Under the Articulatory Phonology framework proposed by Browman and Goldstein (1989), in which articulatory gestures are seen as phonological units, the author seeks to provide an explanation for this phenomenon. This framework states that “gestures involved in syllable onsets tend to couple into an in-phase relationship, while gestures in coda position are left out of phase with respect to surrounding gestures” (Torreira, 2007b, pp. 118-119), i.e., at onsets, articulatory gestures tend to be simultaneous while at codas they tend to be more variable. His proposition is that of a gestural reorganization in which the glottal opening for the aspirated /s/ and the supraglottal closure for the following stop overlap instead of being sequential, as is the case with dialects with pre-aspiration.

Finally, in 2012, Torreira further investigated WAS aspiration before the three voiceless stops /p, t, k/ according to different speech rates and stress patterns. He found that, despite these two factors, VOT did not significantly vary in duration. Therefore, it seems that “the glottal and supraglottal gestures may be phased very closely even in conditions in which we would not expect much articulatory overlap, hence the lack of significant effects of speech rate and stress location on VOT” (p. 61).

In reference to the variability found in WAS aspirated stops, Ruch (2008) researched the production of /st/ in Seville. What she found were nine possible realizations for this sequence: two with sibilants [st], [^{s}t]; four with aspiration [^{h}t], [^{h}t^{h}], [^{s}t^{h}], [t^{h}], one with assimilation [t:], one with complete deletion of /s/ [t], and finally, the new phenomenon that we mentioned in Chapter 1: the affricated [t^{s}]. The most common of these realizations was the post-aspirated stop [t^{h}] (49.1%), followed by the affricated stop [t^{s}] (22%). Additionally, Ruch (2012) conducted a sociophonetic study of the production of /t/ and /st/ in internal-word position with speakers from WAS (Seville) and EAS (Granada), taking into account their gender and their age. She concluded that young speakers produce post-aspiration significantly more frequently than older speakers not only in Seville, but also in Granada. They also produce less pre-aspiration, although this fact was only significant for speakers in Seville. Additionally, she found that female speakers showed greater differences in VOT values than male speakers.

O’Neill (2009) also studied the sequence /st/ in WAS from Seville, narrowing down the effect of aspiration to two most frequent productions: [‘pa^{ɦ}t̪^{h}a] and [‘pat^{h}a], i.e. aspirated stops with or without pre-aspiration. What is interesting is that the author considers the second realization as part of a new set of phonemes in the dialect [p^h, t^h, k^h], working in opposition to their unaspirated counterparts. Instead of being the result of an overlap of gestures, as proposed by Torreira, “these pronunciations correspond to the phonetic realisation of a different sequence of phonemes” (p. 79), i.e., these set of sounds would be phonetic categories in itself and not the result of coarticulatory gestures.

Parrell (2012) corroborates the claims by Torreira of a post-aspiration phenomenon in WAS, but he states that the question of “whether this reduction is an online phonetic process or a phonological one has not been thoroughly investigated” (p. 37).

Finally, the most recent piece of work concerning post-aspirated voiceless stops of Seville is the study carried out by Horn (2013). She investigated the phenomenon in a sentence reading task from various perspectives. First, she studied whether the post-aspiration reported for /t/ also extended to /p/ and /k/. In this regard, she found that place of articulation “is the only robust predictor of the presence of significantly long postaspiration” (p. 81). Post-aspiration also extended to the velar sounds but not to the bilabial sound, opposite to the findings in Torreira (2012). Its duration was significantly shorter for /p/ than for the other two stops. Second, she aimed at analyzing the phenomenon from a social and linguistic perspective. The longest duration of post-aspiration was found when the preceding vowel was stressed and when in word-internal position, once more, in disagreement with Torreira’s claims (2007a, 2012). Although the social factor had no effect in these realizations, there was a tendency for younger women with college education level to reduce sibilance and produce longer post-aspiration. And third, she interpreted these results under the Articulatory Phonology framework. Just as the previous studies, she concluded that there is a negative correlation between the presence of sibilance and post-aspiration.

2.1.2 Fricatives

Fricative sounds are produced when the articulators constrict the passage through which the air escapes. These sounds are continuant, in the sense that "you can continue making them without interruption as long as you have enough air in your lungs" (Roach, 2000, p. 48). When the air passes through the articulators, it creates turbulence due to the size of the passage and the volume velocity of the airflow. Therefore, "the faster the air molecules move, the louder the sound … the narrower the channel, the louder the turbulent noise" (Johnson, 2012, p. 154). Nevertheless, most fricatives are produced when the air hits an obstacle in the passage, i.e., the teeth or the lips, increasing the amplitude of the turbulence. This turbulence noise is represented as a very dark area in the spectrogram. As it was the case with stops, fricatives can also be voiceless /s, f, θ, ʃ, x, h/ and voiced /z, v, ð, ʒ/ (the classification of /h/ is controversial, as we mentioned earlier).

Fricatives can be described according to four characteristics: "spectral properties of the friction noise, amplitude of the noise, duration of the noise, and spectral properties of the transition into and out of the surrounding vowels" (Reetz & Jongman, 2009, p. 189). Sibilant fricatives have a more pronounced spectral shape because the air hits the teeth. Therefore, the alveolar sibilants typically present clear, distinct spectral shapes while labiodental and (inter)dental non-sibilant fricatives display a relatively flat spectrum. Velar fricatives present little energy at higher frequencies since their greatest amount of energy concentrates at lower frequencies; particularly, in the area corresponding to the F2 of the adjacent vowel.

Unlike their voiceless counterparts, voiced fricatives have two sources of energy: not only does it originate from the turbulent noise derived from the constriction of the air passage, but also from the vibration of the vocals folds, which generate low-frequency energy. The spectrograms of both types of fricatives are similar, with the exception that "they contain additional low-frequency energy corresponding to vocal fold vibration and slightly less intensity in the higher frequencies because part of the energy of the airstream serves to make vocal folds vibrate" (Reetz & Jongman, 2009, p. 192).

In Spanish, the fricative sounds are /f, θ, s, x/, to which Quilis (1981) adds the allophones [h] and [ɦ]. In English, the fricative sounds are the voiceless /f, θ, s, ʃ, h/ and their voiced counterparts /v, ð, z, ʒ/. As we reported in Section 1.3.1, Barreiro Bilbao (1994) conducted a cross-sectional study of the acoustic characteristics of Spanish /f, θ, s, x/ and RP English /f, θ s, ʃ, h/. Among the characteristics measured, we find range of frequency, duration, and their spectral peaks. Concerning the range of frequency, she concluded that non-sibilant /f, θ/ present a great amount of dispersion of energy that extends between 1000 Hz and 15 400 Hz. Non-sibilant /x, h/ have a concentration of energy in the lowest area of the spectrum, from 0 Hz to 11 500 Hz. Sibilant /s, ʃ/ show a narrower band of frequency, from 1300 Hz to 14 800 Hz, although with higher intensity. With respect to this, their place of articulation has an effect on their respective frequency. /f, θ/ are articulated at the front of the oral cavity, /s, ʃ/ are articulated in the mid area of the oral cavity, while /x, h/ are articulated at the back of the oral cavity. The fricatives articulated at the back present lower frequency limits than the other sounds. Those articulated in the middle section have low upper limits and higher lower limits. Finally, the fricatives articulated at the front of the oral cavity present higher lower limits and low upper limits.

With respect to the duration of the fricatives, she found that both sets of fricatives had a similar duration according to their place of articulation. However, fricatives in word-internal position were shorter in Spanish, while fricatives in word-initial position were shorter in English. Additionally, the velar sound /x/ had a similar duration to that of /f/, whereas English /h/ was very short.

For the author, differences in spectral peaks are the key characteristic to distinguish these fricatives. This parameter is crucial to explain why the "*trasvase de algunos de estos sonidos … de una lengua a otra conlleva una pronunciación errónea y, en otros casos … no supone cambios importantes a nivel perceptivo o articulatorio*" (p. 477). ["transfer of some of these sounds … of a language to another leads to an erroneous pronunciation and, in other cases … it does not imply important changes on a perceptual or articulatory level"]. She divides them into three groups:

i) Sibilants, which have formants with great amplitude due to the high-pass filter of the oral cavity. Spanish /s/ has a great concentration of energy in one formant from 3515 Hz to 6317 Hz, while English /s/ has this energy from 4336 Hz to 6619 Hz. "*Cuanto más se retrae la punta de la lengua más baja es la frecuencia de dicho formante*" (p. 467) ["The more retracted the tip of the tongue is, the lower the frequency of such formant"]. According to Quilis (1981), the closer the place of articulation is to the front of the oral cavity, i.e., the dental area, the less strident /s/ becomes. In other words, the length of the vocal tract from the point of constriction to the lips is inversely correlated to the frequency of the peak in the spectrum (Hughes & Halle, 1956).

ii) Labiodental non-sibilants have an almost flat spectrum. For Spanish /f/, the greatest information is contained in the first three formants, that is, below 6000 Hz. For English /f/, this information can also be found around 11 300 Hz. /θ/ presents more noise than /f/, without formants. Its information lies in both low and higher frequencies. For Spanish, it peaks up to 9000Hz, while for English it peaks up to 8000 Hz.

iii) Velar and glottal non-sibilants present great energy in the lower area of the spectrum and have a marked coarticulation with the adjacent sounds. Spanish /x/ contains information in the first three formants below 3000Hz. Over 4000 Hz, it only presents noise without formants. English /h/ has five formants up to 8000 Hz.

In fact, the spectrum of the fricatives articulated at the front of the oral cavity, in conjunction with neighboring vowels, see how their spectral peaks in the higher area of the spectrum increase their amplitude; those fricatives articulated in the middle section of the oral cavity suffer a decrease in their F1 and an increase in their F2; and the fricatives articulated at the back of the oral cavity suffer changes in amplitude and their formant frequencies.

The energy of apical /s/ starts at 3500 Hz and reaches the highest point around the center of the spectrum (Martínez Celdrán, 2004; Martínez Celdrán & Fernández Planas, 2007). However, before dental stops /t/ and /d/, sibilance is said to suffer a process of dentalization, to which Quilis (1966) opposes, claiming that a dental allophone would be close to [θ]. Although there seem not to be great differences between apical /s/ and "dental" /s/, some differences in F1 seem to appear, as well as differences between intervocalic /s/ and "dental" /s/. Whether this is a question of an assimilation process or a coarticulatory process, the authors point at a partial assimilation.

As far as the rest of the Spanish fricatives are concerned, García Santos (2002, reported in Martínez Celdrán & Fernández Planas, 2007), state that their perception varies according to their duration. /f/ is perceived when longer than 90 ms; if its duration is shortened to 40-80 ms, it is then perceived as [v], while it is identified as the approximant [ß̞] when its duration is less than 20 ms. Similarly, /θ/ is identified when its duration is longer than 85 ms, while it is perceived as the approximant [ð̞] when its duration is shorter than 35 ms.

Along these lines, Herrero Moreno and Supiot Ripoll (2002) investigated the characteristics than can distinguish voiceless fricatives /f, θ, x/ from the voiced approximants [ß̞, ð̞, ɣ̞] of voiced stops /b, d, g/. In particular, they focused on voicing, noise, and duration as possible influential factors. They found that voicing and noise are not reliable factors to distinguish these sounds; on the contrary, duration counts as the key factor for distinction. In this case, the authors also equate duration and tension. On this aspect, Martínez Celdrán and Fernández Planas (2007) disagree with the notion of duration equated to tension, claiming that tension is not the product of duration but rather an increase in the tension is what leads to longer duration.

Likewise, English /s/ also shows a large amount of energy at high frequencies (Ladefoged & Ferrari Disner, 2012), extending over 10 000Hz and with little energy below 3500 Hz. /ʃ/, in turn, concentrates energy around 3000 Hz, and thus is lower in pitch than /s/. On the contrary, /f/ and /θ/ show energy over a range of frequencies, i.e., greater dispersion, with higher concentration of energy around 3000-4000 Hz for the former and above 8000 Hz for the latter. Their voiced counterparts /z, ʒ, v, ð/, respectively, have less intensity because the movement of the vocal folds to produce voicing diminish the airstream that escapes the mouth. Nevertheless, they have similar energy distributions to their voiceless equivalents. As Ladefoged and Maddison state "the greater frequency of voiceless fricatives in the world´s languages may be due to the fact that the strong low frequency energy that results from voicing tends to mask the lower amplitude frication noise in the higher frequency range" (p. 176).

The question of how to measure and distinguish fricatives using acoustic parameters has been long held. Studies by Jassem (1979), Forrest, Weismer, Milenkovic, and Dougal (1988), and Wrench (1995) point at a better discrimination of sibilant fricatives than non-sibilant fricatives. However, as criticized by Shadle and Mair (1996), "none of these studies has used spectral analysis above 10kHz" (p. 1521). This is a critical question, since some differences can be found between the front fricatives at frequencies higher than 10 000 Hz (Jongman & Sereno, 1995; Shadle, Mair, & Carter, 1996). Nevertheless, even when using, 16 950 Hz as the maximum range for their data, they came to the conclusion that spectral moments are not reliable for the distinction of the English front fricatives [f, θ, s, ʃ].

Jongman, Wayland, and Wong (2000) conducted a large-scale study for the classification of place of articulation in English fricatives in terms of spectrum, amplitude and duration, and the location of these properties along the sound. Alveolar sounds /s, z/, which are articulated at the teeth, show a high-frequency turbulence and a primary spectral peak at higher-frequencies than the other fricative sounds. Labiodental /f, v/ and dental /θ, ð/ do not display a particular highest peak at any given moment. /ʃ, ʒ/ generally show a peak which coincides with the F3 of the following vowel. In this sense, "spectral moments have not been shown to reliably differentiate the nonsibilants" (p. 1254). In their study, however, they found that spectral cues could not only differentiate sibilant from non-sibilant fricatives, but also

/s, z/ from /ʃ, ʒ/, and /f, v/ from /θ, ð/. In terms of amplitude, sibilant fricatives showed higher amplitude than non-sibilant fricatives (/ʃ, ʒ/ > /s, z/ > /f, v/ > /θ, ð/). As to relative amplitude, "defined as the difference between fricative and vowel amplitude in the F3 region for sibilants" (p. 1254), also served to distinguish place of articulation, with the highest relative amplitude for /s, z/, which show their peak above the F3. Noise duration is generally longer in sibilant fricatives than in non-sibilant fricative sounds, and longer in voiceless fricatives than in voiced fricative sounds. While they found that normalized duration could distinguish /s, z/ from /ʃ, ʒ/, it failed to distinguish the other two pair of sounds, which leads the authors to conclude that duration is not a reliable measure to distinguish the place of articulation of fricative sounds.

In their cross-language study of voiceless fricatives in seven languages, Gordon, Barthmaier, and Sands (2002) also found that duration was not a strong parameter to distinguish fricative sounds. What they found is that the place where the constriction occurs is relevant. The further back the constriction is located, as was the case of the velar /x/, the lower the frequency of the fricative. In general, despite variability of cues, spectrum proved to be a reliable parameter for the discrimination of fricatives. Nevertheless, the methodology of this study was questioned by Boersma and Hamann (2008), stating that they "apparently used the incorrect method of Ladefoged (2003), which weighs the frequencies by their intensity values in dB and is therefore sensitive to arbitrary recording settings" (p. 229).

2.1.2.1 Fricatives after WAS aspiration

As described in Section 1.2.1 (Alvar, 1996; Jiménez Fernández, 1999), WAS voiced stops are generally fricatized after the aspiration of /s/ in final position. When aspiration precedes voiced stops, they become fricatives, with a subsequent change in their place of articulation. Bilabial /b/ becomes labiodental [v/f], dental /d/ becomes interdental [ð/θ], and velar /g/ remains velar or becomes glottal [x/ɦ]. Additionally, Spanish voiced approximants [β̞, ð̞, ɣ̞] are allophones of the voiced stops /b, d, g/ in word-initial position preceded by vowel or /s/, which is true not only for CS after /s/, but also for CS and WAS after vowel. Martínez Celdrán and Fernández Planas (2007, p. 208) argue that, unlike fricatives, "*su intensidad es relativamente débil, comparada con las vocales vecinas, y su duración es bastante breve*" ["their intensity is relatively weak, compared to the neighboring vowels, and their duration is rather brief."] When producing approximants, the articulators have a less strict position than the one needed to produce fricatives, given that the tension to produce fricatives is much higher than the one needed to produce approximants (see also Martínez Celdrán, 2004).

Observe Figures 6, 7 and 8, where we can see a clear transition from the voiceless stop [t][12] to the voiced approximant [ð̞] after sibilance, and finally to the fricative [ð] after aspiration.

[12] See Lisker and Ambramson (1964) for the presence of a voice bar in voiceless stops.

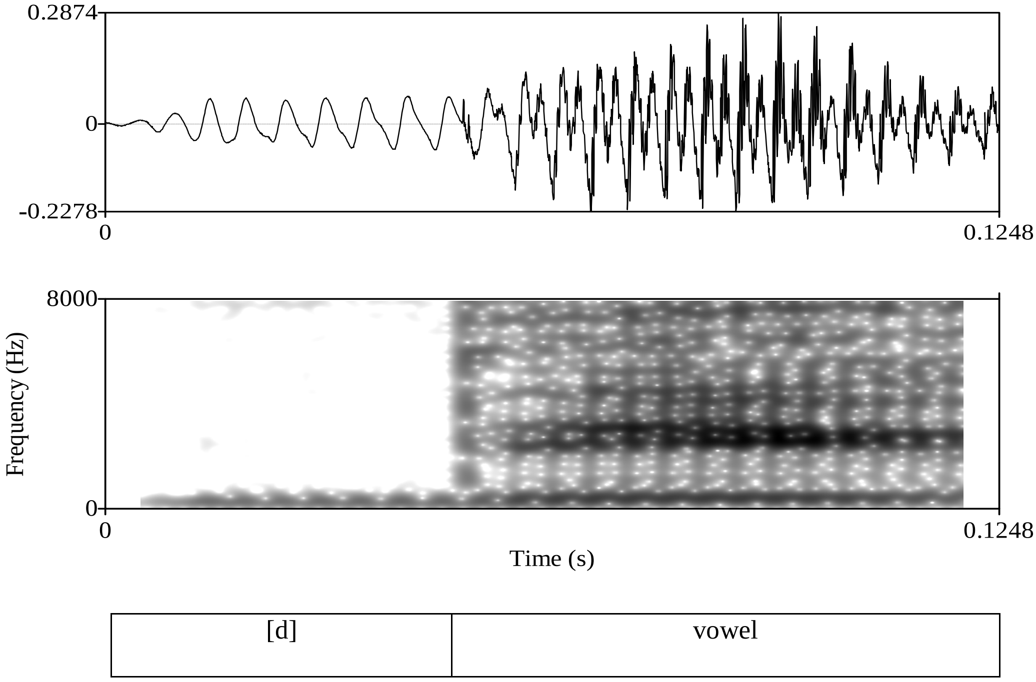


Figure 6. *Word-initial [d] in absolute position*

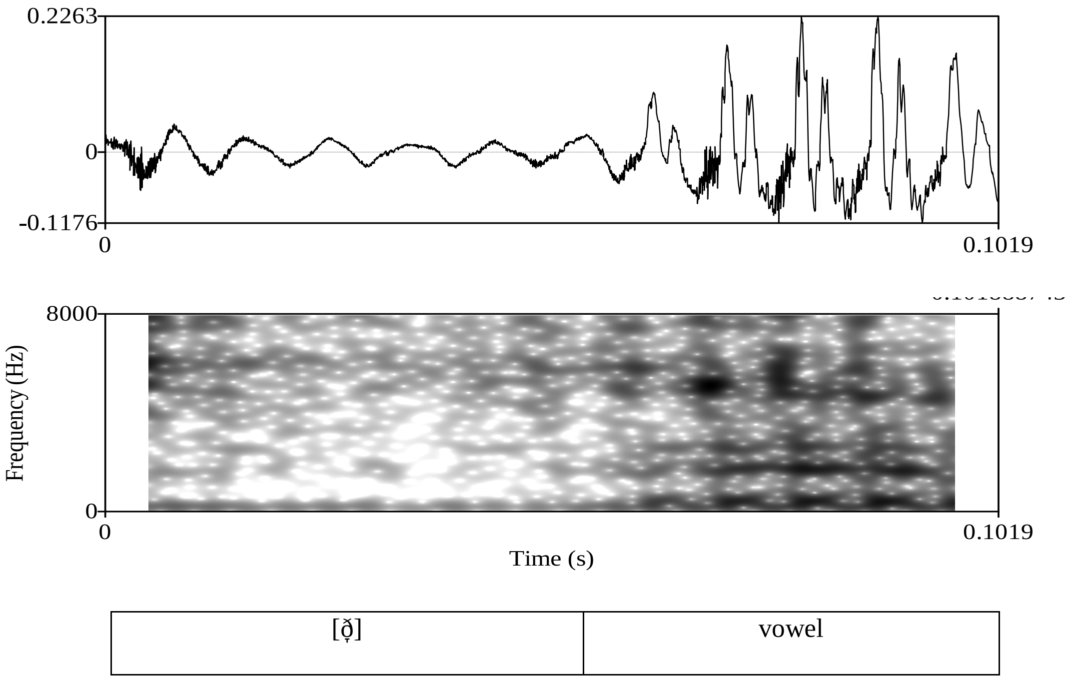


Figure 7. *Word-initial [ð̥] after sibilance*

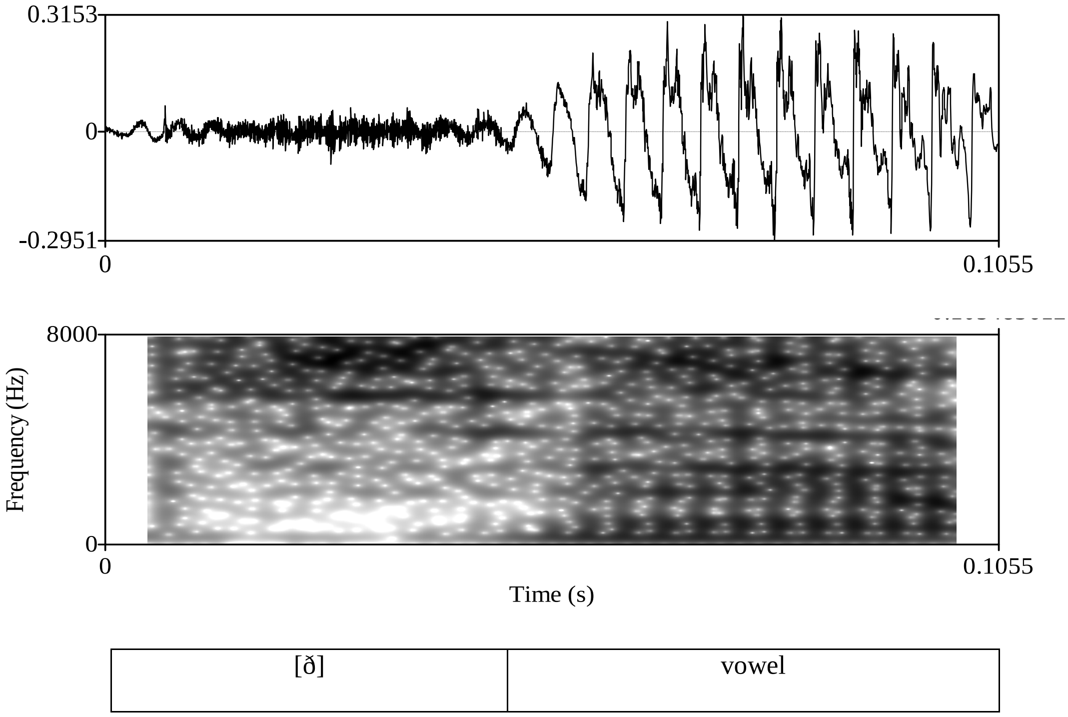


Figure 8. *Word-initial [ð] after aspiration*

To this day, the best-known piece of work related to fricatization of voiced stops after aspiration and to voiced approximants is the study by Romero Gallego (1995). He studied the articulatory gestures that underlie these two sets of sounds in Andalusian Spanish, which are the effect of aspiration and what is known as spirantization, respectively. In terms of manner of articulation, i.e., labial, dental, and velar, he observed that the degree of constriction between the fricatives and the approximants was not different. Instead, primary differences between the two sets of sounds resided in their duration: the fricative sounds were significantly longer than the approximants, for the three manners of articulation.

Martínez Celdrán (2012) observes that there is variability in the degree of constriction of the approximants, although never close enough to cause turbulence. Fricatives, on the contrary, necessarily have the constriction and the tension to cause this noise. Thus, "*la diferencia acústica y perceptiva principal entre fricativa y aproximante consiste en la presencia de turbulencias en la primera, … y su ausencia en la segunda que, por el contrario, presenta estrías regulares de pulsos glotales*" (p. 4). ["the main acoustic and perceptual difference between fricative and approximant consists of the presence of turbulence in the former … and its absence in the latter which, on the contrary, presents regular striation of glottal pulses" (p. 4)].

Figures 9, 10 and 11 below show spectral slices of the three WAS fricatives that result from the aspiration of the voiced approximants.

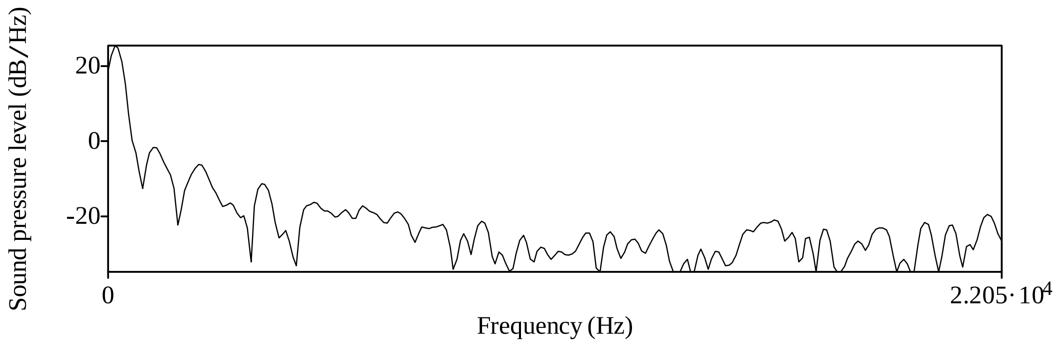


Figure 9. *Spectral slice of [v]*

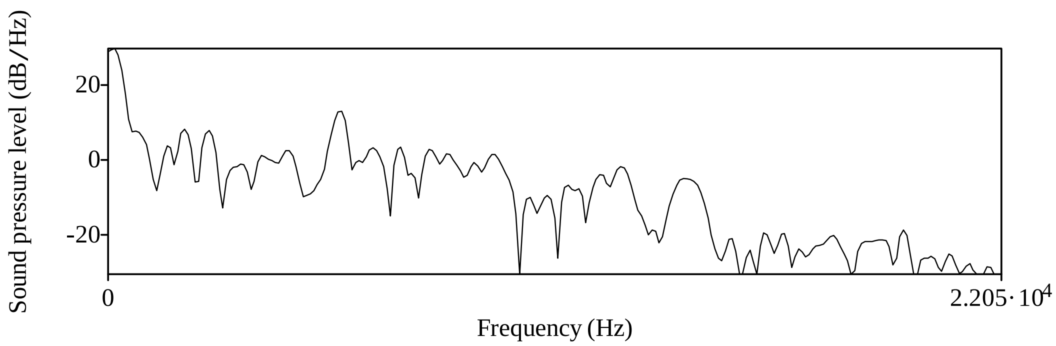


Figure 10. *Spectral slice of [ð]*

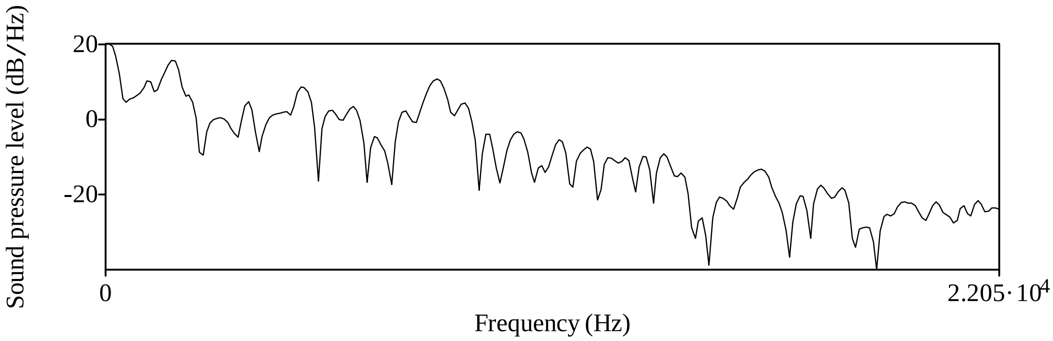


Figure 11. *Spectral slice of [x]*

2.2 Summary

In this chapter, we have focused on one of the branches of phonetics: acoustic phonetics. We have seen how speech sounds can be described in terms of their acoustic properties, particularly as far as stops and fricatives are concerned. Additionally, we have reviewed several studies that investigated the nature of these sounds in WAS, as a result of the aspiration of sibilance given in this dialect. It seems that VOT is a good indicator of the presence of aspiration in WAS voiceless stops, while the spectral moments of fricative sounds have rendered diverse views until Jongman et al.'s (2000) work.

In the following chapter, we cover the area of perceptual phonetics, specifically the area of L2 speech perception, and we explain how the acoustic cues of the sounds, along with the listeners' characteristics, play a role in this process.

CHAPTER 3

L2 SPEECH PERCEPTION

Speech perception in general can be described as the decoding of the acoustic signal in speech into meaningful information for the listener. Native speakers, when processing continuous speech, ignore certain acoustic cues in favor of those that are relevant in their L1, despite age, gender, or rate of speech of the speaker, to “focus on the words being said, and not so much on exactly how they are pronounced” (Johnson, 2012, p. 100). The way we speak guides the way we interpret speech. This leads us to understand sounds according to the language-specific categories that we have learned to use in our L1. Thus, “we hear sounds that we are familiar with as talkers” (p. 107), and our perception is also guided by the linguistic knowledge that we have of our L1, i.e., the phonotactic rules of our native language.

The perception of non-native sounds is said to depend on several factors related to the listener, such as L1, age of learning (AOL[13]), and L2 experience. Initially, L1 listeners will have difficulty with L2 contrasts that are not phonetically contrastive in their L1. Contrasts that are given in the L2 but absent in the L1 may not be distinguished by the listeners. A classic example of this is the perception of English /r/ and /l/ by L1 Japanese listeners as one single L1 category (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975; Best & Strange, 1992; Polka & Strange, 1985 among others). As this contrast is not given in their native language, L1 Japanese listeners are generally unable to distinguish these L2 sounds as separate phonemes. As also found by Flege, Bohn, and Jang (1997), L1 Spanish

[13] This factor will be briefly addressed in Section 3.2.2

listeners in their study assimilated English /i:/ and /ɪ/ to Spanish /i/. Since this contrast is not present in their L1, they matched them to the only phoneme available in their native language. However, English /e/ was assimilated to Spanish /e/ and English /æ/ was assimilated to Spanish /a/, which are two distinct categories in Spanish. L1 Korean and Mandarin listeners also confused /i:/ and /ɪ/, as this contrast is not given in their L1 either. However, that was not the case for L1 German listeners, whose L1 does possess this contrast. Several studies have pointed out at the reliance on durational cues by L1 Spanish listeners in the perception of L2 English vowel contrasts, rather than on spectral cues inexistent in their L1 (Escudero & Boersma, 2004; Escudero, Benders, & Lipski, 2004). This serves as evidence that L1 experience may determine the way certain phonetic cues are used in L2 speech perception. Nevertheless, features that are shared by L1 and L2 on certain segments may not be transferred to new L2 sounds automatically. Consequently, the fact that L1 and L2 share the same features may not necessarily favor perception or learning.

Another factor to be taken into account when examining L2 speech perception is the listeners' experience in the L2, which may lead L2 learners to reorganize their phonetic systems as experience increases. Beginning L2 learners may find difficulties that can be overcome with increasing experience in the language. Bohn and Flege (1990) investigated the perception of English vowels /i:, ɪ, e, æ/ by experienced and inexperienced L1 German listeners. While experience was not an influential factor for the perception of vowels that had similar or identical counterparts in German (/i:, ɪ, e/), it proved to be crucial for the perception of /æ/, which was a new sound for the listeners. The inexperienced listeners performed significantly lower than the experienced listeners in the identification of this L2 sound and seemed to resort to durational cues to distinguish it from /e/ (see also Flege & Liu, 2001; Flege, Takagi, & Mann, 1996 for further effects of experience).

However, some studies have pointed out that experience may not render higher accuracy in some cases. Levy and Strange (2008) found that experience was influential in the perception of L2 French contrasts /u, œ/, /i-y/ and /y-œ/ for experienced and inexperienced L1 American English listeners. However, no differences were found between both groups of listeners for the perception of the contrast /u-y/. Levy (2009) also studied the perception of L2 French vowels by L1 American English listeners with no experience in French, with formal instruction in French, and with formal instruction and immersion in French. She concluded that higher accuracy was found for the most experienced listeners and in bilabial context. In this case, the acoustical similarities between French vowels were not sufficient to explain context-specific assimilation patterns. Instead "it is suggested that native-language allophonic variation influences context-specific perceptual patterns in second-language learning" (p. 1138, see also Levy & Law II, 2010).

To account for these contradictory results, two additional factors need to be taken into consideration in the perception of non-native sounds: the type of contrast under study and the type of acoustic cues of the L2 sounds (Barreiro Bilbao, 2002). Not all contrasts are similarly difficult; other than the L1 background and the L2 experience of the listeners, we should also look at the psychoacoustic salience of the sounds under study, that is, the sounds we perceive and experience as more salient in relation to our physiological capacity (auditory perception) and our phonetic knowledge (categorical perception). As pointed out by Strange and Shafer (2008):

> "… in general, place-of-articulation contrasts in consonants, cued primarily by spectral differences of short duration, may be considered less salient than voicing contrasts, cued primarily by temporal parameters … Contrasts in manner of articulation (e.g., fricative vs. stop) may be considered very salient in that they are differentiated by differences in sound source characteristics." (p. 175)

In a series of studies (Hendrick & Carney, 1997; Hendrick & Younger, 2001; Hendrick & Younger, 2007), the role of relative amplitude and formant transitions of English stops and fricatives in speech perception was investigated. Although the nature of these studies was to investigate perception in hearing-impaired L1 listeners in relation to normal hearing listeners, insightful findings with respect to acoustic cues can be drawn. Studies have shown that in CV sequences manipulating a frequency region of the consonant in the syllable relative to the amplitude of the same frequency region in the following vowel (relative amplitude) influences the perception of place of articulation for fricatives and stop consonants.

In this regard, Chen and Alwan (2003) studied the perception of English stops and fricatives by English L1 listeners in terms of place of articulation: labial /b, p, f, v/ and alveolar /d, t, s, z/, in three vowel contexts /a, i, u/. They found that "the perception of place for plosives and fricatives depends on whether the consonant is voiced or voiceless" (p. 1499), i.e., voiceless consonants were more robust than their voiced counterparts. Later, Alwan, Jiang, and Chen (2011) conducted a similar study in which they found that the identification of the distinction between labial and alveolar stops in noise depends on the manner of articulation and its interaction with voicing.

In a cross-language perception study, Silbert, de Jong, and Park (2005) investigated the perception of English consonants by Korean listeners in terms of voicing, place of articulation (labial/coronal), manner of articulation (stop/fricative), and position in the syllable (initial/final). The Korean language does not have non-sibilant fricatives produced at the front of the oral cavity and neutralizes voicing and manner of articulation in syllable-final position; thus, the identification task tested the effects of L1 specific phonological patterns in the perception of non-native features. The identification of voicing was rather good for labial and coronal stops and fricatives in syllable-initial position, although slightly worse for labial fricatives. In both cases, there was a bias towards voiceless classification. In syllable-final position, they exhibited a poor performance in the identification of voiced labial stops and coronal fricatives, with a tendency to identify the fricative sounds as voiceless sounds. In terms of poor performance, it seems that "being a fricative and being coronal both increase the likelihood that the listeners will call a segment voiceless" (p. 13), resulting in the perception of consonant noise of voiced and voiceless fricatives in a similar way.

A study by Wagner, Ernestus, and Cutler (2006), focused on the role of L1 fricative inventory in the identification of L2 fricatives. They studied how listeners of German, Dutch, English, Spanish, and Polish identified spectrally similar fricatives /θ/ and /f/ in terms of formant transitions with and without manipulation. Since German and Dutch do not have spectrally similar fricatives, they were not affected by the changes in transitions, while listeners of the remaining three languages did. Their conclusion is that all listeners "may be sensitive to mismatching information at a low auditory level, but that they do not necessarily take full advantage of all available systematic acoustic variation when identifying phonemes" (p. 2267).

In a similar study, Cutler, Cooke, Garcia Lecumberri, and Pasveer (2007) investigated the identification of GAE consonants in noise by native listeners, and Spanish and Dutch listeners. With respect to fricatives /f, θ/, due to the similarities of their native inventory, Spanish and English listeners used the same cues, while Dutch listeners deviated more from native performance. Nevertheless, in the presence of noise, when transitional cues are difficult to distinguish, both English and Spanish listeners' identification was affected negatively. In this case, the performance of Dutch listeners was not so affected because they did not rely on formant transition information in the first place, "but relied on the steady-state information in the fricative noise" (p. 1588).

Similar results were found in Barreiro Bilbao (1999), who also researched the perception of fricatives by L2 listeners. In particular, she studied the effect of voicing and place of articulation in the categorization of two English contrasts that are not present in Spanish, that is, /s, z/ and /s, ʃ/. For the first pair of sounds, when the voice bar was removed, the results were random. Thus, Spanish listeners made use of voice to distinguish these two sounds. In the case of the second contrast, Spanish listeners relied on the frequency and amplitude of the fricatives, and not on the F2 transitions, just as Dutch listeners did in the study described above.

Considering all the factors and the research mentioned above, there is still one more aspect of L2 perception that needs to be explored. Most research covers the perception of categorical sounds of mainstream languages; next section covers research concerning the perception, categorization, and identification of dialectal variations of a language.

3.1 Perception of L2 Dialect Variants

Research indicates that categorization and discrimination varies across L2 contrasts and across L1s. L2 learners' perception of L2 contrasts systematically depends on the phonotactic, allophonic, and coarticulatory patterns of their L1. Moreover, highly relevant for this dissertation is the assertion that not only does the L1 of the listener have an effect on the perception of a given L2 sound or contrast, but also "L1 and L2 dialect differences can both systematically affect perception of L2" (Best & Tyler, 2007, p. 19).

This is why, when encountering an unfamiliar L1 dialect, perceptual learning may need to take place. Studies show that preference is given to unmarked dialects or mainstream varieties of a language. Clopper and Bradlow (2006, 2008) studied the intelligibility of dialect variation in noise. In favorable conditions, GAE and Southern English were better identified than Northern and Mid-Atlantic English. However, in unfavorable noise conditions, the intelligibility of GAE was greater than that of the other three dialects, suggesting that dialect information may be conveyed by aspects of the signal that are relatively vulnerable to perceptual disruption by noise.

Sumner and Samuel (2009) also demonstrated this higher accuracy in identification of mainstream features of a language. Furthermore, they also found that being familiar with a dialect renders greater identification of its features. They researched word recognition in dialectal variation and the role of experience in perception and representation. With a series of tasks involving priming they targeted the perception and production of r-dropping in New York City (NYC) dialect, opposed to GAE full realization of *–er* > [ɚ]. Listeners in this study were i) speakers of NYC dialect, ii) speakers familiar with the dialect, and iii) speakers of GAE unfamiliar with the dialect. They came to the conclusion that dialect production is not always representative of dialect perception and representation; listeners familiar with but not speakers of NYC dialect performed similarly to speakers of the dialect in perception tasks. Thus, experience seems to strongly affect a listener's ability to recognize spoken words, although variants that are not regionally-marked are preferred overall.

If we take this to the domain of L2 acquisition, differences in phoneme inventory between L1 and L2 pose a higher difficulty than L1 differences; L2 learners require exposure to, training in, and use of the L2 to attain the new features. One of the most recent works on L1 cross-dialectal perception is the study by Tuinman et al. (2011), which focused on the perception of British English intrusive [ɾ] by speakers of American English, who accurately perceived vowel-initial words despite intrusive [ɾ]. Nevertheless, these results are in contrast with the findings for the same materials presented to proficient L2 listeners (Tuinman et al., 2007), whose responses showed that they perceived intrusive [ɾ] as word-initial /r/. Although L1 dialect variation is not equivalent to L1-L2 differences, the results broadly showed a robust ability by L1 listeners to adjust to variation within the same language.

A study by Cutler, Smits, and Cooper (2005) had also explored this dialect variation within the same L1 with the addition of subjects from an L2. They studied the identification of American English vowels in open and closed syllables by speakers of American English, Australian English, and Dutch. Both groups of English speakers clearly outperformed Dutch speakers; nevertheless, vowel tenseness judgment was more variable for Australian English speakers due to cross-dialectal differences. When speech input mismatches the native dialect, the difficulty is very much less than that which arises when speech input mismatches the native language in terms of the repertoire of phonemic categories available.

When we move towards L2 perception by listeners of different L1 dialects, we find works such as that by Chádková and Podlipský (2011), who studied the perception of Dutch /i:, ɪ/, characterized by spectral differences, by listeners of two dialects: Bohemian Czech and Moravian Czech, which have the same contrast. The first one is also based on spectral differences whereas the second one is based on durational differences. As predicted, Bohemian Czech speakers assimilated the Dutch contrast to two L1 categories while Moravian Czech speakers assimilated the L2 contrast to a single category, /ɪ/, supporting the claim that different L1 dialects can render different assimilation patterns of the same nonnative contrast.

More recently, Escudero and Williams (2012) studied the perception of Dutch vowels by speakers of Peruvian Spanish (from Lima) and Peninsular Spanish (from Madrid), whose results indicate that acoustic differences in the native dialect can be more influential than proficiency in the L2. Peninsular Spanish speakers outperformed Peruvian Spanish speakers despite being less proficient in Dutch. Therefore, experience in this case does not seem to be most relevant for perception; results show that L1 dialect prevails.

Moving towards our dialects under study, we found that research on AAE has been especially directed towards the description of the language in fields such as variation and change, grammar, phonology, lexicon and use, ethnic identity, education, origins and history, and recently hip hop culture (Alim, 2004; Baugh, 2000, 2004a; Billings, 2005; Fasold, 1972; Green, 2004b; Morgan, 2001; Mufwene, 2003; Poplack, 2000; Spears, 2001; Wolfram et al., 2001; Zeigler, 2001, among others). In the field of speech perception and production, research on AAE has traditionally focused on its implications for education, particularly for reading and writing among AAE-speaking children. In any case, research is mainly restricted to

L1 studies (Craig & Washington, 2004; Laing, 2003; Seymour, Bland-Stewart, & Green, 1998; among others).

Felder (2006) studied the perception of final voiced and voiceless stops produced by AAE speakers, by both experienced AAE listeners and GAE listeners with little or no AAE experience. The words were presented in two medial sentence contexts, followed by either a vowel or by the voiceless fricative consonant /f/, and subjects were given response alternatives. Both AAE and GAE groups performed similarly, identifying final voiced stops as consonant deletion or as voiceless stops. Gender of speaker was also influential; listeners perceived the female speaker to devoice stops in both contexts while they perceived the male speaker to delete final stops in /f/ sentence context. The author points at the inability to determine if these findings are the result of individual differences, a reflection of AAE rules, or gender-based differences.

Previously, Felder and Strange (2000) had studied the discrimination of AE contrasts between /θ, ð/, which do not occur in Haitian Kreyol, and /t,d/ or /f,v/ in initial, intervocalic, and final position by bilingual speakers of Haitian/AAE and dialectal speakers of AE/AAE. Haitian speakers substitute /t, d, v/ for /θ, ð/ whereas AAE speakers' realizations vary according to context. Results indicate that perceptual errors are related to the substitutions and realizations characteristic of each variety and thus dependent on L1 constraints.

Consistent with these results are the findings in Sligh and Conners' (2003) study of the relation of dialect to phonological processing and its implications for reading. They tested GAE and AAE speakers on the completion of word-initial and word-final consonant clusters in which one of the members of the cluster was deleted. As predicted, GAE listeners outperformed AAE listeners in word-final clusters while

AAE listeners performed better in word-initial deletion, probably as a result of AAE reduction of final clusters in speech (see also Kile, 2007).

Studies on the perception of dialectal variation in Spanish are not abundant (see Boomershine, 2006; Díaz-Campos & Navarro-Galisteo, 2009; Face & Menke, 2009; Rose, 2010), although there is an increasing interest in the perception of the sociophonetic variants of this language. Even when aspiration of implosive /s/ is reported in diverse Spanish dialects, the study of the perception of this feature seems to be limited, especially by L2 learners.

Perhaps the most relevant research work for this dissertation is the study by Schmidt (2011) of the aspiration of implosive /s/ in citation-form words of Argentinian and Venezuelan Spanish by speakers of GAE, in relation to their level of proficiency (levels 1-5). Although the object of her study is a different dialect than WAS, the one that we used for this dissertation, the feature of implosive /s/ aspiration is given in both dialects. Results indicate that not until level 3 do listeners start to identify this dialectal feature and not until level 5 do listeners perform similarly to native speakers of the dialect. The author (Schmidt, 2009) previously found that familiarity with a Spanish dialect increases the identification accuracy of dialectal features, although the unmarked dialect still renders better results, as mentioned above.

How can these results be accounted for? Current models of cross-language speech perception attempt to explain how non-native sounds are perceived by speakers with or without experience with the language. In the following section, we revise the main theories of these models with the purpose of applying them to our current study.

3.2 Models of L2 Speech Perception

3.2.1 Native Language Magnet Model

Developed by Kuhl (1992, 1993a, 1993b, 1994) and Kuhl and Iverson (1995), the Native Language Magnet (NLM) model is based on the premise that “exposure to language early in life produces a change in perceived distances in the acoustic space underlying phonetic distinctions, and this subsequently alters both the perception of spoken language and its production” (Kuhl & Iverson, 1995, p. 122). Primarily intended to account for L1 speech perception in infants’ first year of life, before their acquisition of lexicon and contrastive phonology, the implications of this model can also be applied to adult L2 speech perception.

The organization of phonetic categories seems to be around prototypes, i.e., good exemplars of these categories, which act as perceptual magnets for the rest of the sounds in their category, attracting them and reducing the perceptual distance between them (Iverson & Kuhl, 1996). This magnet effect is species specific and, by 6 months of age, it is affected by exposure to a given language, giving way to the *warping* of the acoustic space underlying phonetic perception, which increases with language exposure. Thus, as perceptual distance around a prototype is reduced, discrimination sensitivity to acoustic differences close to the prototype is also reduced.

While infants are able to discriminate pairs of L1 phonetic segments across boundaries but rather fail to discern differences between phonetic units within a given phonetic category, this ability to identify changes in category “plays a role in infants’ abilities to organize their category representations” (Kuhl & Iverson, 1995, p. 139). Indeed, these general auditory processing mechanisms that identify perceptual boundaries are said to be the base for language-specific magnet effects (for evidence against the magnet effect, see Frieda, Walley, Flege, & Sloane, 1999; Sussman & Lauckner-Morano, 1995; Thyer, Hickson, & Dodd, 2000).

In the field of adult L2 acquisition, NLM posits that a sound in the L2 that is similar to a sound in the L1 will be difficult to identify as different while those that are different will be easily discriminated. In other words, the closer an L2 phonetic unit is to a prototype in the L1, the more it will be assimilated to and undistinguishable from it. Nevertheless, the model points at training as a method to increase discrimination of L2 contrastive sounds. Iverson and Evans (2009) studied the interference of L1 German (18 vowels) and Spanish (5 vowels) in the perception and learning of English vowels. German subjects outperformed Spanish listeners after five training sessions; however, with ten additional sessions, both groups performed similarly and were able to retain the information learned.

Training seems to involve changes at a higher order level, which implies that listeners also draw on memory and attention. When attention is directed towards sound cues that are relevant to perform the categorization task, the distance between the tokens is said to be increased, whereas the distance along irrelevant cues is reduced. As far as memory is concerned, alterations in the task that involve memory load do not seem to affect the influence of the prototypes.

How are these representations stored in memory? NLM offers two possibilities; namely, either as individual instances or as abstract summaries of these instances, but does not choose one explanation. What the model posits is that "speech representations are initially auditory, but they become polymodal as infants acquire information (both visual and motor)" (Kuhl & Iverson, 1995, p.147). This notion of articulatory dimensions is also supported by Best's (1995) and Strange's (2011) models of cross-language perception

3.2.2 Speech Learning Model

The Speech Learning Model (SLM) was developed by Flege (1995) with the aim to understand how speech learning works in relation to age of learning (AOL) and particularly to L2 production and ultimate attainment of L2 pronunciation. It focuses on experienced listeners (i.e. bilingual speakers), from infants to adults, and postulates that language-specific aspects of L1 speech segments are stored in phonetic categories within our long-term memory, the processes of which can also be applied to L2 learning. The model also posits that the existent categories and the new ones that are formed evolve throughout our life span to reflect L1 and L2 sounds which coexist in a common phonological space and change as L2 experience broadens.

In this respect, this model rejects the assumption that errors in production are caused after a critical period for speech learning based on neurological maturation. Instead, it points at perceptual errors as a common denominator in production errors and accented speech. In fact, Flege, Freida & Nozawa (1997) conducted a pioneer study in which bilinguals' amount of L1 use was found to influence accented speech more than AOL (see also Flege & MacKay, 2004; Flege, MacKay, & Piske, 2002, Flege, Schirru, & MacKay, 2003).

The principles governing L1 acquisition may not be equally applicable to L2 learning but, as mentioned above, they remain intact throughout life. Experienced listeners will perceive L2 sounds in terms of their L1 sounds (at first); therefore, their perception will not be the same as that of native listeners'. This does not imply that an L2 learner cannot establish further L2 categories; as learners' experience with the L2 increases, so do their chances to discriminate similar L1-L2 sounds and establish new L2 categories independent of L1 representations (see studies that are contradictory with this notion: Levy, 2009; Levy & Strange, 2008).

SLM's postulates are as follows (Flege, 1995, p. 239):

P1 The mechanisms and processes used in learning the L1 system, including category formation, remain intact over the life span, and can be applied to L2 learning.

P2 Language-specific aspects of speech sounds are specified in long-term memory representations called *phonetic categories*.

P3 Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category.

P4 Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space.

SLM focuses on the phonetic level under the assumption that L1 and L2 sounds are related at a position-sensitive allophonic level rather than at a phonemic level. This model also assumes a bidirectional L1-L2 interference by which sounds in both languages linked to one another influence one another (see H6 below), in agreement with Grosjean's (1998) claim that the two language systems of a bilingual's are constantly engaged.

The hypotheses are the following (Flege, 1995, p. 239):

H1 Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level.

H2 A new phonetic category can be established for an L2 sound that differs phonetically from the closest L2 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds.

H3 The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.

H4 The likelihood of phonetic differences between L1 and L2 sounds, and between L2 sounds that are noncontrastive in the L1, being discerned decreases as AOL increases.

H5 Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones). Eventually, the diaphones will resemble one another in production.

H6 The phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if: 1) the bilingual's category is "deflected" away from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's.

H7 The production of a sound eventually corresponds to the properties represented in its phonetic category representation.

This mechanism of *equivalence classification* seen in H5 is a process by which an L2 sound can be perceived as identical, similar, or new with respect to an existing L1 sound. The L2 sound will be assimilated to the L1 sound if it is perceived as identical or similar, whereas a new category will be formed for the L2 sound if it is perceived as less similar or new (however, it is unclear what the terms 'similar' and 'less similar' exactly refer to.)

Concerning the perception of non-native contrasts, SLM predicts that if two contrasting L2 sounds are perceived as similar to one L1 sound, then discrimination will be difficult. At the same time, if one of the L2 sounds is dissimilar to any L1 sound, then equivalence will not take place and a new phonetic category will be likely formed, so both perception and production can be carried out relatively accurately. Therefore, "the greater the perceived distance of an L2 sound from the closest L1 sound, the more likely it is that a separate category will be established for the L2 sound" (Flege, 1995, p. 264).

3.2.3 Perceptual Assimilation Models

The Perceptual Assimilation Model (PAM), developed by Best (1995), focuses primarily on the perception of nonnative sounds by naïve listeners (i.e. monolingual) with no experience in the L2. This model presents a direct-realist view of speech perception based on gestural information which, unlike SLM, "is *not* built up from an analysis of simple acoustic features" (Best, 1995, p. 177) but detected from speech directly and actively by means of integrated perceptual systems. L2 sounds "tend to be perceived according to their similarities to, and discrepancies from, the native segmental constellations that are in closest proximity to them in native phonological space" (Best, 1995, p. 193).

Monolingual speakers can not only distinguish phonemes but also within-category phonetic variations, rating them as good or poor exemplars of the category. This idea reflects the notion of *warping* that we have seen in NLM. According to PAM, assimilation of an L2 phone can follow any of these three patterns: i) the L2 phone can be assimilated to an L1 category as a good exemplar, an acceptable exemplar, or a deviant exemplar of that category; ii) the L2 phone can be classified as uncategorizable, i.e., recognized as speech but not an exemplar of any given L1 category; and iii) the L2 phone may not be assimilated to speech.

Additionally, the model establishes six possible types of perceptual assimilation for nonnative contrastive sounds that differ in terms of difficulty: i) if the contrastive L2 sounds are assimilated to two different L1 categories (Two Category or TG type), then discrimination will be excellent; ii) if the contrastive L2 sounds are assimilated as equally acceptable or equally deviant exemplars of one single L1 category (Single Category or SG type), discrimination will be difficult (above chance level); iii) if the contrastive L2 sounds are assimilated to one single L1 category but their goodness to fit differs (Category Goodness or CG type), discrimination will be moderate to very good. Additionally, iv) when one of the L2 sounds is not perceived as similar to any L1 category (Uncategorized-Categorized or UC type), discrimination is expected to be very good.; v) if none of the L2 sounds are assimilated to any L1 category (Uncategorized-Uncategorized or UU type), discrimination will range from poor to very good; finally vi) if the L2 sounds are so different than any L1 sound that they are not perceived as speech at all (Non-Assimilable or NA type), discrimination will range from good to very good (for a study in which a revision of the UC type is suggested, see Guion, Flege, Akane-Yamada, & Pruitt, 2000).

3.2.3.1 Perceptual Assimilation Model-L2

Furthermore, Best and Tyler (2007) developed the PAM-L2 to explain speech perception by late L2 learners and to additionally review SLM from PAM's perspective. We must take into account that by the term *L2 learner,* they understand "people who are in the process of actively learning an L2 to achieve functional, communicative goals, that is, not merely in a classroom for satisfaction of educational requirements" (p. 16).

On the one hand, the problem these authors see with a foreign language acquisition (FLA) environment is the L1-accented input that learners may receive along with the different dialectal varieties of the L2 language which can interfere with perception. In addition, a further limitation is the usual scenario of FLA being an educational requirement and not a process of active learning to achieve communicative and functional skills, as opposed to SLA learners. On the other hand, unlike naïve speakers, FLA learners are exposed to the L2; thus, the authors encourage research on perceptual adjustment to L2 contrasts in FLA settings as opposed to SLA contexts, which is what we did in this dissertation.

Whereas its predictions of the perceptual assimilation of L2 contrasts by experienced listeners are similar to those posed about equivalence classification in the SLM and perceptual assimilation by naïve listeners in the PAM, the three models differ in one key aspect: PAM-L2 adds the phonological level of both L1 and L2 to judgments of L1-L2 similarity and dissimilarity; thus, perceptual assimilation can occur at the phonological, phonetic, or gestural/articulatory level.

This addition stems from the inclusion of L2 learners into this model who, unlike naïve listeners in PAM, have knowledge of the phonetic and phonological aspects of their L2. At the same time, this knowledge depends on their developmental stage and lexicon[14] acquired, making the phonological level a lexical-functional one where "listeners may identify L1 and L2 sounds as functionally equivalent (assimilated phonologically)", which does not necessarily imply that "the associated phones are perceived as identical at the phonetic level" (Best & Tyler, 2007, p. 26). For example, such is the case of French /r/ > [ʀ], which American English learners of French assimilate to English /r/ > [ɾ] at a functional level.

Late L2 learners, like naïve speakers, may also present difficulty in assimilating L2 contrasts which are not distinctive in their L1, especially if they have limited experience with the L2. However, as experience and familiarity with the L2 increases, so does the perception and production of the L2. PAM-L2 enumerates the following four possibilities for the perception of L2 contrastive sounds (Best & Tyler, 2007, pp. 28-30):

1. Only one L2 sound is perceptually assimilated to a given L1 phonological category, as in UC type. In this case, discrimination will have little difficulty. Alternatively, there exists the case in which the learner perceives an L2 sound as phonetically deviant from their L1 sound but yet phonologically and phonotactically similar on a lexical and functional level, and thus equates them phonologically.

2. Both L2 sounds are perceived as equivalent to the same L1 phonological category, but one is perceived as being more deviant than the other. This instance corresponds to the CG assimilation contrast. The good exemplar will be assimilated to the L1 category while it is estimated that, with L2 experience, the deviant exemplar

[14] PAM-L2 considers that perceptual assimilation is more likely to succeed for listeners with limited L2 vocabulary; otherwise, incomplete perceptual learning before vocabulary expansion may give way to fossilization.

can move from a perceived phonetic variant of the good exemplar to a new phonological category.

3. Both L2 sounds are perceived as equivalent to the same L1 phonological category, but as equally good or equally poor examples of that category. In this case, it is an SC assimilation type, in which both L2 sounds will be assimilated to the L1 category and discrimination will be difficult.

4. No L1-L2 phonological assimilation. In this case, the L2 sounds will be uncategorized by the listener if they cannot be assimilated to any L1 phoneme but rather share characteristics of several L1 phonological categories.

One limitation that the authors point out is that "some aspects of sensitivity to phonetic variation are related to similarities between nonnative stimuli and native speech patterns, but others reflect language-universal perceptual tendencies. The implications of these experience-tuned vs. universal phonetic sensitivities have not yet been fully resolved" (Best & Tyler, p. 18). We will see how Strange (2011) addresses this issue in the next section.

3.2.4 Automatic Selective Perception Model

As a consequence of the models described in the previous sections, Strange (2006, 2011) developed the Automatic Selective Perception (ASP) working model to determine the mechanisms of speech processing that take place in the perception of L1 and L2, using neurobiological studies for the purpose. The focus is on adult naïve L1 listeners -category that also comprises beginning L2 learners- and on late L2 learners residing in a non-native country.

Much like PAM, ASP is based on the direct-realist, ecological view of speech perception as "a purposeful, information-seeking activity whereby adult listeners detect the most reliable acoustic parameters that specify phonetic segments and sequences in their native language (L1)" (Strange, 2011, p. 456). By this mechanism, adult L1 speakers resort to what she terms *selective perception routines* (SPRs) to detect relevant information for recognizing phonological sequences in their L1, which become automatic with the mastery of the language. In contrast, late L2 learners "must employ greater attentional resources in order to extract sufficient information to differentiate phonetic contrasts that do not occur in their native language" (p. 456). Therefore, L1 interference with L2 perception is seen as the attunement of L1 SPRs to the incorrect information in the L2 input.

In this model, two modes of perception are described: the *phonological mode* and the *phonetic mode.* "These are "ways of perceiving" determined by an interaction of the listeners' knowledge, purpose and intentions, the complexity of the stimulus materials, and the demands of the task to be accomplished" (Strange, 2011, p. 460). The *phonological mode* is employed by adult listeners to process continuous L1 speech, whether by speakers of the same variety or of dialects of the language familiar to the listener. The context-dependent phonetic variations are ignored in favor of the semantic message of the utterance, using automatic and robust SPRs even in non-optimal conditions. The *phonetic mode,* on the other hand, is context-dependent and implies attentional focus to allophonic details and to those phonetic and phonotactic patterns necessary in their native dialect or language. It is also slower and may suffer in non-optimal conditions. Strange, Bohn, Trent, and Nishi (2004) and Strange, Bohn, Nishi and Trent (2005) studied the perceived similarity of German [u:] and [y] to American English vowels by naïve speakers of American English. Overall, the two vowel sounds were assimilated to their L1 [u:]; however, in citation-form /hVp/ contexts, [y] was classified as a poorer example of L1 [u:], while in sentence-embedded /bVp/, /dVp/ and /gVk/ contexts, both German sounds were seen as good exemplars of L1 [u:], most likely because American English back rounded vowels are fronted in these contexts and become more similar to German front rounded vowels.

Perception also depends on the design of experiment tasks: auditory salience[15] and perceptual salience[16] of the L2 sounds, memory and attention of listeners can all be targeted by the manipulation of the stimulus materials and the type of task employed in the experiment (see the Tetrahedral Model for Speech Perception Experiments by Strange, 1992).

When the task and the stimuli are simple (citation words) and instructions direct listeners to pay attention to certain phonetic aspects, both naïve listeners and L2 learners can distinguish non-native L2 contrasts and determine similarities and dissimilarities between L1 and L2 sounds. However, as the complexity of the task increases, e.g. listeners must understand the semantic message of the utterance, so does the cognitive demand, and performance may suffer as listeners may resort to their L1 SPRs. Indeed, "as the complexity of the discrimination task increases, performance outcomes begin to reflect not only basic auditory sensory capabilities but increasingly the cognitive processes involved in categorization (including implicit labeling of presented stimuli)" (Strange & Shafer, 2008, p. 161).

Even when listeners have enough experience in the L2 to have established L2 SPRs, these still may not be as automated as their L1 SPRs as "immersion experience alone may not be sufficient for L2 learners to develop and automate these SPRs" (Strange, 2011, p. 464). Instead, training for L2 learning is suggested as it can lead to

[15] "The magnitude of the obligatory physiological response to a change from one to another contrasting lexical segment, tone, or sequence of segments in a normal hearing listener" (Strange, 2011, p. 458).
[16] "Behavioral and physiological response "strength" that varies as a function of linguistic experience, as well as experimental manipulations of attentional focus" (Strange, 2011, p. 458).

the development of new SPRs to improve the detection of the most reliable cues in the L2.

3.3 Current Study

What about the perception of two dialect allophones of the same phonological category? Initially, native speakers familiar with both L2 variants would assimilate both allophones to the same category (SG type) while native speakers unfamiliar with one of the L2 dialects would also assimilate both allophones to one single category but with differences in goodness-of-fit (CG type). As we saw at the beginning of this chapter, preference is given to the unmarked features of a language; thus, the marked allophone would be perceived as more deviant than the unmarked one. Native listeners may successfully discriminate two allophones "when the experimental task allows reliance on pre-existing mental representations of sounds" (Celata, 2007). Nevertheless, the perception of an allophonic contrast is generally less accurate than the perception of a phonemic contrast (Boomershine, Hall, Hume, & Johnson, 2008). The key point is that, in both types of assimilation, PAM and PAM-L2 consider the two contrastive sounds to be phonologically distinctive, but fail to determine how perception is carried out when the two sounds are allophones of one single category. The question pertinent to this study is how L1 listeners identify two dialect variants of the same L2 category, one of which is unfamiliar to them.

The studies reviewed in this chapter suggest that the L1 dialect of the listener exerts a great influence on their discrimination and categorization of L2 segments. Thus, this study tested the perception of two dialect variants of implosive /s/ in Spanish, namely, aspiration [h] found in WAS and sibilance [s] characteristic of CS, by native speakers of two American English dialects, GAE and AAE, whose L1 dialects differ in the use of final /s/ as a marker of plurality and verb agreement.

3.3.1 Predictions and Research Questions

Listeners in this study are L2 learners of Spanish who, even at the elementary stage, are presumed to know that /s/ is phonologically distinctive in the L2 as it differentiates plural nouns from singular nouns and second-person verbs from third-person verbs in the present tense. What they ignore, especially when contact with an aspirating dialect has never occurred, is that [h] is a legitimate allophone of /s/ in certain Spanish dialects and it marks the same distinctions as [s].

A similar sound to the allophone under study [h] occurs in English as a contrastive sound in initial position but not as a legitimate variant of /s/ in implosive position, as is the case in WAS (and other varieties of Spanish). Even when aspirated /s/ and English [h] are acoustically and articulatorily similar to each other, listeners may not assimilate these two sounds, precisely due to the phonotactic biases of their L1.

Can these listeners extract enough information from aspiration to identify it as functionally equivalent to [s]? They key may be in their experience with the L2 and their familiarity with the L2 dialectal feature. In this case, since we studied listeners of elementary and intermediate Spanish (levels 1 and 2) with no experience with aspirating dialects, the answer may be they cannot. It is in these levels where we can best determine if L1 dialect plays a role in perception. Thus, our first research question is as follows:

> Q1: Do AAE and GAE listeners differ in their ability to identify WAS aspiration of final /s/ in plural nouns and second-person verbs?

Contrastively, syllable-final /s/ is found as a legitimate sound in both GAE and AAE. Following the cross-language models reviewed, we predict that GAE listeners will assimilate CS [s] to GAE [s]. However, AAE speakers can regularly omit final /s/ from plural nouns and third-person verbs and, as we saw in the studies by Johnson (2005), and de Villiers and Johnson (2007), at least AAE children do not understand /s/ as an agreement marker, while GAE children do. Does this transfer to adulthood and to the perception of L2 features? Consequently, we pose our second research question:

> Q2: Do AAE and GAE listeners differ in their ability to identify CS sibilance in final /s/ in plural nouns and second-person verbs?

Additionally, we have seen how context can affect the perception of stimuli and can render variation of results. For this reason we also wanted to explore how the syntactic and the phonetic contexts of the target variants can influence the perception of aspiration and sibilance. Our third research question is as follows:

> Q3: How do syntactic and phonetic contexts influence stimuli perception?

Finally, we have seen that as experience with an L2 increases so does the identification and categorization of L2 sounds and contrasts. In Schmidt's study (2011), there was no significant difference in [h] identification accuracy for level 1 and 2 listeners, it was not until level 3 that listeners began to identify [h] as a legitimate realization of implosive /s/, and not until level 5 that they performed similarly to native Spanish speakers. In this current study, we included listeners of elementary (level 1) and intermediate (level 2) Spanish of two different L1 dialects. In spite of not having enough experience with the target language, does proficiency level or years of instruction play a role in identification of the variables in this case? Our last research question is the following:

> Q4: What role do L2 proficiency and L2 instruction play in perception with respect to L1 dialect?

3.4 Summary

In this chapter, we have seen how different factors can affect L2 speech perception. Factors that depend on the L2 listener are their L1 background, the age of acquisition of the L2 (given that they are not naïve listeners), and their L2 experience. Factors that depend on the L2 stimuli used are their acoustic characteristics and the type of sounds and contrasts included. Additionally, we have seen how dialectal differences can play a key role in the perception and identification of sounds.

A subsequent review of the main models of L2 speech perception agree that characteristics that are specific to the listener's L1 affect the perception of L2 sounds. While the NLM and the SLM draw on phonological categories stored in memory, the PAM and PAM-L2 as well as the ASP prefer a direct-realist gestural perception. Nevertheless, their proposals seem insufficient to accommodate dialectal variations. One exception that may be crucial for our study is the H1 proposal of the SLM, which states that "sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level" (Flege, 1995, p. 239), at least with respect to the identification of aspiration. The identification of sibilance, given the link between phonetic and morphological implications, may need further models to be explained.

In the following chapter, we describe the methods employed in the design of the experimental task and the procedure to gather the data for this study.

CHAPTER 4

METHOD

In this chapter, we describe the method followed in the design of the experiment to answer the questions and test the hypotheses posed in the previous chapter. We also provide a description of the speakers and the participants in the study, the materials and instruments employed, and the procedure followed. Additionally, data coding and data analysis methods, as well as the acoustic analyses performed, are also reported here.

4.1 Participants

4.1.1 Speakers

The recordings of the four speakers that we employed for this current experiment were selected from the corpus of stimuli recorded by 8 Spanish speakers. For this current study, we selected one female WAS speaker (WASF1) and one male WAS speaker (WASM2) from Seville (Western Andalusia), one female CS speaker (CSF2) from Bilbao (Basque Country), and one male CS speaker (CSM2) from Cuenca (Castile). All of them had higher-level education[17] (M_{age} = 37.5).

Additionally, given that AAE speakers understand GAE, and that the opposite is not always true (Rickford, 1975), a native speaker of GAE was recorded and

[17] Speaker WASM2 was pursuing his B.A. at the time when the recordings were carried out.

analyzed to measure the common ground between CS, WAS, and GAE in terms of the voiceless fricative sibilant /s/.

4.1.1 Listeners

Two hundred and six listeners were recruited for participation in this current study: 99 AAE learners of elementary Spanish (AAE1), 27 AAE learners of intermediate Spanish (AAE2), 34 GAE learners of elementary Spanish (GAE1), and 56 GAE learners of intermediate Spanish (GAE2).

Recruitment took place in the USA by means of personal contact with the experimenter and arrangements with the Spanish instructors of the participants. All participants satisfied the following general requirements: i) native speakers of American English, ii) students of Spanish at university level, iii) no speech or hearing disorder, iv) no previous stay in a Spanish-speaking country for over 3 months.

The criteria to classify a participant as a speaker of AAE were i) informal conversations with the participants by the experimenter and a second trained expert who, being acquainted with this dialect, attested their dialect use, ii) answers to the Spoken English Questionnaire (see Section 3.3.2 below), iii) being in the same Spanish section as these speakers, taking into account that “contact with a different dialect of the L1 could conceivably cause perceptual changes” (Best & Tyler, p. 18). Assignment to level of Spanish proficiency was made attending to the section in which participants were enrolled at the time of testing (elementary Spanish and intermediate Spanish sections). All instructors reported using sibilance in the classroom, despite their Spanish dialect of origin.

Participants in the AAE1 group (M_{age} = 20.68) were students enrolled in elementary Spanish courses during the fall semester of 2012, with an average of 2.21

years of instruction. Two participants were excluded from the study for reporting a hearing disorder. None reported a stay in a Spanish-speaking country for 3 months or more. Nine of them reported speaking a language other than English and four reported another language at home.

Participants in the AAE2 group (M_{age} = 20.76) were students enrolled in intermediate Spanish courses during the fall semester of 2012, with an average of 3.48 years of instruction in Spanish. Two participants reported a stay in a Spanish-speaking country of 3 months or more. One of them reported speaking a language other than English and this same listener also reported speaking another language at home.

The GAE1 group of participants (M_{age} = 21.60) was enrolled in elementary Spanish sections during the spring semester of 2013, whose average of instruction was 2.07 years. Three participants reported a stay in a Spanish-speaking country for over 3 months. One of them also reported a speech or hearing disorder. Four participants reported speaking a language other than English and five reported another language at home.

Finally, participants in the GAE2 group (M_{age} = 18.98) were students enrolled in intermediate Spanish courses during the fall semester of 2012, with an average of 3.04 years of instruction in the L2. Two participants reported a stay in a Spanish-speaking country of 3 months or more. Four of them reported speaking a language other than English and two of these reported another language at home.

Therefore, after discarding the number of speakers who reported hearing or speech disorders and those with a previous stay in a Spanish-speaking country, the total number of subjects for each of the groups was as follows: AAE1 (N = 97), AAE2 (N = 25), GAE1 (N = 30), GAE2 (N= 54).

4.2 Materials

For both this current experiment and a pilot test previously carried out, we compiled a list of Spanish words divided into 4 categories: twenty singular nouns (SN), 20 plural nouns (PN), 20 third-person verbs (3PV), and 20 second-person verbs (2PV) in the present tense. The nouns were embedded in carrier sentences while the verbs were inserted in content sentences, both types containing between 8 and 10 syllables each (see Appendix C). These sentences derived from those produced by Cervera and González-Álvarez (2010, 2011), which in turn were based on the Speech Perception in Noise (SPIN) sentences elaborated by Kalikow et al. (1977). The SPIN test consists of several sets of sentences in English, in which the target word is in final position, and contains interspersed high-probability and low-probability sentences. That is, sentences in which the target word can be predicted from the context and sentences in which the target word cannot be predicted from the context, respectively. The novelty in our present experiment is that target words were embedded in initial or medial position within the sentence, not in final position, to avoid the type of neutralization previously described in O'Neill's study (2005).

Target words mainly consisted of two syllables, in which the nucleus of the last syllable was an open vowel /a, e, o/. They were followed by a word starting by either one voiced stop /b, d, g/, one voiceless stop /k, p, t/, one nasal or one lateral /m, n, l/, or by an open vowel /V/. These 10 phoneme contexts appeared in 2 different sentences within each category.

For example, target words ending in vowel before context /t/ were:

(SN)	*Digo mano torpemente*	*Digo coche torpemente*
	(I say hand awkwardly)	(I say car awkwardly)

(3PV) *Tiene terreno en el campo* *Está tomando mucha verdura*

(He has land in the countryside) (He is eating a lot of vegetables)

Figure 12 below shows the waveforms of WAS and CS /t/ after vowel:

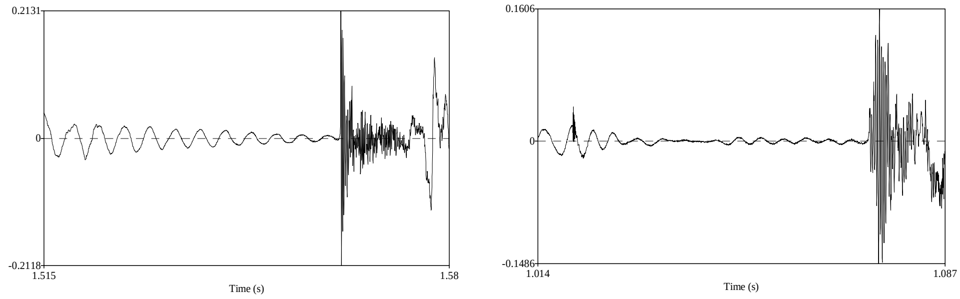


Figure 12. *WAS and CS /t/ after vowel*

Target words ending in the morphological marker *–s* before /t/ were:

(PN) *Digo colas torpemente* *Digo amigos torpemente*

(I say tails awkwardly) (I say friends awkwardly)

(2PV) *Deberías tener más cuidado* *Necesitas tiempo para pensar*

(You should be more careful) (You need time to think)

Table 13 displays the waveforms for /t/ after aspiration and sibilance:

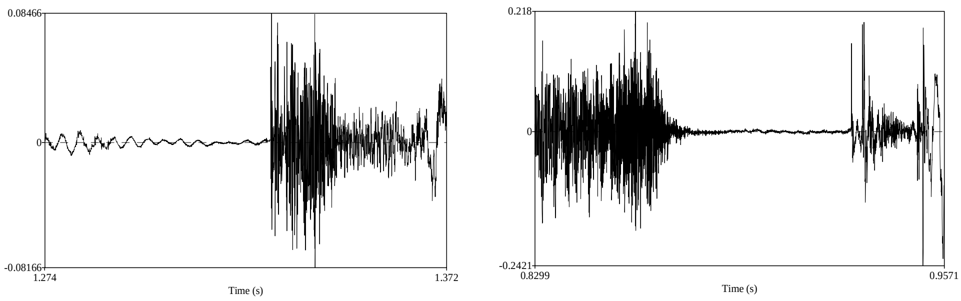


Figure 13. *WAS aspiration and CS sibilance before /t/*

From the 10 phonetic contexts that followed the target words, 6 of them have identical counterparts in English (/p, k, m, n, l, V/) in terms of place and manner of articulation, while the remaining 4 (/b, d, g, t/) have similar but not identical counterparts in English. Since stimuli consisted of sentences, we need to consider a few allophonic variations that occur in Spanish due to the influence of the preceding sounds in connected speech. Voiced stops /b, d, g/ in word-initial position preceded by vowel or /s/ become voiced approximants [β̞, ð̞, ɣ̞] (Garrido Almiñana, Machuca Ayuso, de la Mota Gorriz, 1998). This is true for CS after vowel and /s/ and for WAS after vowel. When WAS aspiration precedes these voiced stops, they become fricatives [v, ð, x]. Additionally, while /t/ is an alveolar stop in English, it is a dental stop in Spanish (see Table 3 below). The rest of the phonemes share place and manner of articulation with their English counterparts; however, WAS voiceless stops carry post-aspiration, while nasals and the lateral sound are geminated.

Table 3

Dialectal allophones of Spanish consonants in word-initial position after [s] and [h]

	CS			WAS		
		Place	Manner		Place	Manner
b	β̞	bilabial	approximant	v	labiodental	fricative
d	ð̞	dental	approximant	ð	interdental	fricative
g	ɣ̞	velar	approximant	x	velar	fricative
p	p	bilabial	stop	p^h	bilabial	stop
t	t	dental	stop	t^h	dental	stop
k	k	velar	stop	k^h	velar	stop
m	m	bilabial	nasal	^{h}m.m	bilabial	nasal
n	n	alveolar	nasal	^{h}n.n	alveolar	nasal
l	l	alveolar	lateral	^{h}l.l	alveolar	lateral
V	V	glottal	open	hV	glottal	open

4.2.1 Recording

The 4 sets of sentences were recorded twice by four speakers of CS (two males, two females) and four speakers of WAS (two males, two females) at a 44.1 kHz and 16 bps sampling rate in a recording booth at the Phonetics Laboratory of the University of Seville (Spain), using a Marantz Professional PMD671 solid-state recorder and a Shure SM48 microphone, under the presence of the experimenter. Speakers were instructed to read as naturally as possible, as if they were talking to a friend at a normal conversational rate.

Originally, this set of stimuli was added three levels of noise (30dB, 55dB, 65dB) with Akustyk for Praat (Plitcha, 2010), to be used in the pilot test only. With the pilot test we explored the extent to which aspiration and sibilance were subject to disruption by noise, as we will see in Section 5.1.1.4. As evidence suggested, at least the GAE listeners obtained native-like scores for sibilance in all noise conditions, and their identification of aspiration was generally less accurate than that of sibilance but increased with level of proficiency, generally despite noise condition. Thus, the effect of noise here may be confounded with proficiency level. Nevertheless, the evidence that was most interesting for our purposes came from the AAE listeners. Therefore, we eliminated the noise factor in our current experiment for this dissertation and focused on the performance of lower-level participants of both L1 dialects.

4.3 Procedure

For this current experiment, we selected the four best exemplars out of the eight speakers from our corpus: one female and one male speaker per L2 dialect. CSM1 and WASM1 were discarded due to intonation and speech rate deviations in comparison with the rest of speakers. Subsequently, speakers CSF1 and WASF2 were eliminated from this present study in order to have one exemplar speaker from each gender and L2 dialect. The resulting 320 sentences were converted to mp3 format and included in a test mounted in the experimenter's university webpage supporting HTML, PHP, and MySQL.

The experimental task was devised by a computer technician specifically for the purpose. This application developed in the Laboratory of Phonetics at the University of Seville is used to gather a great amount of data, which would not be possible otherwise. Participants must go through five sections to complete the test. The first section gathers general information for the sampling attributes of the experiment, such as age, gender, etc. The second section gathers linguistic information about the listeners' L1 use in informal conversation, and aims at compiling data on L1 dialect use. The third section is a training exercise in which five samples of stimuli from the corpus appear, one at a time, with their corresponding solution. In the fourth and final section of the experiment, participants reproduce each individual stimulus twice before choosing an answer. The number of stimuli that appear in each test is fixed (60 sentences, in this case) but the order and type of stimulus is randomly presented by the application. Finally, in the last section the application asks the participant to confirm the submission of the results, and thanks the listener for their participation.

The experimental task is programmed in a within a single webpage; therefore, during the completion of the task, the participant does not browse from one page to the next. This simple detail makes participants unable to use the browser to go back or go next, and lose the information provided up to that moment. Additionally, the Javascript functions that manipulate the webpage are invisible to the user, even to experienced programmers.

Thus, this web application was able to originate a different test for each of the participants in the study. Thanks to this randomness, we can have an unlimited number of stimuli in the corpus because they all have the same probability to be pooled by the program. Furthermore, the application records the listener's reaction times to each stimulus. Figure 14 below shows a screenshot from the experimental task.

Figure 14. *Screenshot of the identification task in the current study*

Experiments were generally run in one of the following two settings: language classroom or at home. AAE listeners of both elementary and intermediate levels of Spanish took the test in language laboratories at a US university where the experimenter was present. These laboratories had 30 computer stations where students completed the identification task individually, using headphones. GAE listeners of both elementary and intermediate levels took the test at their US institutions, under the direction of their instructors. No monetary compensation was given to the participants but they were granted extra credit for their participation.

Participants received written instructions in English that they would listen to sentences in Spanish and would need to select the sentence that they heard from the two forced-choice written options given. When the target word was a plural noun or a second-person verb, the alternative option offered the same sentence with the same target word without the final *–s*, and vice versa. For example, if the sentence played was *Nunca comes nada dulce* (You never eat anything sweet), the two options given were *Nunca come nada dulce* (He/She never eats anything sweet) and *Nunca comes nada dulce* (You never eat anything sweet), so the correct option could not be inferred from reading the sentences alone. These instructions were presented in an informed consent document (see Appendix D) and repeated in the test itself.

Listeners performed a self-paced sentence identification task in which each participant listened to a separate set of 60 sentences randomly chosen from the corpus, with no feedback provided. As a training method, the test played five sentences and showed the correct answer, so that they became familiar with the task and could adjust volume settings. Participants had to listen to each sentence twice in order to proceed to the next one, and were allowed unlimited time to complete the test, although a total duration of 15-20 minutes was estimated.

Additionally, participants were required to sign the aforementioned informed consent form and fill out two initial questionnaires included in the test: a Language Background Questionnaire (see Appendix E) and a Spoken English Questionnaire (see Appendix F), both aimed at making a detailed profile of the listeners for classification and interpretation of the findings in this study.

4.3.1 Language Background Questionnaire

The language background questionnaire gathered information about age and gender of the participants, birthplace of the participants and their parents or guardians, languages spoken at home and outside home, accent or dialect spoken by the participants, whether the participants had ever stayed in a Spanish-speaking country for over 3 months, dialect of Spanish currently exposed to, years of Spanish instruction, Spanish level (this question was later excluded; level was determined by section attending at the time of testing, as reported earlier), other languages in which participants were fluent, and hearing or speech disorders reported.

4.3.2 Spoken English Questionnaire

The Spoken English Test listed 13 questions designed to test for dialectal features included among the most stable and rising in AAE speech (Wolfram, 2004). Specifically, the test looked into: copula absence + *V-ing*, habitual *be* + *V-ing*, third-person *–s* absence, copula absence + adjective, negative inversion, possessive *they*, existential *they*, noun plural absence, resultative *be done*, cluster reduction before vowel, regular past tense *–ed* deletion before vowel, and r-lessness before vowel.

From the 122 AAE listeners, 49% of them reported using some of the features in this test. Table 4 below shows the percentages reported by these listeners for each of the elements tested.

Table 4

Percentage of AAE usage reported by AAE listeners

	Percentage of AAE listeners
copula absence + *v-ing*	21
habitual be + *v-ing*	5
third person –*s* absence	12
copula absence + adjective	23
negative inversion	34
possessive *they*	18
existential *they*	32
noun plural absence	5
resultative *be done*	18
cluster reduction before vowel	39
-*ed* deletion before vowel	13
r-lessness before vowel	48

4.4 Data Coding

Data were gathered by the program immediately after submission into an Excel sheet displaying all information submitted by each participant, i.e., their answers to the linguistic background and spoken English questionnaires and the 60 stimuli they listened to in order of appearance together with their score (1 = correct, 0 = incorrect). The following information was entered into a file using IBM SPSS Statistics 20: listener ID, age of listener, gender of listener (1 = female; 2 = male), other languages spoken at home (1 = yes; 2 = no), other languages spoken at school (1 = yes; 2 = no), dialect of listener (1 = GAE; 2 = AAE), stay in a Spanish-speaking country (1 = yes; 2 = no), years of instruction (1 = less than 1 year; 2 = 1-3 years; 3 = 3-5 years; 4 = more than 5 years), level of instruction (1= elementary; 2 = intermediate; 3 = high-intermediate; 4 = advanced; 5 = proficiency). At this point, participants who reported speech or hearing disorders were excluded so this variable was no longer present.

We then included the following characteristics for each stimulus in order of presentation (1-60): speaker dialect (1 = WAS; 2 = CS), speaker gender (1= female; 2 = male), sentence type (1 = 3PV, 2 = 2PV, 3 = SN, 4 = PN), phonetic context (1 = [b]; 2 = [d]; 3 = [g]; 4 = [k]; 5 = [p]; 6 = [t]; 7 = [m]; 8 = [n]; 9 = [l]; 10 = [V]), score (1 = correct; 2 = incorrect), reactions times, and place of testing.

On a separate SPSS file, we also included accuracy percentages (0-100%) per participant of [h] perception (*aspiration*), [s] perception (*sibilance*), [V] perception in WAS sentences and [V] perception in CS sentences, dialect and level to which they belonged (1 = AAE1, 2 = AAE2, 3 = GAE1, 4 = GAE2), and place where they took the test (1 = computer classroom; 2 = home).

Additional SPSS files were created for the classification of stimuli according to their acoustic characteristics. For voiceless stops, we indicated gender of speaker (1 = female, 2 = male), L1 dialect of speaker (1 = WAS, 2 = CS), type of phonetic context (1 = sk, 2 = sp, 3 = st, 4 = k, 5 = p, 6 = t), duration of preceding vowel (ms), closure duration values (ms), and Voice Onset Time values (ms). For fricative sounds, we also indicated gender and L1 dialect of speaker as in the previous file, type of phonetic context (1 = sb, 2 = sd, 3 = sg, 4 = h, 5 = s), duration of previous vowel (ms), fricative intensity (dB), duration (ms), Center of Gravity (Hz), dispersion (Hz), kurtosis, skewness, and spectral peak (Hz).

.

4.5 Statistical Analysis

Accuracy results were obtained by dividing the number of correct answers per listener and variable by the number of stimuli they received from each variable. Thus, a participant that listened to 30 sentences where aspiration was present, and identified 20 of these sentences correctly, had an accuracy score of 66.67%. A general level of significance of $p < .05$ was assumed for all tests. However, when applicable, levels of significance were also expressed as $p < .01$, $p < .005$, and $p < .001$.

We initially performed a Kolmogorov-Smirnov test to check for normal distribution. As not all groups showed normal distribution, we applied Spearman rank order correlations to determine the correlation between participant characteristics and variables tested. The initial characteristics we tested were i) stay in a Spanish-speaking country, ii) languages other than English at home, iii) languages other than English at school, iv) languages other than English spoken. Subsequently, participants who displayed influential factors were removed from the results. In the absence of normal distribution in most of the groups, we then proceeded to run non-parametric tests to analyze the results. For each group, we ran Wilcoxon tests (non-parametric equivalent to paired two-sample *t*-tests) comparing their intra-group performance in aspiration and sibilance first, and then between vowel identification in WAS and CS sentences. We then ran a Kruskal-Wallis test (non-parametric equivalent to ANOVA) to compare performance across all groups, with subsequent Mann-Whitney tests (non-parametric equivalent to unpaired two-sample *t*-tests) between pairs of groups. A third analysis was directed towards the syntactic context in which the target words were embedded and the phonetic contexts that followed [h] and [s], and finally, the years of instruction in Spanish that each group received. We explored the overall performance by all L2 learners and the performance by group of learners, with special emphasis on the analysis of aspiration and sibilance in terms of stops and fricative sounds.

4.6 Acoustic Analysis

In light of the results obtained from the identification task, an acoustic analysis with Praat (Boersma & Weenick, 2009) was carried out of individual aspiration and sibilance tokens according to the contexts in which the most significant findings were found, i.e., before voiceless stops (WAS aspirated stops [p^h, t^h, k^h] and CS fricative [s]), before voiced approximants (WAS fricatives [v, ð, x] and CS fricative [z], and in intervocalic position (WAS [h] and CS fricative [s]).

4.6.1 Voiceless Stops

4.6.1.1 Previous vowel

The duration of the vowels before the target stops was measured with spectrograms, waveforms, and listening to the recordings. The onset of the vowel was placed after a change in formant transitions was observed, while the offset was marked at the beginning of the closure period. Analyses of the first three formants using formant tracking (Maniwa, Jongman, & Wade, 2009) and of the Harmonics-to-Noise Ratio were carried out using Praat.

4.6.1.2 Closure duration

The duration of the closure of the target stops was measured by means of spectrograms and waveforms. The onset was marked at "the end of aspiration noise or, if none existed, the decrease in formant intensity at the end of the vowel associated with closure to the release burst" (Parrell, 2012, p. 39), while the end was placed at the release of the target stop.

4.6.1.3 Voice Onset Time

Voice onset time (VOT) was manually measured by the author and another trained expert three-way: by means of waveforms, spectrograms, and by listening to the recordings. The authors cited in Section Torreira (2012) measured the VOT "from the beginning of the release of the target stop to the first visible voicing period of the upcoming vowel" (p. 55). For Horn (2013), "the endpoint was marked at the downward zero-crossing of the first full period of the following vowel" (p. 36), a measurement that is also followed by Parrell (2012) and Torreira (2007b). As this last author states:

> Even though this method cannot be considered entirely faithful to the events in the speaker's glottis, the signal being the result of overlapping supraglottal and glottal gestures, it appeared to be a consistent way of measuring VOT in the absence of articulatory data. (p. 115)

Finally, Torreria (2007a), O'Neill (2009) and Ruch (2010) do not specify how VOT was measured in their respective studies.

In our case, the author and the trained expert found that the beginning of voice did not coincide with the beginning of the following vowel in several cases, as was the case with the authors above. Instead, voice started towards the end of the stop. If the offset was marked at the first downward zero-crossing before at the first full pulse was completed in these instances, we could hear part of the plosive on the following vowel. Figure 15 exemplifies this case.

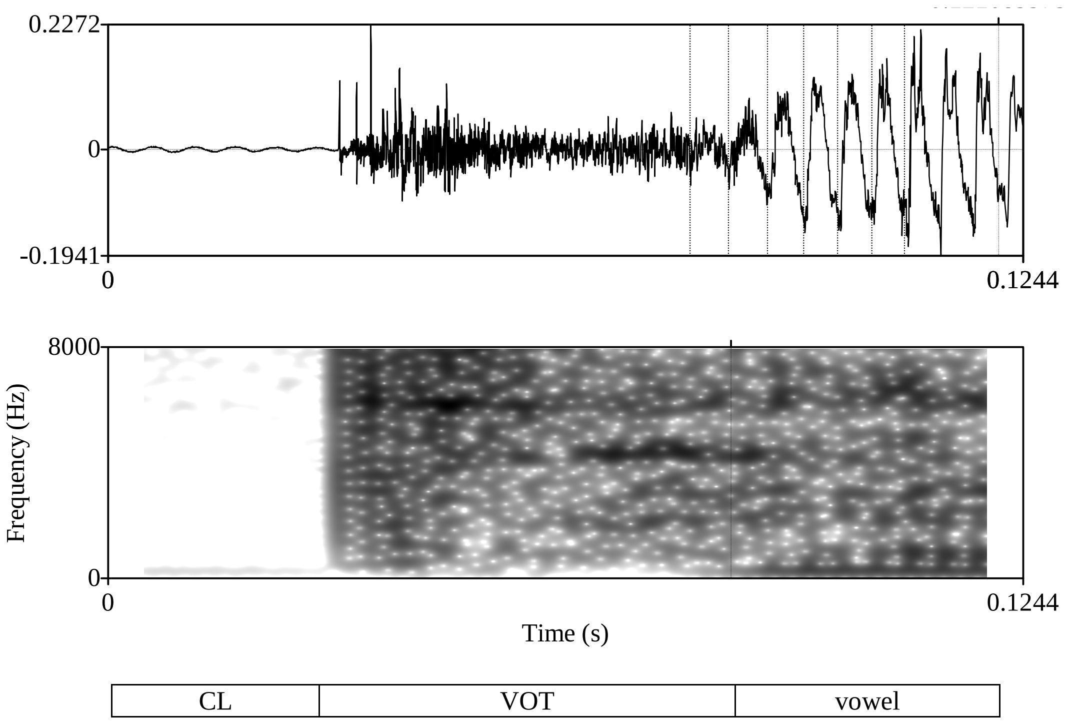


Figure 15. *Waveform and spectrogram of WAS [t^he]*

For this reason, we allowed up to the second pulse to be included in the VOT. We are aware that this does not match the definition of VOT per se; nevertheless, we feel that the measurement would not be accurate if we proceeded otherwise. As Lisker and Abramson (1964) state, “there is a danger of giving primary emphasis to an instrumentally detectable acoustic disturbance that, in the situation, can have no auditory consequences” (p. 416). Figure 16 shows an example of GAE [t^h], recorded in the same laboratory and under the same conditions as our stimuli, which matches the standard definition of VOT.

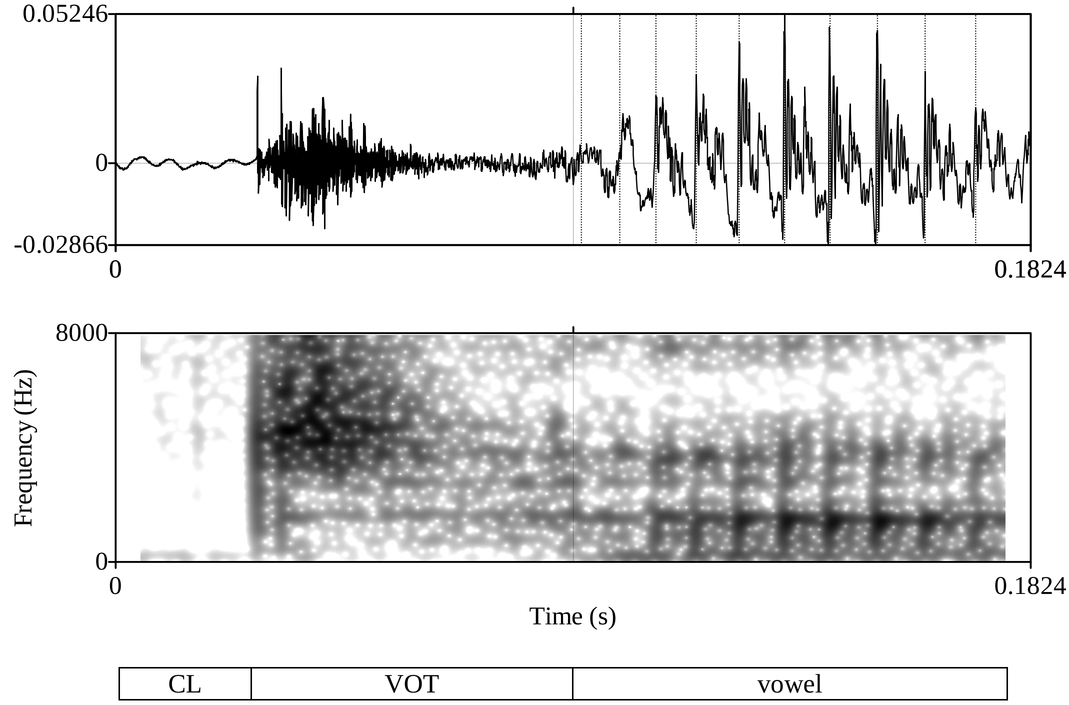


Figure 16. *Waveform and spectrogram of GAE [t^{h}aɪ]*

4.6.2 Fricatives

4.6.2.1 Spectral information

To measure the acoustic cues of fricative sounds, we focused on the spectral moments mentioned in the studies reported in Section 2.1.2.1. For this purpose, we used a script that can be run with Praat, adapted by the laboratory technician at the University of Seville. The spectrum was set at 18 000 Hz, in light of Barreiro Bilbao’s (1994) findings and our findings from the analysis of GAE and CS /s/. After reviewing several methods, we used a high-pass filter to eliminate the section of the target sounds under 500 Hz (Cicres, 2013; Jongman et al., 2000; Koenig, Shadle, Preston, & Mosshammer, 2013; Maniwa et al., 2009) to avoid the masking of higher frequencies. Additionally, 50% at the center of the fricatives was selected for the analysis, since frequencies at the beginning and end of the sound are influenced by the adjacent sounds. Apart from accounting for intensity (dB) and duration (ms), the elements measured were the following, as described by the previous authors and Styler (2013):

The first spectral moment is the center of gravity (COG). It corresponds to the frequency that divides the spectrum into two halves such that the amount of energy in the top half (higher frequencies) is equal to that in the bottom half (lower frequencies). The COG measures the mean concentration of energy of a sound. A sound with a lot of high-frequency energy will have a high value for COG.

The second spectral moment is dispersion (also termed variance and standard deviation). It provides a measure of whether the energy is concentrated mainly in a small band around the COG or spread out over a wide range of frequencies; thus, it measures the distance of the frequencies with respect to the COG.

The third spectral moment is called kurtosis. It refers to the shape of the data distribution. If kurtosis value is 0 (this value has no unit of measurement), the data matches the Gaussian distribution. If it has positive values, the higher these are, the higher the peakedness (clearly defined spectrum) of the sound. If it has negative values, the higher these are, the flatter the distribution.

Finally, the fourth spectral moment is skewness (asymmetry). It quantifies how symmetrical the distribution is with respect to the COG. As with kurtosis, this value has no unit of measurement. When the value is 0, it means the distribution of energy is symmetrical along the sound. When it has positive values, it means the energy is concentrated in lower frequencies. When the values are negative, it means the energy is concentrated in higher frequencies.

Additionally, we extracted the spectral peak location of each sound manually, that is, the local maximum of their spectrum, which, as stated in Section 2.1.2, is believed to be a very reliable cue to distinguish the place of articulation of the sounds.

4.7 Rationale for the Identification Task

Following Strange and Shafer (2008), a justification for the use of this experimental design is provided. We used an identification task, instead of a categorization or a discrimination task, because in this type of task "the stimuli must be compared against internal representations of phonetic/phonological categories" (p. 183). In our task, several instances of the target words representing the L2 variants under study were embedded in sentences with multiple phonetic and phonotactic contexts, and presented in random order and one at a time. The use of a mixed list prevents the listeners from anticipating "the context in which the target phones will occur" (p. 183). The use of real words instead of nonsense items, despite the fact that lexical and phonetic effects might be confounded, ensued from the assertion that "experiments using real world materials more accurately reflect the receptive problems of L2 learners" (p. 163). Embedding target words in sentence contexts instead of using citation-form stimuli avoids the problem of the results not being easily generalizable to real-world situations of understanding continuous speech. The fact that our stimuli were not conversational per se, but rather read stimuli, still can render results that "may be generalized to some real-world situations, such as the language classroom" (p. 163).

4.7.1 Pilot Study

As a preliminary study, we used 56 sentences from the set of stimuli that was added noise (8 speakers): seven sentences per speaker dialect and target word type. In this case, we only used /p, t, k/, /b, d, g/ and vowel as following sounds for L2 learners. Initially, the pilot test had 80 stimuli, as we also included /m, n, l/ as following sounds, but reduced the number of stimuli due to the duration of the test, which was discouraging for participants given the design of the platform in which it was mounted. Figure 17 shows a screenshot of the task.

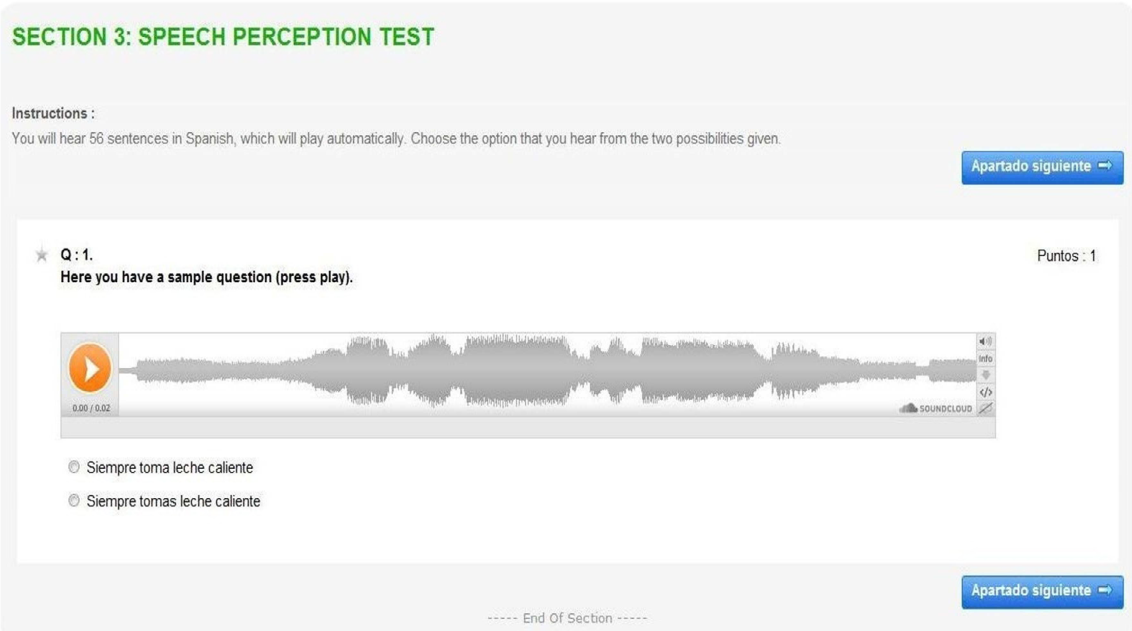


Figure 17. *Screenshot of the identification task in the pilot study*

4.7.1.1 Speakers

Speakers in this pilot test were the four speakers in our current study with the addition of another set of four speakers (one male and one female per L2 dialect). The additional four speakers were a male speaker (CSM1) from Toledo (Castile), a female speaker (CSF1) from northern Cordova (Northern Andalusia, at the border with Castile), who retained sibilance, a male speaker (WASM1) from Seville, and a female speaker (WASF2) from Seville (Western Andalusia). Three of them had higher-level education[18] (M_{age} = 28.25), with the exception of speaker WASM1.

4.7.1.2 Listeners

Twenty-four native Spanish listeners participated in this pilot identification task with the initial 80 stimuli, while 53 L2 learners of Spanish participated in the task with the final 56 sentences, either under the presence of the examiner or another trained instructor, or at home, during the spring semester of 2012. These listeners were classified according to their reported proficiency in the L2: Levels 1 and 2 were labeled under "low", listeners of Levels 3 and 4 were named "mid" and listeners of Level 5 were termed "high"; and according to L1 dialect: AAE and GAE. Based on the findings from this pilot test, our current experiment only focused on elementary (Level 1) and intermediate (Level 2) L2 learners. We also examine these preliminary results in the following chapter.

[18] Speakers CSM1 and WASF1 were pursuing their B.A. at the time when the recordings were carried out.

CHAPTER 5

RESULTS

In this chapter, we present the results obtained from the pilot test and the subsequent identification task for this current study to answer the research questions stated at the end of Chapter 2. We first introduce the preliminary results for the native Spanish listeners and the L2 learners that participated in the pilot test, and then provide a review of the performance of the participants in the current identification task, in terms of accuracy identification of aspiration and sibilance in general, also according to the amount of instruction received by the listeners, and subsequently according to the syntactic and the phonetic contexts of the target words. Acoustic analyses are subsequently provided in search of an explanation for the results.

5.1 Results of Perception

5.1.1 Pilot Study

5.1.1.1 Native listeners

Twenty-four native listeners (NL) of WAS participated in the pilot experiment. Their lowest accuracy score was for aspiration (*M* = 91.67, *SD* = 9.58), while their highest score was for sibilance (*M = 100*, *SD* = 0), with percentages of *M* = 98.51, 97.62; *SD* = 2.96, 5.45 for vowel in WAS sentences and vowel in CS sentences, respectively.

Wilcoxon tests showed that the perception of sibilance was significantly higher than that of aspiration for this group of listeners [Z = -3.21, p = .001]. Taking into account that the stimuli was presented in noise, as we will see in Section 4.1.4, aspiration seems to be vulnerable to disruption by noise, at least for these group of NL of Spanish. Precisely, it was at all levels of noise that NS presented differences (65dB: Z = -2.07, p < .05; 55dB: Z = -2.94, p < .005; 30dB: Z = -2.49; p < .05) between aspiration and sibilance. In fact, NL21, NL23 and NL24 showed remarkably lower scores in the identification of aspiration. This could be the main reason for such results. Finally, their identification of sentences ending in vowel was similar in both WAS and CS conditions.

5.1.1.2 L2 listeners

Fifty-three L2 learners of Spanish, AAE low (*n* = 24), GAE low (*n* = 6), GAE mid (*n* = 19), GAE high (*n* = 4), took part in this initial test. Twenty-one had stayed in a Spanish-speaking country at the time of testing, none of which were AAE listeners, while 30 had not. Overall performance was as follows: accuracy in aspiration was markedly poorer (*M* = 32.28, *SD* = 28.47) than in all other conditions, followed by sibilance (*M* = 82.92, *SD* = 29.14), CS vowel (*M* = 83.42, *SD* = 27.76), and WAS vowel (*M* = 85.31, *SD* = 20.84). Table 5 shows the identification percentages obtained by each listener group per variable.

Table 5

Mean accuracy percentages for all groups and variables

		aspiration		sibilance		WAS vowel		CS vowel	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
group	AAE low	25.32	25.88	62.64	33.64	76.78	22.81	73.81	25.33
	GAE low	21.53	18.15	100	0	83.93	23.60	96.43	4.12
	GAE mid	32.98	24.65	99.56	1.91	92.50	11.23	95.36	8.44
	GAE high	86.81	11.37	100	0	97.14	3.91	91.43	11.74

5.1.1.3 Preliminary analysis

After applying a Kruskal-Wallis statistical test, we found significant differences across all L2 groups for aspiration [$\chi^2(3) = 12$, $p < .01$]; sibilance [$\chi^2(3) = 28.7$, $p < .001$]; WAS vowel [$\chi^2(3) = 8.03$, $p < .05$], and CS vowel [$\chi^2(3) = 10.33$, $p < .05$].

Wilcoxon statistical tests applied to each L2 group individually to extract intra-group performance revealed that the perception of sibilance was also significantly higher than the perception of aspiration for all L2 learner groups: [AAE low: Z = -3.64, $p < .001$; GAE low: Z = -3.08; $p < .005$; GAE mid: Z = -5.56, $p < .001$; GAE high: Z =
-2.48, $p < .05$].

We then proceeded to analyze how AAE listeners compared with the three GAE groups in terms of aspiration and sibilance. For this purpose, we first considered whether AAE listeners who expressed overt AAE features in the Spoken Language Questionnaire ($n = 14$) and those who did not ($n = 10$) showed evidence of similar or different identification of aspiration and sibilance. In this case, there were no statistically significant differences (aspiration: U = 47, p = .19; sibilance: U = 39.5, p = .07).

A Mann-Whitney test revealed that the perception of sibilance between AAE and GAE low-level groups was significantly higher for GAE listeners (U = 18, $p < .005$) but no differences were found for sibilance between these two groups (U = 71, p = .96). The same statistical test also revealed that the perception of sibilance between AAE listeners and GAE mid was also significantly higher for GAE listeners (U = 61, $p < .001$), but similar between both groups for aspiration accuracy (U = 175.5, p = .20). Upon comparison with GAE high listeners, accuracy for both aspiration (U = 1, p < .001) and sibilance (U = 12, p < .05) was again found to be significantly higher for the GAE listeners.

Subsequently, we analyzed the performance between the groups of GAE listeners. A comparison between GAE low and GAE mid showed that the perception of aspiration and sibilance was similar between both groups (U = 42.5, p = .36; U = 54, p = .88). When comparing GAE mid with GAE high, it was evident that the accuracy of aspiration identification (U = 2, p = .001) was higher for the most proficient learners but similar between the two groups for sibilance (U = 36, p = .91). Likewise, the performance between GAE low and GAE high listeners was also significantly favorable to the second for aspiration only (U = 0, p = .01), but identical between both for sibilance (U = 12, p = 1).

We then compared the results of those who had stayed in a Spanish-speaking country and those who had not. A Mann-Whitney test showed that all differences were statistically significant, with higher accuracy for those who had stayed in a Spanish-speaking country before: [aspiration (U = 194.5, p = .01), sibilance (U = 147, p < .001), vowel WAS (U = 177, p < .005), and vowel CS (U = 176, p < .005). However, we have to consider that none of the AAE speakers (low-level) had ever stayed in a Spanish-speaking country while 13 out of the 39 GAE speakers did (at mid- and high-levels, but not at low-level).

Finally, we compared the performance in aspiration and sibilance identification between L2 learners and NL. Mann-Whitney tests revealed that the identification of aspiration by NL was significantly better than that by the rest of the groups of L2 learners except for the GAE high group (AAE low: U = 3, p < .001; GAE low: U = 0, p < .001; GAE mid: U = 11.5, p < .001; GAE high: U = 31, p = .29). In the perception of sibilance, however, no significant differences were found between GAE listeners and NL, but AAE low seemed to be significantly less accurate than NL ($U = 79$, $p < .001$).

These preliminary results indicate that identification accuracy of aspiration for GAE listeners gradually increased with level of proficiency, rendering statistically significant differences between mid- and high-level learners. Native-like performance for GAE listeners was achieved at high-level of Spanish, with no differences in either aspiration or sibilance identification between these listeners and NL. Likewise, all L2 learners in this pilot study showed native-like performance in their identification of sibilance, but not in aspiration. While these results were predictable, a striking finding was the fact that significant differences in the identification of sibilance were found between GAE low and AAE low listeners in favor of the GAE listeners, suggesting that L1 dialect features may influence perception in this case.

5.1.1.4 The effect of noise

As stated in the description of the stimuli employed for the pilot test, three levels of noise were added to the sentences in the task: 65dB, 55dB, and 30dB, the influence of which we analyze here. As we can see in Table 6 below, the L2 listeners' overall identification of sibilance in the three conditions was similar [$\chi^2(2) = .21$, $p = .90$] while their identification of aspiration as a group was conditioned by the level of noise [$\chi^2(2) = 9.7$, $p < .01$].

Table 6

Overall identification percentages of aspiration and sibilance in the three noise conditions

	noise65dB		noise55dB		noise30dB	
	M	*SD*	*M*	*SD*	*M*	*SD*
aspiration	24.53	33.43	36.16	36.06	36.14	30.31
sibilance	83.49	29.80	83.02	31.23	82.26	32.02

Figure 18 shows the performance of both low-level groups. At first sight, the figure already indicates what statistics can corroborate: GAE low listeners significantly outperformed AAE low listeners only in the identification of sibilance for the three levels of noise (65dB: U = 27, p < .05; 55dB: U = 30, p <.048; 30dB: U = 27, p <. 05).

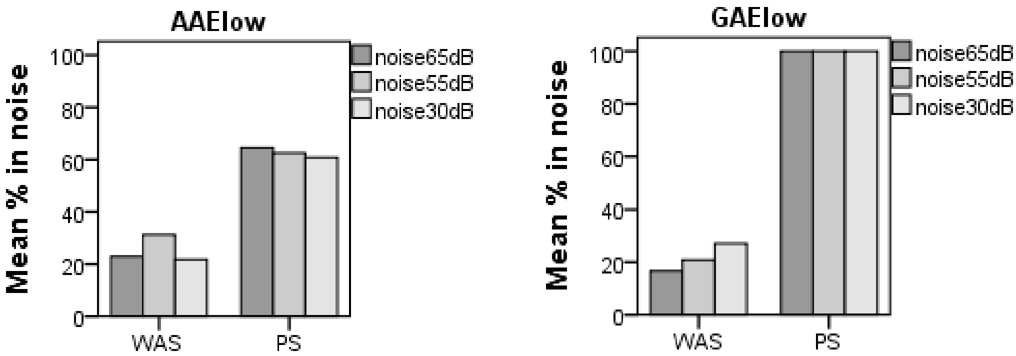


Figure 18. *Identification percentages of aspiration and sibilance for AAE and GAE low-level groups according to noise level*

GAE mid (Figure 19) was the only group for which noise was an influential factor in their identification of aspiration (χ^2 (2) = 19, p < .001), which increased as level of noise decreased. Additionally, their performance was similar to that of GAE low listeners for both sibilance and aspiration at the three levels of noise, and significantly more accurate than that of AAE low listeners for sibilance (65dB: U = 92.5, p < .001; 55dB: U = 95, p < .001; 30dB: U = 85.5, p < .001) and for aspiration only at 30dB (U = 111.5, p < .005).

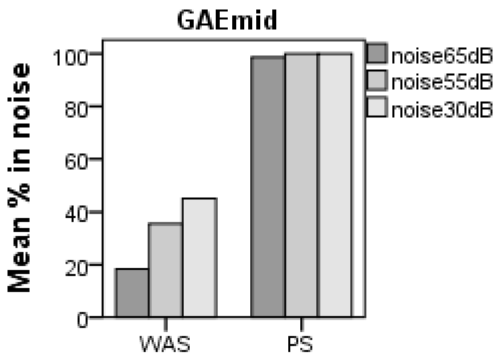


Figure 19. *Identification percentages of aspiration and sibilance for GAEmid listeners according to noise level*

In comparison with GAE high listeners, both groups performed similarly for sibilance but GAE mid identified aspiration significantly more poorly than GAE high listeners at the three levels of noise (65dB: U = 8, $p < .01$; 55dB: U = 6.5, $p < .01$; 30dB: U = 2, $p < .005$).

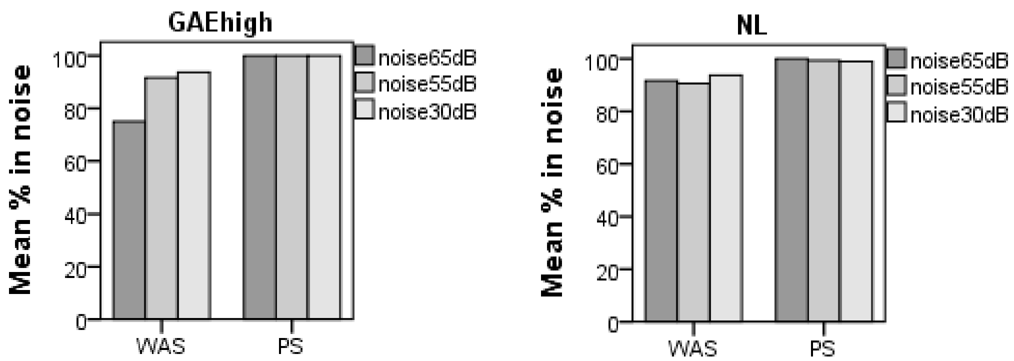


Figure 20. *Identification percentages of aspiration and sibilance for AAEhigh and NL according to noise level*

There were no significant differences between GAE high and NL for any level of noise or target L2 feature, i.e., the performance of GAE high was similar to that of NL of Spanish (Figure 20).

So far, these analyses confirm what was stated in the previous section. All GAE listeners and NL performed similarly in the identification of sibilance, in spite of noise level. GAE listeners at low- and mid- level also performed similarly in the identification of aspiration, but it was not until high-level that GAE showed a significant improvement in the identification of aspiration, similar to that of NL, also regardless of noise level.

AAE listeners, on the other hand, also identified aspiration in a similar manner to GAE low and GAE mid participants, with the exception that, at the lowest level of noise (30dB), GAE listeners of mid-level performed significantly better. The main difference here is that AAE listeners identified sibilance significantly less accurately than all groups of GAE listeners, including their low-level counterparts, in all three noise conditions.

Our aim is to investigate L2 dialect speech perception, particularly in the lowest levels of learning without exposure to the target features, when language-specific patterns of perception are more likely to be reflected. Therefore, we deemed it necessary to discard the use of noise in our following experiment given that its effect was irrelevant for these groups of L2 listeners in the pilot test.

5.1.2 Current Study

As stated in Section 4.5, we initially ran statistical analyses to determine whether certain characteristics played a role in perception: i) stay in a Spanish-speaking country, ii) languages other than English at home, iii) languages other than English at school, iv) languages other than English spoken. The only characteristic that we found to be a significant factor was the stay in a Spanish-speaking country, which was inversely correlated with the perception of sibilance ($r = -.165$, $p < .05$); therefore, these participants were excluded from the study.

5.1.2.1 Overall identification within groups

As Figure 21 shows, the identification of the L2 dialect variant under study, aspiration, was significantly poorer for L2 learners as a whole than the identification of the mainstream variant, sibilance. Nevertheless, the perception of both vowel conditions seemed to be the highest for the participants.

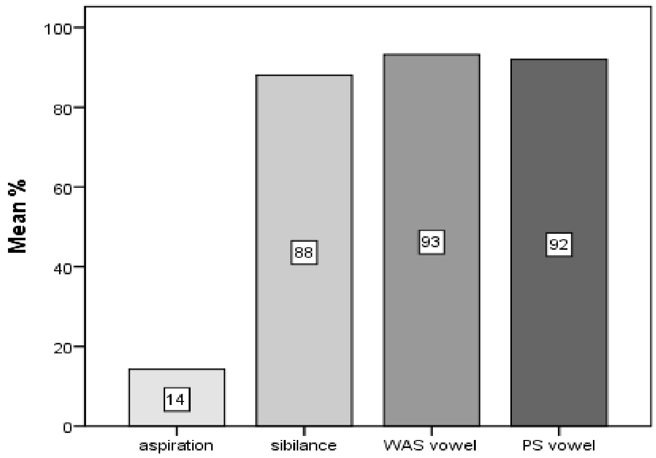


Figure 21. *Overall* a*ccuracy percentages for all variables*

We now proceed to analyze the performance of each group of listeners individually.

AAE1

Taking into account that statistical tests revealed no significant differences between members of this group who expressed overt features of AAE in the Spoken English Questionnaire (*n* = 45) and those who did not (*n* = 52), we considered all of them within the same dialect group (aspiration: U = 1 034, p = .32; sibilance: U = 1 125.5, p = .75).

For AAE listeners of elementary level (*N* = 97), performance was clearly poorer in the identification of aspiration (*M* = 13, *SD* = 16.37) than that of sibilance

(*M* = 82.36; *SD* = 17.34). Their highest scores were for CS vowel and WAS vowel (*M* = 89.37, 91.52; *SD* = 15.26, 12.65, respectively), as shown in Figure 22. A Wilcoxon test corroborated that sibilance was significantly better identified than aspiration [Z = -8.51, p < .001] but no differences were found between accuracy percentages of identification of the two vowel conditions. However, in spite of the relatively high scores for sibilance, its identification was still significantly lower than that of WAS vowel [Z = -5.06, p < .001] and CS vowel [Z = -4.24, p < .001].

The identification of aspiration for this group ranged from 0% to 84.62%, with 34% of the listeners rendering 0% correct answers and 4% of the listeners identifying aspiration above 50%. For sibilance, identification ranged from 23.53% to 100%, with 90.7% of the listeners obtaining 50% or more correct answers and 19.6% of them with 100% correct answers.

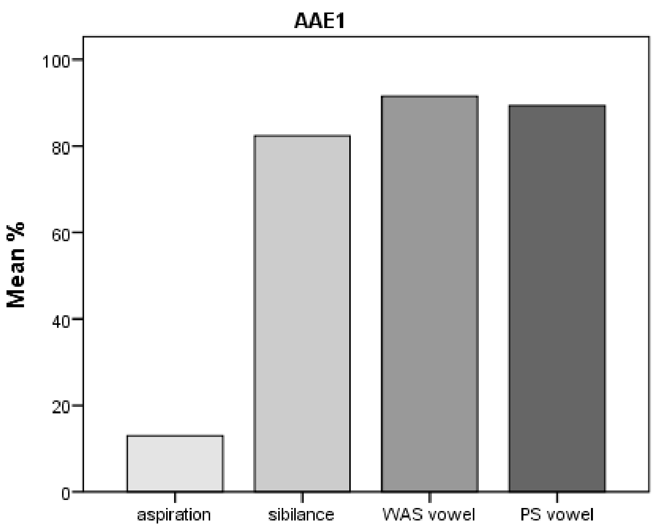


Figure 22. *Mean accuracy for all variables (AAE1)*

GAE1

GAE listeners of elementary level (*N* = 30) also performed significantly lower in the identification of aspiration (*M* = 8.41, *SD* = 10.10) than sibilance (*M* = 92.92, *SD* = 8.67), as Figure 23 shows. Their highest scores were also for WAS vowel (*M* = 97.50, *SD* = 5.10) and CS vowel (*M* = 96.92, *SD* = 5.86). A Wilcoxon test corroborated that the perception of sibilance was also significantly higher than the perception of aspiration for this group of listeners (Z = -4.78, p < .001).

Aspiration identification for this group ranged from 0% to 31.81%, with 40% of the listeners obtaining 0% correct answers and none of them reaching 50%. On the contrary, their identification of sibilance ranged from 69.23% to 100%, with 43% of the listeners obtaining 100% correct answers and all of them above 50%.

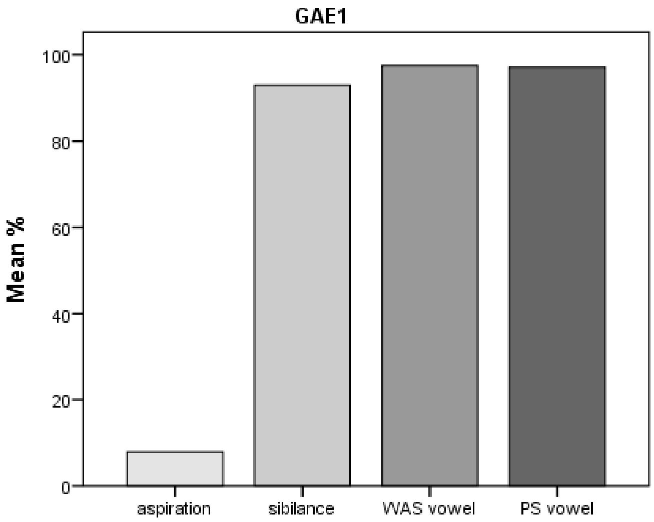


Figure 23. *Mean accuracy for all variables (GAE1)*

AAE2

As was the case with AAE1 listeners, no statistically significant differences were found between participants in this group who reported overt AAE features in the questionnaire (*n* = 10) and those who did not (*n* = 15); therefore, they were considered members of the same group of listeners (aspiration: U = 62.5, p = .50; sibilance: U = 59.5, p = .40).

AAE listeners of intermediate level (*N* = 25) also showed low performance in the identification of aspiration (*M* = 20.3, *SD* = 21.29), as opposed to sibilance (*M* =

89.94, *SD* = 15.34). The perception of WAS vowel (*M* = 92.26, *SD* = 12.69) as well as the identification of CS vowel (*M* = 93.76, *SD* = 12.79) were the highest scores for this group (Figure 24). A Wilcoxon test revealed that sibilance was significantly higher than aspiration [Z = -4.29, p < .001] but no differences were found between both vowel conditions or between sibilance and any of these vowel conditions.

The identification of aspiration for this group ranged between 0% and 81.82%, with 16% of the listeners rendering 0% correct answers and 4% of the listeners identifying aspiration above 50%. For sibilance, identification ranged between 43.75% and 100%, with 92% of them obtaining 50% or more correct answers and 40% of them obtaining 100%.

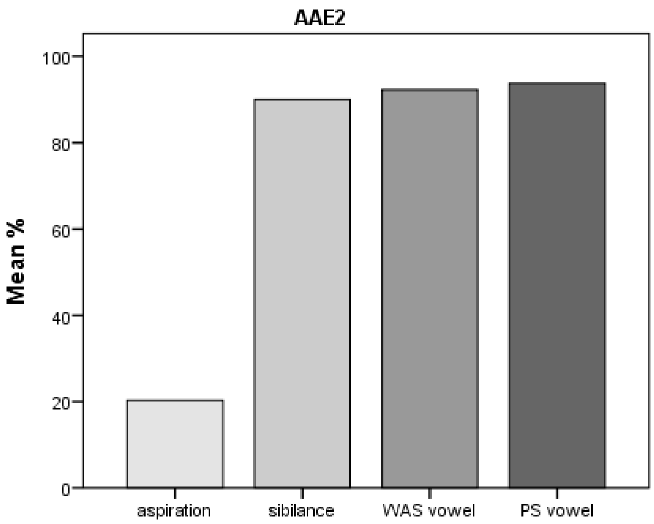


Figure 24. *Mean accuracy for all variables (AAE2)*

GAE2

The identification of aspiration (*M* = 16.67, *SD* = 22.63) by GAE speakers of intermediate level (*N* = 54) was also lower than of sibilance (*M* = 94.61, *SD* = 12.42), as shown in Figure 25. They showed a slightly higher performance for sibilance than for CS vowel (*M* = 93.66, *SD* = 17.33), but slightly lower than WAS vowel (*M* = 94.95, *SD* = 14.5). A Wilcoxon test revealed that the identification of aspiration was significantly lower than that of sibilance [Z = -6.28, p < .001] but no significant differences were found between any of the other conditions.

For this group, the identification of aspiration ranged between 0% and 100%, with 33.3% of them obtaining 0% correct answers, 9.3% of them identifying 50% or more of the sentences correctly, and 3.7% of them answering 100% correctly. For sibilance, identification ranged from 33.33% to 100%, with 96.3% of them obtaining 50% or more correct answers and 68.5% of them rendering 100% identification.

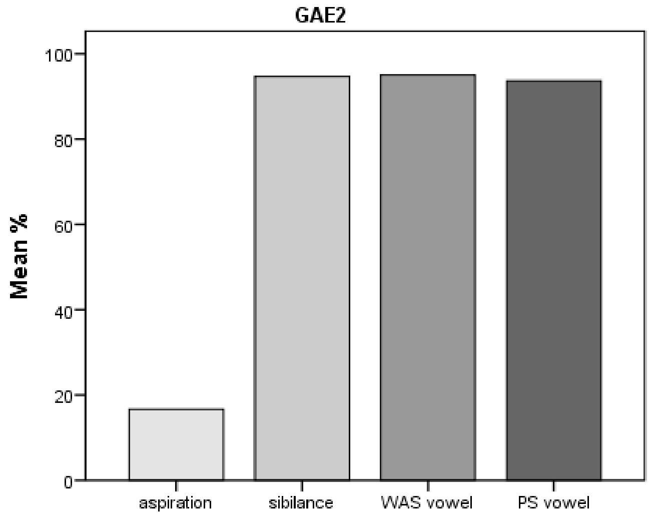


Figure 25. *Mean accuracy for all variables (GAE2)*

5.1.2.2 Identification of aspiration and sibilance across groups

The perception of aspiration was below 25% for all L2 learner groups. Figure 26 below shows the accuracy percentages for the four groups of learners.

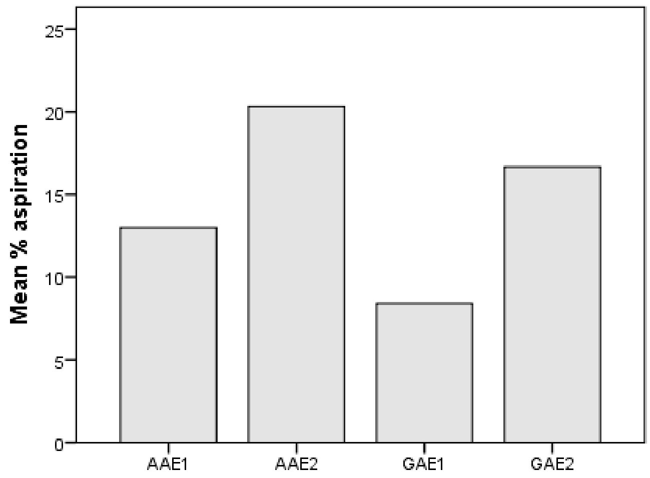


Figure 26. *Perception of aspiration by all groups of listeners*

We analyzed performance (a) by dialect group, (b) by proficiency level, and (c) across groups:

(a) Mann-Whitney tests revealed that aspiration identification was significantly higher for AAE2 listeners than for AAE1 listeners ($U = 895.5$, $p < .05$) but no significant differences were found between GAE1 and GAE2 listeners ($U = 650.5$, $p = .13$), although GAE2 listeners were more accurate than GAE1. Thus, level of proficiency proved to be a significant factor in the perception of aspiration for AAE listeners.

(b) At both levels of proficiency, AAE listeners identified aspiration more accurately than their GAE counterparts. However, differences were not statistically significant (elementary: $U = 1\ 247$, $p = .23$; intermediate: $U = 551.5$, $p = .19$).

(c) AAE1 listeners also perceived aspiration similarly to GAE2 listeners ($U = 2\ 456.5$, $p = .52$). In any case, what was clear is that AAE2 listeners performed significantly more accurately than GAE1 listeners ($U = 218$, $p < .01$).

The perception of sibilance was well above 80% for all groups. Figure 27 below shows the accuracy percentages for the four groups of learners together.

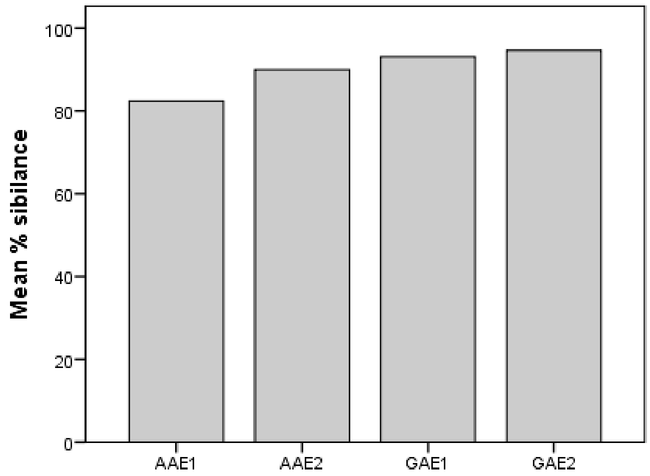


Figure 27. *Perception of sibilance by all groups of listeners*

As with aspiration, we analyzed performance (a) by dialect group, (b) by proficiency level, and (c) across groups:

(a) The identification of sibilance was again significantly higher for AAE2 listeners than for AAE1 listeners (U = 821.5, $p < .05$), and also higher for GAE2 than for GAE1 listeners (U = 620.5, $p = .05$)[19]. In this case, level of proficiency proved to be a significant factor in the perception of sibilance.

(b) Among listeners within the same level of proficiency, we found that GAE1 listeners identified sibilance significantly better than AAE1 listeners (U = 879.5, $p = .001$) and that GAE2 listeners also performed significantly better than AAE2 listeners (U = 477.5, $p < .05$). For sibilance, GAE listeners seemed to have an advantage over AAE listeners at the two levels of Spanish.

[19] Although a level of significance of < .05 was assumed, this is on the verge of statistical significance.

(c) In fact, AAE2 listeners' identification of sibilance was similar to that of GAE1 listeners', i.e., L1 dialect prevailed over L2 proficiency in this case.

In light of these findings, the identification of aspiration [h] and sibilance [s] across listener groups in this current study can be summarized as follows: as expected, aspiration was significantly less accurately perceived than sibilance by each individual group of listeners. Although AAE listeners outperformed GAE listeners, this was not statistically significant (U = 4 798, p = .43). What was significant was that intermediate level listeners performed significantly better than elementary level listeners (U = 4 220.5, p = .05). The case of sibilance identification was different. It seemed to work to the advantage of GAE listeners (U = 2 879, p <. 001), and of intermediate level listeners (U = 3 011, p < .001). For both L2 variants, elementary level listeners obtained the lowest scores overall; GAE1 for aspiration and AAE1 for sibilance, which coincides with the findings from our previous pilot study.

Order of Stimuli

At this point, we analyzed how the order of presentation of the stimuli in the identification task affected the identification of the stimuli. In general, the order of stimuli did not have a correlation with the identification of the target stimuli. Nevertheless, although weak, some correlations were found. There was a positive correlation between the identification of aspiration in second-person verbs for GAE1 listeners (N = 271, r = .119, p = .05). Likewise, for GAE2 listeners, there was also a positive correlation between their identification of WAS third-person verbs and the order of the test. This implies that these listeners identified these tokens worse as the order of the test progressed. For AAE1 listeners, There was a negative correlation between their identification of WAS third-person verbs (N = 733, r = -.113, p < .005), as well as CS third-person verbs (N = 758, r = -.097, p < .01) and CS singular nouns (N = 686, r = -.121, p < .005). This implies that, as the order of the stimuli progressed, their identification of these tokens improved. Finally, there was no significant correlation for the AAE2 group of listeners.

5.1.2.3 Results by years of instruction

In this section, we explore the effect of number of years of instruction on perception, although we should take into account that the type of such instruction was not measured in this experiment. As stated in Section 3.4, years of instruction were classified into four groups (less than 1 year, 1 to 3 years, 3 to 5 years, and more than 5 years). Kruskal-Wallis tests revealed that the number of years of instruction was not significant for any of the groups of listeners individually (AAE1: $\chi^2(3) = 3.25$, p = .36; AAE2: $\chi^2(2) = 4.10$, p = .13; GAE1: $\chi^2(2) = 1.9$, p = .39; GAE2: $\chi^2(3) = 2.14$; p = .54).

Nevertheless, as shown in Table 7, the perception of aspiration decreased with amount of instruction for AAE1 listeners and increased with years of instruction for GAE1 listeners. Additionally, the identification of sibilance progressively increased with instruction for both groups at elementary level of Spanish. At the elementary level of Spanish, it was AAE listeners with less than 1 year of instruction who identified aspiration best and those with more than 5 years of instruction identified it worst. At intermediate level, AAE listeners with 3-5 years of instruction obtained the highest scores for aspiration and GAE listeners with less than 1 year of instruction obtained the lowest scores (taking into account that there were no participants with less than 1 year of instruction in the AAE group).

Table 7

Perception of aspiration and sibilance by years of instruction and listener group

				aspiration		sibilance	
			N	*M*	*SD*	*M*	*SD*
AAE1	years of instruction	< 1	20	15.10	15.05	80.33	18.73
		1 to 3	59	13.57	17.39	80.83	18.79
		3 to 5	14	9.66	15.24	88.61	11.24
		> 5	4	5.88	11.77	93.18	13.63
AAE2	years of instruction			-	-	-	-
		< 1	0				
		1 to 3	9	15.71	26.05	90.08	15.67
		3 to 5	11	23.51	21.49	91.93	17.11
		> 5	5	21.52	11.39	85.31	12.31
GAE1	years of instruction						
		< 1	14	7.01	10.64	92.10	9.34
		1 to 3	8	6.49	7.17	92.73	7.93
		3 to 5	8	12.79	11.46	94.97	8.86
		> 5	0	-	-	-	-
GAE2	years of instruction						
		< 1	2	3.13	4.42	97.22	3.93
		1 to 3	27	14.37	18.63	93.53	15.92
		3 to 5	17	21.70	31.62	96.71	5.48
		> 5	8	17.11	12.80	93.12	11.65

For sibilance, GAE1 listeners with 3-5 years of instruction were the most accurate and AAE1 with less than 1 year of instruction were the least accurate. At intermediate level, GAE listeners with less than 1 year of instruction obtained the highest scores (although $n = 2$), followed by those with 3-5 years of instruction. AAE listeners with more than 5 years of instruction obtained the lowest scores.

Upon analyzing identification performance across groups and years of instruction, we observed that no significant differences in perception were found between the two GAE groups or between AAE2 listeners and either of the GAE groups. Statistical differences worth mentioning were observed between AAE1 listeners and the other three groups.

Perhaps the most salient finding was that, with less than 1 year of instruction, AAE1 listeners significantly outperformed GAE1 participants in the identification of aspiration ($U = 83.5$, $p < .05$), and GAE1 listeners outperformed AAE1 participants in the perception of sibilance, which was on the verge of significance ($U = 87$, $p = .06$). With 1-3 years of instruction, the three groups perceived sibilance significantly better than AAE1 listeners (GAE1: $U = 128.5$, $p < .05$; GAE2: $U = 308$, $p < .001$; AAE2: $U = 155$, $p < .05$). GAE2 listeners also outperformed AAE1 listeners in the perception of sibilance with 3-5 years of instruction ($U = 64.5$, $p < .05$). Additionally, AAE2 listeners also perceived aspiration significantly better than AAE1 listeners with 3-5 years of instruction ($U = 33.5$, $p < .05$).

Therefore, as to the number of years of instruction, it seemed to particularly affect AAE1 listeners in comparison with the rest of the groups. Once more, at elementary level with less than 1 year of instruction, in which we could regard listeners as truly “naïve” in the language, we observed that AAE listeners significantly outperformed GAE participants in the perception of aspiration, while GAE listeners identified sibilance significantly better than AAE participants.

5.1.2.4 Results by syntactic context

The perception of aspiration and sibilance is described in this section in relation to the syntactic context in which target words were embedded: content sentences with second-person verbs in the present tense (2PV), in which the morphological marker *–s* determines verb person, and carrier sentences with plural nouns (PN), in which the morphological marker *–s* determines plurality. Examples of such sentences are as follows:

(2PV) Nunca comes nada dulce (*You never eat anything sweet*)

(PN) Digo perros por la tarde (*I say dogs in the afternoon*)

Aspiration

Overall, aspiration was significantly better identified by all listeners when target words were second-person verbs (14.12%) than plural nouns (13.49%) ($Z = -11.43$, $p < .001$). In spite of this general trend, AAE2 listeners (Figure 28) perceived aspiration in verbs significantly better than in nouns ($Z = -5.63$, $p < .001$).

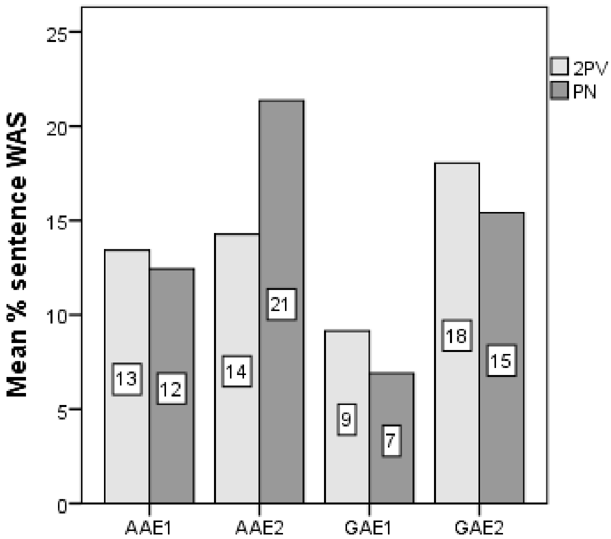


Figure 28. *Perception of aspiration in syntactic context by listener group*

Analyses in terms of (a) listener dialect, (b) level of proficiency, and (c) across groups revealed the following:

(a) The perception of aspiration in 2PV between the two AAE groups was statistically similar ($U = 118\ 835$, $p = .63$), whereas aspiration in PN was significantly higher for AAE2 ($U = 55\ 755$, $p < .001$). GAE2 listeners performed significantly better than GAE1 in both contexts (2PV: $U = 19\ 440$, $p < .001$; PN: $U = 41\ 310$, $p < .001$).

(b) At elementary level of proficiency, AAE listeners outperformed GAE participants in both contexts (2PV: $U = 81\ 480$, $p < .001$; PN: $U = 93\ 120$, $p < .001$). At intermediate level, GAE2 performed significantly better in 2PV sentences ($U = 47\ 925$, $p < .001$) while AAE2 listeners were more accurate in PN sentences ($U = 41\ 850$, $p < .001$).

(c) Both intermediate groups outperformed elementary groups of the opposite dialect in both contexts.

In general, all groups identified the morphological marker *–s* when realized as aspiration better in second-person verbs than in plural nouns, with the exception of AAE2 listeners, who identified nouns more accurately. In fact, they outperformed the rest of the groups in the identification of the plurality marker *–s*. Likewise, GAE2 were the most accurate for the identification of the second-person verb marker *–s*. As we have seen before, GAE1 listeners were less accurate than the rest of listeners at identifying the morphological marker *–s* when realized as aspiration in both verbs and nouns.

Sibilance

The identification of sibilance by all listeners as a group was, on the contrary, significantly higher in carrier sentences (90.79%) than in content sentences (83.88%) ($Z = -21.32$, $p < .001$). This pattern was followed by each individual group of listeners, as shown in Figure 29.

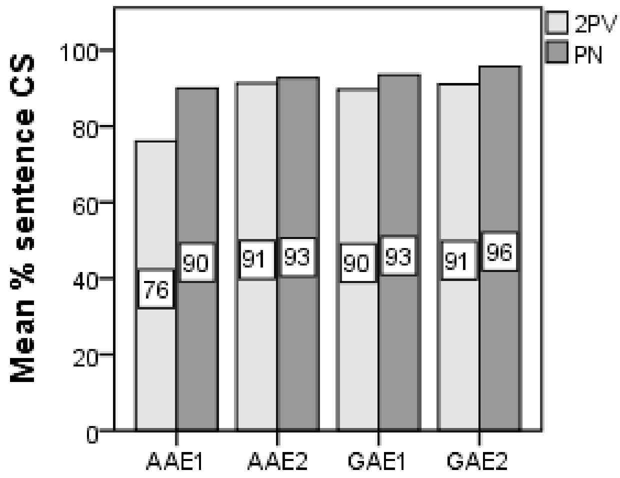


Figure 29. *Perception of sibilance in syntactic context by listener group*

Upon analyzing performance in terms of (a) listener dialect, (b) level of proficiency, and (c) across groups, we found:

(a) Within the AAE group, intermediate listeners significantly outperformed elementary listeners in both contexts (2PV: U = 19 400, p < .001; PN: U = 61 837.5, p < .001). For GAE listeners, elementary level participants outperformed intermediate level listeners in 2PV context (U = 66 420, p < .001) but both performed similarly in PN context (U = 80 190, p = .81).

(b) At elementary level of Spanish, GAE listeners' perception was significantly more accurate than that by AAE participants in both contexts (2PV: U = 46 560, p < .001; PN: U = 58 200, p < .001). At intermediate level, both GAE and AAE listeners performed similarly in 2PV context (U = 62 775, p < .112) while GAE listeners were more accurate in PN sentences (U = 58 050, p = .001).

(c) Across groups of listeners, the GAE2 group outperformed AAE1 listeners in both syntactic contexts (2PV: U = 26 190, p < .001; PN: U = 57 618, p <

.001), while the GAE1 group performed significantly better than AAE2 listeners in PN sentences (U = 31 500, p < .001) but similarly in 2PV contexts (U = 34 875, p = .15).

In this case, all groups identified the morphological marker –*s* in plural nouns better than in second-person verbs. In general, all groups performed similarly, with the exception of AAE1 listeners, who were the least accurate at identifying the morphological marker –*s* in both nouns and verbs.

To check if these patterns are also true for singular nouns and third–person verbs, we also analyzed these two types of target words in the two L2 dialects. As Table 8 shows, all groups identified singular nouns in CS sentences significantly better than third-person verbs (AAE1: Z = -9.85, p < .001; AAE2: Z = -5, p < .001; GAE1: Z = -5.48, p < .001; GAE2: Z = -7.35, p < .001), just as they identified plural nouns significantly better than second-person verbs in CS sentences. However, in the case of WAS sentences, we found that both AAE1 and GAE1 listeners identified singular nouns significantly better than third-person verbs (AAE1: Z = -9.85, p < .001; GAE1: Z = -5.48, p < .001), as opposed to their identification of WAS aspiration in second-person verbs, which was significantly more accurate than in plural nouns. GAE2 listeners were consistent in the sense that they also identified WAS verbs ending in vowel significantly better than nouns (Z = -7.35, p < .001). AAE2 listeners, however, identified third-person verbs significantly better that plural nouns this time (Z = -5, p < .001).

Table 8

Identification of WAS and CS third-person verbs and singular nouns

	WAS		CS	
	3PV	SN	3PV	SN
	M	*M*	*M*	*M*
AAE1	89.63	93.40	87.60	92.71
AAE2	95.24	93.01	95.61	97.30
GAE1	97.12	98.28	95.79	98.83
GAE2	95.32	93.20	91.78	93.43

Therefore, all groups consistently identified CS nouns more accurately than verbs, whether ending in [s] or vowel, while WAS sentences rendered several outcomes. Both elementary-level groups identified verbs ending in [h] better than nouns, and nouns ending in vowel better than verbs. AAE2 listeners were the opposite: aspiration was better identified in nouns than in verbs, while verbs ending in vowel were better identified than nouns. GAE2 listeners identified verbs in both conditions more accurately than nouns.

Reaction Times

We also measured the reaction times (RTs) of the L2 listeners, i. e., how long they took to choose an answer after listening to each stimulus. Table 9 shows their RTs in milliseconds (ms) for each type of sentence by L2 dialect.

Table 9

Reaction times (ms) in both conditions in the four syntactic contexts by group of listeners

		GAE1	GAE2	AAE1	AAE2
		M	*M*	*M*	*M*
3PV	WAS	1142	1319	1307	1108
	CS	1061	980	911	826
2PV	WAS	1447	1711	2088	1684
	CS	839	986	977	936
SN	WAS	988	1231	1227	1120
	CS	834	939	805	764
PN	WAS	1551	1400	1753	1307
	CS	867	965	1065	1155

In general, RTs were significantly higher for WAS sentences than for CS sentences in all syntactic contexts, regardless of presence or absence of aspiration or sibilance in the target stimuli, as shown in Table 10.

Table 10

Wilcoxon test and statistical probability values

		GAE1	GAE2	AAE1	AAE2
3PV	Z	-2.97	-2.99	-5.90	-3.51
	p	< .005	< .005	< .001	< .001
2PV	Z	-5.82	-6.31	-8.97	-5.23
	p	< .001	< .001	< .001	< .001
SN	Z	-2.46	-3.93	-6.42	-2.45
	p	< .05	< .001	< .001	< .05
PN	Z	-4.62	-5.46	-8.27	-2.83
	p	< .001	<. 001	< .001	< .005

5.1.2.5 Results by of phonetic context

As we saw in Section 1.2.1, the aspiration of implosive /s/ in WAS causes certain changes in the following sounds. Concerning the phonetic context and the speakers in this study, we will now revisit those changes:

i) Fricatization of voiced stops:

/b/ → [v], /d/ → [ð], /g/ → [x]

ii) Aspiration in voiceless stops:

/p/ → [p^h], /t/ → [t^h], /k/ → [k^h]

iii) Reduplication and gemination of nasal and lateral sounds:

/m/ → [^{h}m.m], /n/ → [^{h}n.n], /l/ → [^{h}l.l]

iv) Appears as full [h] before vowel:

/V/ → [hV]

An analysis of phonetic context showed that GAE1 listeners generally identified aspiration in all contexts less accurately than the rest of the groups, except for aspiration before [t^h], which GAE2 listeners identified worst. Overall, intermediate level listeners identified all contexts best except for aspiration before [p^h], for which AAE1 listeners rendered the highest scores. In fact, the perception of aspiration followed by the three voiceless stops by AAE1 listeners was statistically the same [$\chi^2(2) = .00$, $p = 1$]. Nonetheless, individual-group performance showed that the perception of aspiration before [ð] was the most accurate for AAE1 and GAE2, while AAE2 and GAE1 listeners perceived aspiration before [t^h] most accurately, as shown in Table 11 below.

Table 11

Perception of aspiration in phonetic context by listener group

			group				
			AAE1	AAE2	GAE1	GAE2	**TOTAL**
WAS	phonetics	v	13.75	27.53	10.43	22.15	**18.47**
		ð	26.10	25.00	16.33	37.94	**26.34**
		x	11.25	11.33	2.86	12.55	**9.50**
		k^h	12.75	7.15	6.67	14.70	**10.32**
		p^h	12.56	7.15	4.55	10.78	**8.76**
		t^h	16.92	27.65	18.06	12.16	**18.70**
		^{h}m.m	16.34	21.91	8.26	17.22	**15.92**
		^{h}n.n	8.77	16.59	10.17	16.86	**13.10**
		^{h}l.l	8.64	27.06	0	14.02	**12.43**
		V	2.25	6.82	2.94	8.95	**5.24**

AAE1 listeners identified sibilance before all phonetic contexts less accurately than the rest of the groups. The highest scores before [β̞], [ɣ̞], and [V] were for GAE1 listeners, surpassed by GAE2 participants in the identification of [s] before [ð̞], [m],

[n], and [l], and by AAE2 listeners in the identification of [s] before voiceless stops. Individually, AAE1 and AAE2 listeners identified sibilance before [t] most accurately, while GAE1 listeners identified sibilance before [β̞], [ɣ̞] and [V] best, and GAE2 listeners before [k] (see Table 12 below).

Table 12

Perception of sibilance in phonetic context by listener group

			group				
			AAE1	AAE2	GAE1	GAE2	**TOTAL**
CS	phonetics	β̞	77.04	94.74	100	95.82	**91.90**
		ð̞	80.88	84.75	78.58	89.81	**83.51**
		ɣ	80.95	97.37	100	95.75	**93.52**
		k	87.99	97.73	97.62	96.29	**94.91**
		p	81.65	97.62	92.76	93.17	**91.30**
		t	89.18	97.83	94.22	91.62	**93.21**
		m	80.85	82.49	84.85	93.77	**85.49**
		n	80.91	89.58	83.54	90.15	**86.05**
		l	76.81	92.04	85.58	92.87	**86.83**
		V	79.57	87.50	100	90.75	**89.46**

Place of articulation

An analysis of phonetic context in terms of the place of articulation of the sounds following [h] was carried out. According to the new categories established, a restructuration of the allophones that resulted from the influence of [h] was done as follows:

i) Bilabial [p^h], [^{h}m.m]

ii) Dental [t^h]

iii) Velar [x], [k^h]

iv) Alveolar [^{h}n.n], [^{h}l.l]

v) Labiodental [v]

vi) Interdental [ð]

vii) Glottal [h]

Figure 30 shows the pattern of perception followed by all listeners as a group in terms of the seven categories established above. Overall, the perception of aspiration before the interdental sound was the highest (*M* = 27.64, *SD* = 9.67), followed by labiodental (*M* = 17.14, *SD* = 6.42) and dental sounds (*M* = 17.14, *SD* = 8.25), bilabial sounds (*M* = 13.17, *SD* = 5.84), alveolar sounds (*M* = 11.53, *SD* = 7.22), velar sounds (*M* = 11.04, *SD* 5.08), and glottal (*M* = 4.66, *SD* = 4.01).

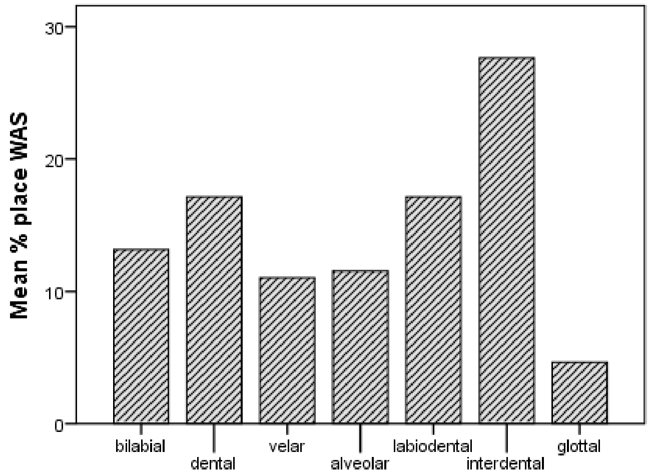


Figure 30. *Overall perception of aspiration according to place of articulation*

The perception of aspiration before the seven places of articulation by each individual group of listeners was as shown in Figure 31: AAE1 listeners' highest accuracy percentage was for aspiration before interdental (26.1%) and lowest before glottal (2.25%). This corroborates the results reported in Table 6, which indicate that AAE1 listeners' highest percentage was for aspiration before [ð] and their lowest before vowel. For this group, their identification of aspiration before bilabial and labiodental sounds was statistically the same (Z = 0, p = 1). For AAE2 listeners, the perception of aspiration before dental sounds was the most accurate (27.65%), statistically similar to their identification of aspiration before labiodental (27.53%) (Z = 0, p =1), while their lowest was before glottal (6.82%). Once more, this supports their highest scores for aspiration before [t^h] and [v], and their lowest before vowel. Additionally, the perception of aspiration before interdental and alveolar sounds was found to be statistically the same for this group, i.e., no significant differences were found ($Z = 0$, $p = 1$).

The perception of aspiration before dental was also the highest for GAE1 listeners (18.06%), while their lowest score was before glottal (2.94%). This partly corroborates the results observed in Table 7. GAE1 listeners' highest score was indeed for aspiration before [t^h]; however, their lowest score was before [^{h}l.l], followed by aspiration before vowel. The fact that [^{h}n.n] is also alveolar increased their score for aspiration before this place of articulation. Additionally, their perception of aspirated /s/ before velar and alveolar sounds was similar ($Z = -1.80$, $p = .072$), and before velar and glottal sounds it was statistically the same ($Z = 0$, $p = 1$). Finally, GAE2 listeners' highest accuracy percentage was also for aspiration before interdental sounds (37.94%), whereas their lowest score was before glottal (8.95%), supporting their highest score for aspiration before [ð] and their lowest score before vowel. For this group, the perception of aspiration before bilabial-velar and bilabial-alveolar sounds was statistically the same ($Z = 0$, $p = 1$).

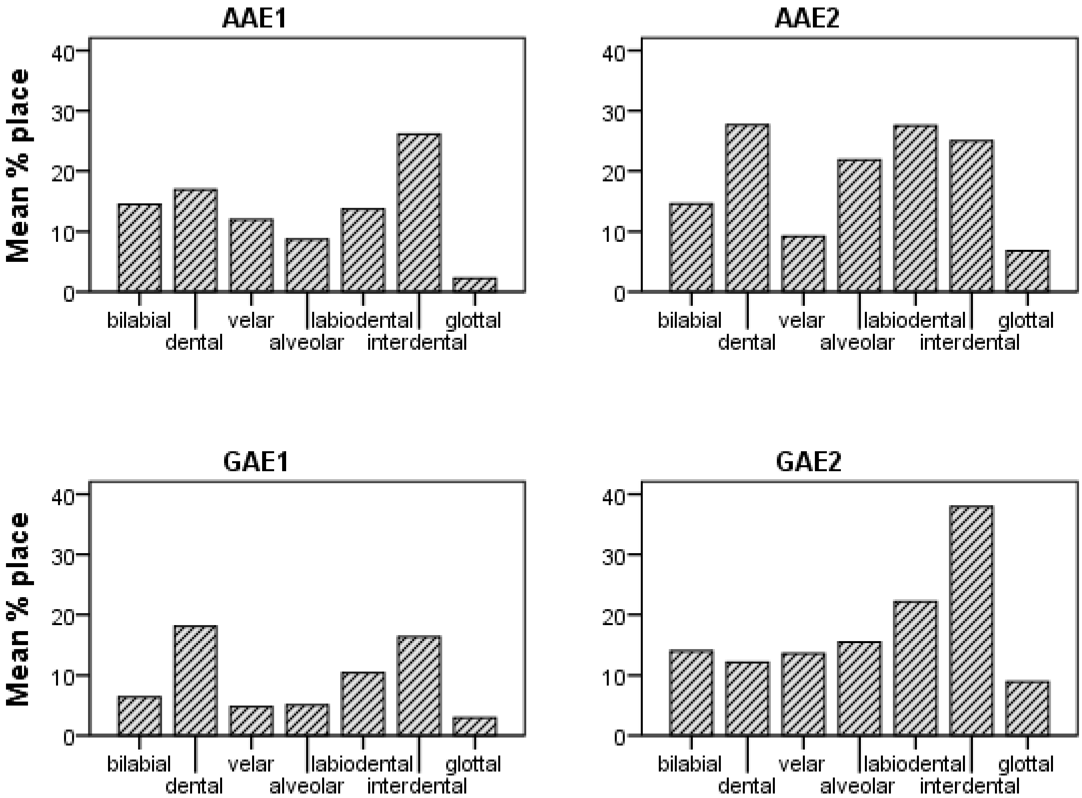


Figure 31. *Perception of aspiration according to place of articulation by listener group*

The analysis of the performance across the four groups of listeners was done in terms of (a) listener dialect, (b) level of proficiency, and (c) across groups:

(a) AAE2 listeners were significantly more accurate than AAE1 listeners in the perception of aspiration before dental (U = 2 425, $p < .001$), alveolar (U = 0, $p < .001$), labiodental (U = 0, $p < .001$), and vowel sounds, but not before velar (U = 2 425, $p < .001$), in which AAE1 listeners performed better. They both identified aspiration before bilabial (U = 4850, p = 1) and interdental sounds (U = 2 425, p = 1) in a similar manner. GAE2 listeners were significantly more accurate than GAE1 listeners for all conditions except for aspiration before dental sounds, for which GAE1 performed better (U = 1 620, $p < .001$).

(b) At elementary level, AAE listeners' performance was more accurate than that of GAE listeners' for all conditions except for aspiration before dental and vowel, in which both groups of listeners performed similarly ($U = 5820$, $p = 1$). Once more, the two groups at intermediate level rendered diverse results. GAE2 listeners were more accurate than AAE2 in aspiration before velar ($U = 8100$, $p < .001$) and interdental ($U = 0$, $p < .001$) sounds, whereas AAE2 listeners were more accurate before dental ($U = 1350$, $p < .001$), alveolar ($U = 8100$, $p < .001$), and labiodental ($U = 1350$, $p < .001$) sounds. Both groups performed similarly before bilabial contexts ($U = 9450$, $p = .07$) and before vowel ($U = 2700$, $p = 1$).

(c) Additionally, AAE1 identified aspiration before bilabial ($U = 36\ 666$, $p = .01$) and dental ($U = 5238$, $p < .001$) sounds more accurately than GAE2 listeners. AAE2 listeners outperformed GAE1 listeners in all contexts.

In this case, the analysis of the perception of aspiration according to the place of articulation of the following sounds revealed that all groups of listeners identified aspiration before vowel worse than before any other place of articulation, particularly GAE1 and AAE1 listeners. AAE1 and GAE2 listeners individually identified aspiration before interdental better than any other condition. In fact, GAE2 listeners obtained the highest score among the four groups for this context. On the contrary, GAE1 and AAE2 individually identified aspiration before dental best, especially AAE2, who obtained the highest score for this condition overall. Particularly interesting is the fact that GAE2 listeners, in this case, identified aspiration before dental sounds less accurately than any other group.

The classification of phonetic context after sibilance according to place of articulation was done as follows:

i) Bilabial [β̞], [p], [m]

ii) Dental [ð̞], [t]

iii) Velar [ɣ̞], [k]

iv) Alveolar [n], [l]

v) Glottal [V]

As a group (Figure 32), highest accuracy was for sibilance before velar sounds (*M* = 91.17, *SD* = 6.85), followed by dental sounds (*M* = 87.48, *SD* = 5.48), bilabial sounds (*M* = 86.90, *SD* = 7.62), glottal (*M* = 86.44, *SD* = 7.37), and finally before alveolar sounds (*M* = 84.45, *SD* = 5.96).

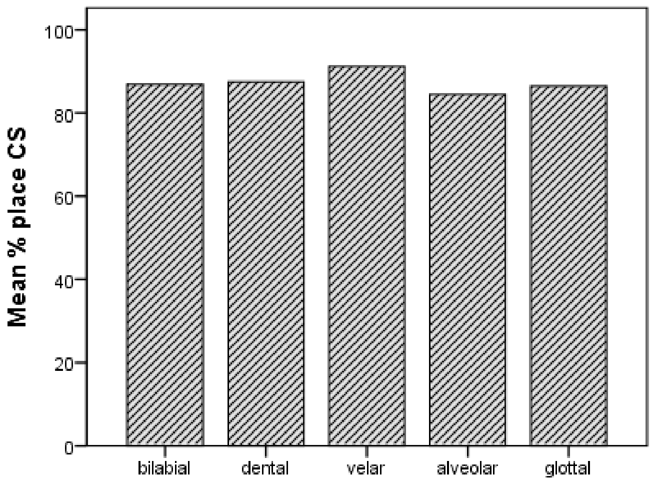


Figure 32. *Overall perception of sibilance according to place of articulation*

Figure 33 shows that the perception of sibilance according to place of articulation for AAE1 listeners was statistically similarly accurate before dental (85.03%) and velar sounds (84.47%), followed by bilabial sounds (79.85%), and also similar before glottal (79.57%), and alveolar sounds (78.86%). So far, this corroborates the findings seen in Table 4 above, in the sense that AAE1 listeners' highest score was for sibilance before [t] and their lowest score for sibilance before [l]. For AAE2 listeners, their highest score was for sibilance before velar sounds (97.55%), followed by bilabial sounds (91.62%), dental (91.29%), alveolar sounds (90.81%), and glottal (87.50%). Statistically, the perception of sibilance before the last three types of contexts was the same (Z = 0, p =1). Place of articulation alone fails to explain this group of listeners' highest [t] and lowest scores [m].

The perception of sibilance before glottal was the most accurate for GAE1 listeners (100%), followed by sibilance before velar sounds (98.81%), bilabial sounds (92.54%), dental (86.40%), and alveolar sounds (84.56%). For this group, the perception of [s] before dental and alveolar sounds was statistically the same (Z = 0, p = 1). In this case, these results support the findings that these listeners perceived sibilance before vowel, [β̞] and [ɣ̞] most accurately while they identified [ð̞] the least accurately. For GAE2 listeners, their highest score was for velar sounds (96.02%), followed by bilabial sounds (94.25%), alveolar sounds (91.51%), glottal (90.75%), and finally dental (90.71%). In this case, the identification of dental and glottal sounds, as well as the perception of glottal and alveolar sounds, was statistically the same. Once more, place of articulation corroborates their highest identification of sibilance before [k] and their lowest score before [ð̞].

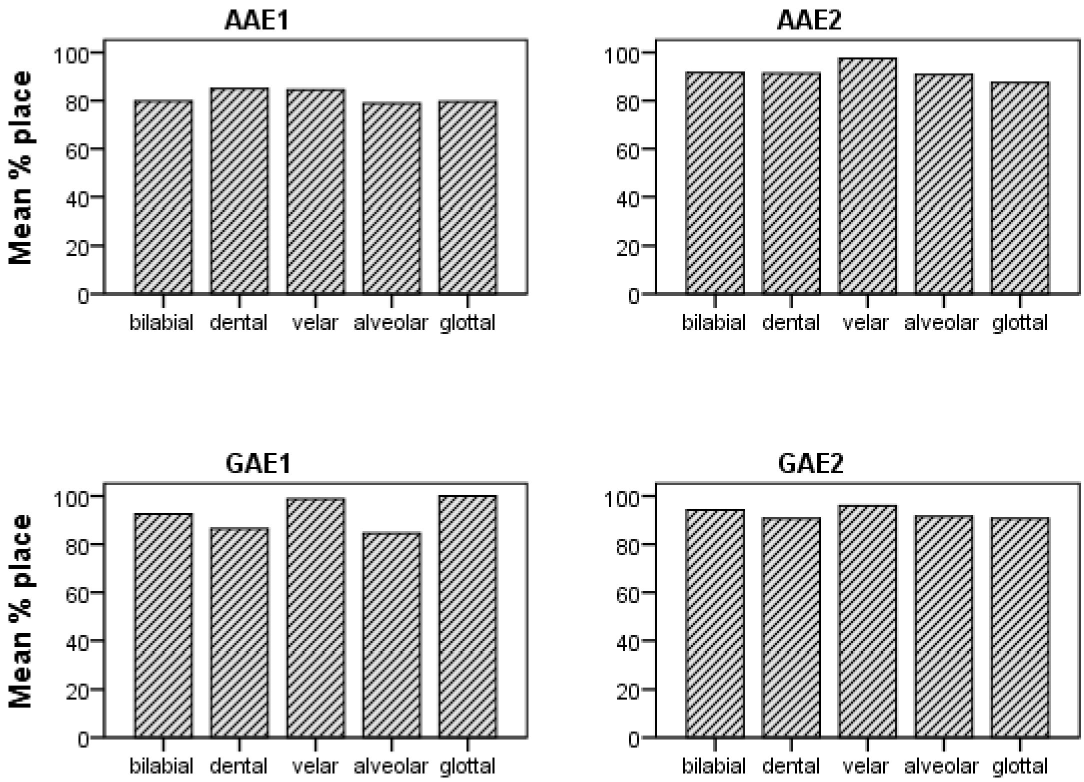


Figure 33. *Perception of sibilance according to place of articulation by listener group*

An analysis of the performance across the four groups of listeners in terms of (a) listener dialect, (b) level of proficiency and (c) across groups revealed that:

(a) AAE2 listeners outperformed AAE1 listeners for all contexts. Both groups of GAE participants perceived dental sounds similarly (U = 3 240, p =1), while GAE2 identified bilabial (U = 4 860, p <.001) and alveolar (U = 0, p <.001) sounds significantly better than GAE listeners, and these identified velar and glottal sounds more accurately than their intermediate counterpart (U = 0, p <.001).

(b) At elementary level of proficiency, GAE listeners outperformed AAE participants for all contexts (U = 0, p <.001) except for sibilance before dental sounds, which both groups identified similarly (U = 5 820, p =1). At intermediate level, GAE listeners identified sibilance before alveolar (U =

1 350, < .001) and glottal (U = 0, <.001) sounds more accurately while AAE listeners performed better before velar sounds (U = 0, p <.001). Their perception of sibilance before dental sounds was statistically the same (U = 2 700, p = 1) and their identification of bilabial sounds was not significantly different, although AAE listeners obtained higher scores (U = 5 400, p = .16).

(c) Across groups, GAE2 listeners outperformed AAE1 listeners for all contexts. AAE2 participants identified sibilance before dental (U = 750, p <.001) and alveolar (U = 0, p < .001) sounds more accurately than GAE1 listeners, while these identified sibilance before velar (U = 750, p < .001) and glottal (U = 0, p <.001) sounds more accurately. Their perception of sibilance before bilabial sounds was not statistically significant (U = 3 000, p = .22), despite the higher scores obtained by the GAE1 group.

In this case, the perception of sibilance according to place of articulation revealed that both GAE2 and AAE2 listeners individually identified sibilance before velar sounds best, but worst before vowel, and dental and alveolar sounds. Likewise, AAE1 listeners identified sibilance before alveolar sounds and vowel worst, but before dental and velar best. GAE1 listeners also identified sibilance before dental and alveolar worst but before vowel best. What they all have in common is their lowest identification of sibilance before alveolar sounds and that three out of the four groups identified sibilance before velar sounds best. Particularly interesting is the fact that GAE1 listeners outperformed the rest of the groups in the identification of sibilance before velar sounds and vowel.

Manner of articulation

An analysis of phonetic context in terms of the manner of articulation of the sounds following [h] was carried out. According to the new categories established, a restructuration of the allophones that resulted from the influence of [h] was done as follows:

i) Stop [k^h], [t^h], [p^h]
ii) Fricative [v], [ð], [x]
iii) Nasal [^{h}m.m], [^{h}n.n]
iv) Lateral [^{h}l.l]
v) Open [hV]

Figure 34 shows the overall performance in terms of manner of articulation of the sounds in the phonetic context following aspiration. Fricatives were the sounds best perceived (M = 18.39, SD = 10.03), followed by nasals (M = 14.06, SD = 5.02), voiceless stops (M = 13.03, SD = 7.28), and finally the lateral sound (M = 11.02, SD = 9.09), and open (M = 4.66, SD = 4.01).

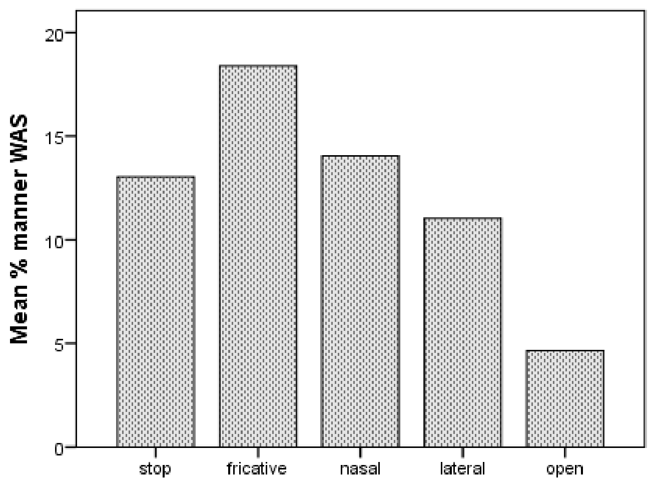


Figure 34. *Overall perception of aspiration according to manner of articulation*

Upon analysis of performance by individual groups of listeners, we observed (Figure 35) that AAE1 listeners perceived aspiration before fricatives best (17.03%), followed by stops (14.07%), nasals (12.56%), lateral (8.64%), and open (2.25%). As was the case with place of articulation, manner of articulation also supports their highest score for aspiration before [ð] and their lowest before vowel. AAE2 listeners showed higher accuracy for aspiration before the lateral sound (27.06%) than before fricatives (21.29%), nasals (19.25%), and lastly stops (13.98%) and open (6.82%). In this case, manner of articulation fails to support their highest scores for aspiration before [t^h] but it explains their lowest score before vowel.

GAE1 listeners' highest score was for aspiration before fricatives (9.87%), followed by stops (9.75%), nasals (9.22%) and open (2.94%). They did not perceive aspiration before the lateral sound (0%). Nevertheless, there were no significant differences between their perception of aspiration before fricatives, stops and nasals for this group ($Z = 0$, $p = 1$). Manner of articulation also corroborates their highest score for aspiration before [t^h] and, unlike place of articulation, it now explains their lowest score before [^{h}l.l]. Finally, GAE2 listeners also showed the highest identification accuracy for aspiration before fricatives (24.21%), followed by nasals (17.04%), the lateral sound (14.02%), and finally stops (12.55%) and open (8.95%). As was the case with place of articulation, manner of articulation also supports their highest scores for aspiration before [ð] and their lowest score before vowel.

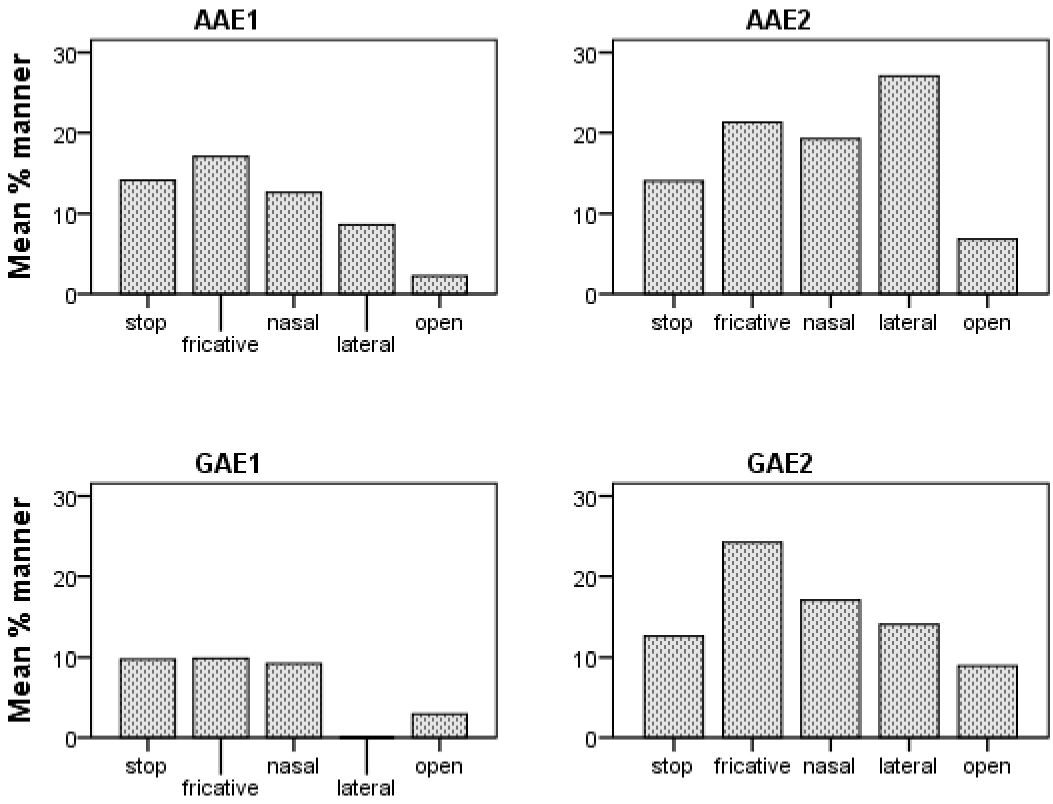


Figure 35. *Perception of aspiration according to manner of articulation by listener group*

An analysis of the performance across the four groups of listeners was again carried out in terms of (a) listener dialect, (b) level of proficiency, and (c) across groups:

(a) Aspiration before voiceless stops was better perceived by AAE1 listeners than by AAE2 listeners ($U = 33\ 950$, $p < .001$), while these identified aspiration before fricatives ($U = 29\ 100$, $p < .001$) and open better than AAE1 listeners. In this case, AAE2 listeners also showed significantly higher accuracy for aspiration before nasals ($U = 4\ 850$, $p < .001$) and the lateral sound ($U = 0$, $p < .001$). For GAE listeners, intermediate level listeners performed significantly better than elementary level listeners for all conditions.

(b) AAE1 listeners significantly outperformed GAE1 listeners in the perception of all conditions except for open, for which both groups performed similarly ($U = 5\ 820$, $p = 1$). GAE2 listeners perceived stops ($U = 20\ 250$, $p < .005$) and open better than AAE2 listeners, while they both perceived fricatives and nasals similarly. For the lateral sound, AAE2 listeners were significantly more accurate.

(c) Additionally, AAE1 listeners outperformed GAE2 listeners in the perception of voiceless stops ($U = 10\ 476$, $p < .001$). Once more, AAE2 listeners performed significantly better than GAE1 listeners for all contexts.

Individually, we can see some similarities between the four groups of listeners. AAE1, GAE1 and GAE2 identified aspiration before fricative sounds best while AAE1, AAE2 and GAE2 identified aspiration before vowel worst. Now, the lowest score for GAE1 was for aspiration before the lateral sound, while AAE2 listeners' highest identification accuracy was precisely for aspiration before the lateral sound. In general, the highest identification accuracy across all groups was for intermediate listeners, with the exception of aspiration before stops, which AAE1 listeners identified better than any of the other groups.

The perception of [s] in phonetic context was classified according to manner of articulation as follows:

i) Stops [p], [t], [k]

ii) Approximants [β̞], [ð̞], [ɣ̞]

iii) Nasals [m], [n]

iv) Lateral [l]

v) Open [V]

Overall performance showed that the highest accuracy was for stops (*M* = 90.86, *SD* = 5.18), followed by approximants (*M* = 86.80, *SD* = 8.26), open, (*M* = 86.44, *SD* = 7.37), nasal (*M* = 84.89, *SD* = 4.84), and then the lateral sound (*M* = 84.14, *SD* = 7.31), shown in Figure 36.

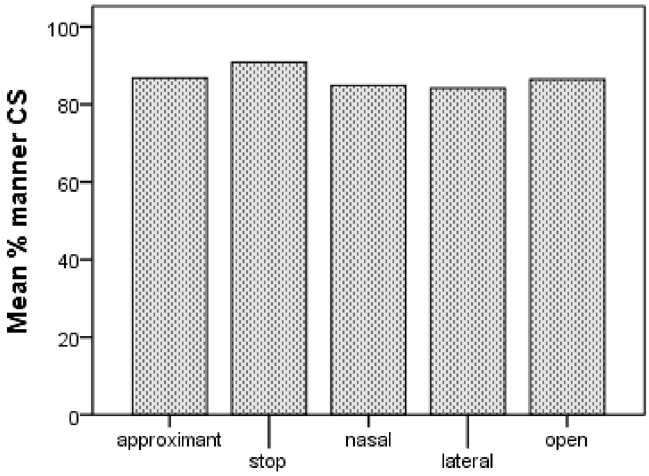


Figure 36. *Overall perception of sibilance according to manner of articulation*

Figure 37 shows that AAE1 listeners identified sibilance before stops most accurately (86.27%), followed by nasals (80.88%) and approximants (79.62%), the identification of which was statistically similar, open (79.57%), and lateral (76.81%). As was the case with place of articulation, manner of articulation additionally supports their highest score for sibilance before [t] and their lowest before [l]. For AAE2 listeners, stops were also their highest score (97.72%), followed by approximants (92.28%), lateral (92.04%), and open (87.50%) and nasal (86.04%), which were similarly perceived. Unlike place of articulation, manner of articulation is now able to explain their highest score for sibilance before [t] and their lowest before [m].

GAE1 participants identified open best (100%), subsequently followed by stops (94.87%), approximants (92.86%), lateral (85.88%), and nasals (84.20%). In this case, and contrary to place of articulation, manner of articulation only accounts for their highest score before vowel, but fails to clearly account for their scores before [β̞] and [ɣ̞] or [ð̞]. Finally, GAE2 listeners perceived approximants (93.79%) and stops (93.69%) in a similar manner, followed by lateral (92.87%), and nasal (91.96%) and open (90.75%), the identification of which was also statistically similar. For this group, manner of articulation also explains their highest score for sibilance before [k] but fails to support their lowest score before [ð̞].

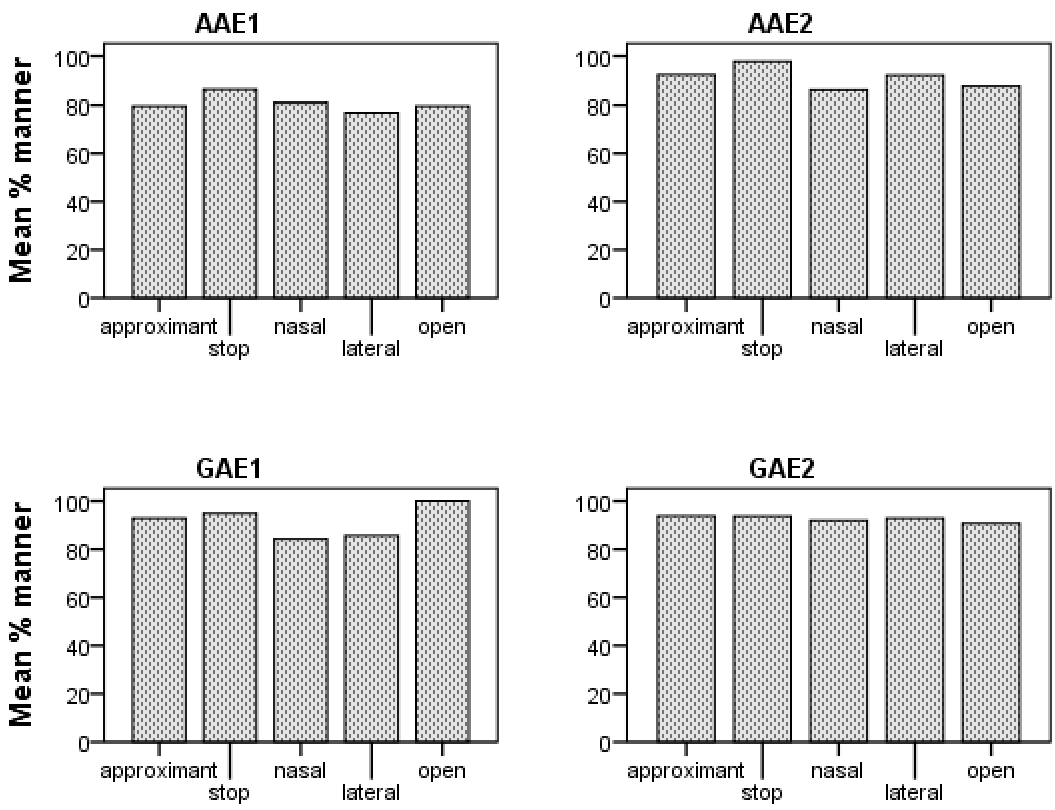


Figure 37. *Perception of sibilance according to manner of articulation by listener group*

An analysis of the performance across the four groups of listeners in terms of (a) listener dialect, (b) level of proficiency, and (c) across groups revealed:

(a) Within the AAE group, intermediate level listeners outperformed elementary level listeners in all cases. For GAE listeners, elementary level participants performed significantly better that intermediate level listeners in stops ($U = 4\ 860$, $p < .001$), approximants ($U = 4\ 860$, $p < .001$), and open ($U = 0$, $p < .001$), while GAE2 performed better in lateral and nasal sounds ($U = 0$, $p < 001$).

(b) At elementary level, GAE listeners outperformed AAE listeners in all cases. At intermediate level, GAE participants performed more accurately in nasals, lateral, and open ($U = 0$, $p <.001$), while AAE listeners identified stops better ($U = 0$, $p < .001$). Their perception of approximants was statistically similar ($U = 5\ 400$, $p = .16$).

(c) GAE2 outperformed AAE1 listeners in all conditions, while GAE1 listeners performed significantly better than AAE2 participants in the identification of aspiration before approximants ($U = 2\ 250$, $p < .001$) and open ($U = 0$, $p < .001$), and both performed similarly for aspiration before nasals ($U = 1\ 500$, $p = 1$).

In this case, the identification of sibilance according to the manner of articulation of the following sounds revealed that all groups identified sibilance before stops most accurately, with GAE2 additionally identifying sibilance before approximants similarly. Nevertheless, it was the AAE2 listeners who obtained the highest score before stops overall. Both GAE2 and AAE2 individually identified sibilance before nasals and open worst, while GAE1 listeners also identified sibilance before nasals worse than before any other context, and AAE1 identified sibilance before lateral worst. In spite of this, GAE1 listeners outperformed the rest of the groups in the perception of sibilance before approximants and vowel.

The identification of aspiration and sibilance according to the place and manner of articulation of the following sounds rendered some interesting results. Independently of certain differences seemingly due to either proficiency level or L1 dialect of the listeners, broadly speaking, the identification of aspiration tended to be particularly favorable before sounds articulated at the teeth, and remarkably poor before vowel. On the contrary, the identification of sibilance tended to be favored when followed by stops and particularly by velar sounds, while less accurate before sounds articulated at the front of the oral cavity, with the exception of the dental stop, which favored identification as much as the velar stop.

As we saw in Section 1.3.1, [h] is articulated at the back of the oral cavity (whether in the glottis, larynx, pharynx, or velum). According to the results covered in this section, it seems that the identification of aspiration is more likely when followed by a sound that is articulated at the front of the oral cavity. Likewise, [s] is articulated at the front of the oral cavity (alveoli), the identification of which seems to be more favorable, with exceptions, when followed by sounds articulated at the back of the oral cavity.

Thus, overall, L2 listeners as a group identified aspiration before interdental [ð] and sibilance before velar [k] and dental [t] most accurately than in the rest of contexts. On the other hand, they identified intervocalic aspiration [h] and sibilance before [ð̞] less accurately than in the rest of contexts.

By context, before voiceless stops, aspiration was best identified before [t] and worst before [p]. Sibilance was best identified before [t] (AAE) and [k] (GAE) and worst identified before [p], although identification percentages were rather similar. Before voiced approximants, sibilance was more accurately identified before [ɣ̞] (and also [ß̞] for GAE) while it was remarkably less accurately identified before [ð̞]. Before fricatives, aspiration was best identified before [ð] and worst identified before [x].

Upon comparison between L2 groups, the identification of sibilance before CS stops and approximants was consistently higher for GAE listeners than for AAE listeners ($p < .001$). The identification of aspiration in stops and fricatives rendered different results. GAE listeners identified fricative [v] and [ð] significantly better than AAE listeners ($p < .001$), while AAE listeners identified aspiration in stops [p^h] and [t^h] significantly more accurately than GAE listeners. Both groups of listeners identified aspiration in fricative [x] and aspirated [k^h], i.e., velar sounds, similarly ($p > .05$), but less accurately than in the rest of the contexts. In intervocalic position, GAE listeners outperformed AAE listeners in both cases: sibilance [s] and aspiration [h] ($p < .001$).

In the next section, we will see the acoustic characteristics of the sounds that have been analyzed in terms of place and manner of articulation here.

5.2 Results of Acoustic Analysis

As we stated in Section 4.6, we conducted an acoustic analysis of voiceless and fricatized stops after aspiration, as well as [h] in intervocalic position. First, we have analyzed the vowels after which the aspiration of /s/ was produced, in terms of their first three formants and their Harmonics-to-Noise Ratio (HNR), following Maniwa et al. (2009) and Boersma and Weenik (2009), respectively. Vowels ending in aspiration displayed a mean F1 at 570 Hz (*SD* = 227.70), F2 at 1929 Hz (*SD* = 324.08), and F3 at 2896 Hz (*SD* = 282.30). Their mean HNR was 11.90 dB (*SD* = 1.32), which was higher before voiced sounds (14.90 dB, *SD* = 1.51) than in intervocalic position (10.70 dB, *SD* = 1.05) and before voiceless sounds (8.80 dB, *SD*

= 1.10), suggesting higher pre-aspiration before the voiceless stops than before the fricatized approximants. Figure 38 below shows an example of aspiration from the extract *bebes agua* (you drink water):

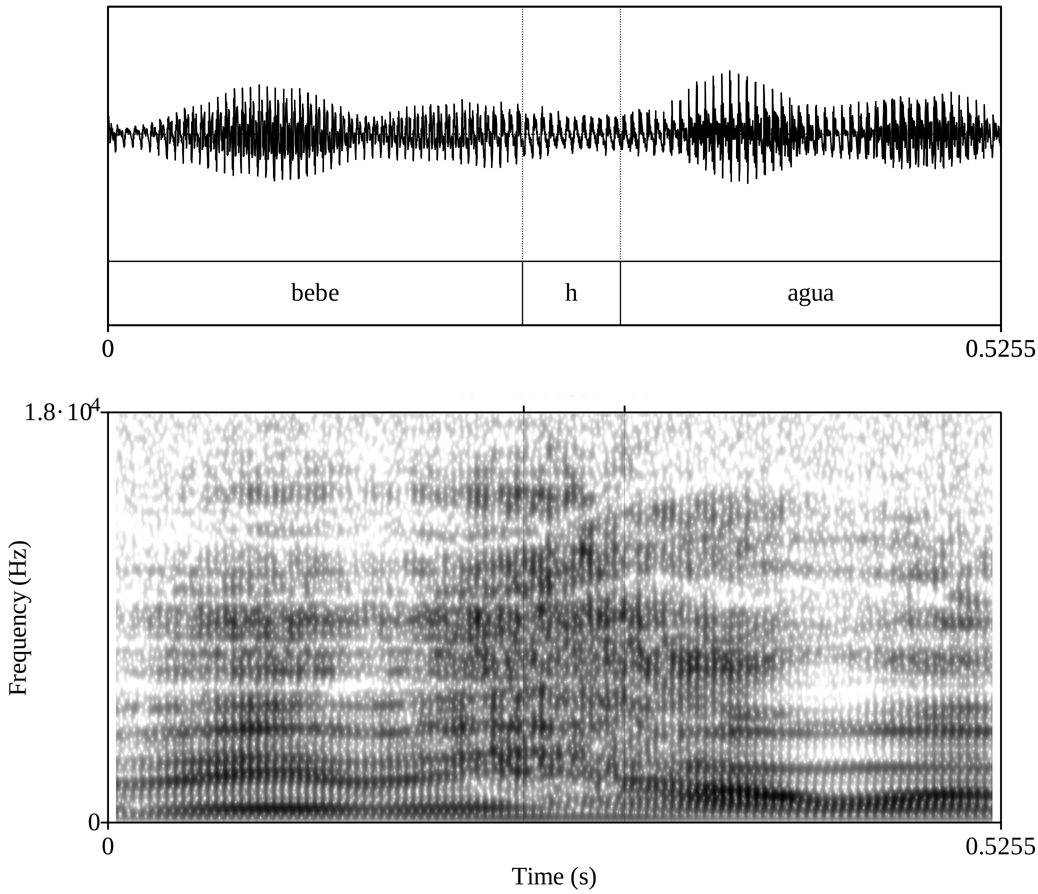


Figure 38. *Waveform and spectrogram of "bebe[h]agua*

We also analyzed the sibilant fricatives [s] and [z] before voiceless stops and voiced approximants, respectively, in an attempt to explain the results summarized in the previous paragraphs. As mentioned in Section 4.6.2.1, our reasons for analyzing our target stimuli from 500 Hz to 18 000 Hz were derived from the minimum and maximum frequency ranges observed after the analysis of CS and GAE sibilance. The minimum frequency range for the Spanish sibilants was 2298 Hz and the maximum was 14 517 Hz. For the English sibilants, the maximum was 15 735 Hz, while the minimum was 3373 Hz. In Figure 39, we can observe how WAS aspiration before [t^h] differs from CS sibilance before [t] and GAE [s] before [t]. In spite of the cepstral smoothing at 500 Hz, the WAS sound (red line) presents great energy in the first formants, whereas CS [s] (green line) and GAE [s] (blue line) present a peak at approximately the same frequency, with a second peak for the CS fricative.

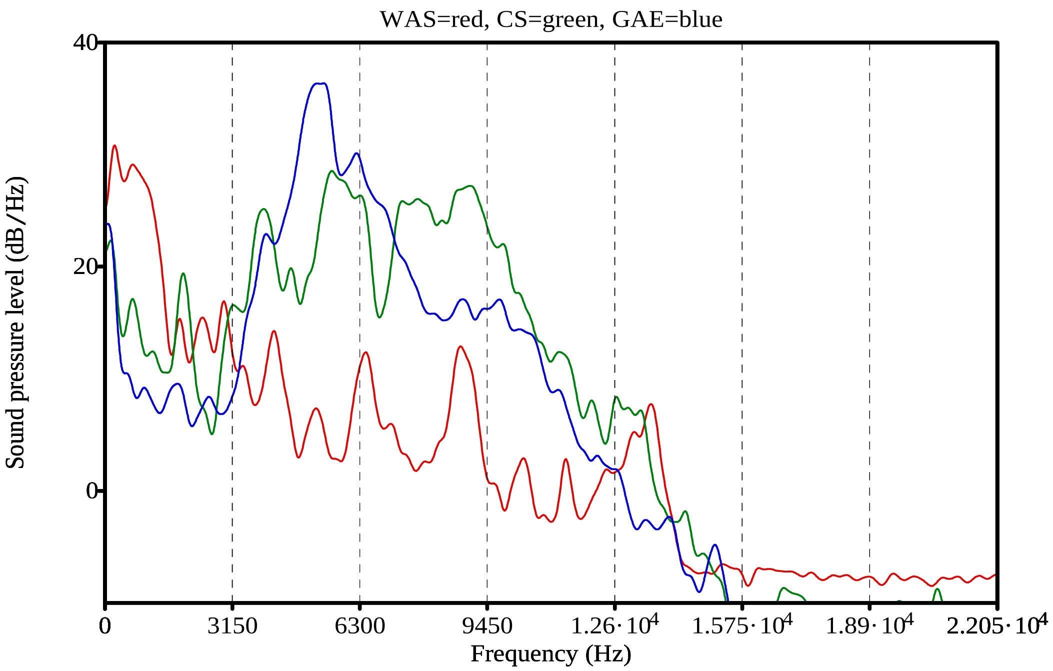


Figure 39. *Spectral slice of WAS [h], CS [s], and GAE [s] before [t]*

In this case, we also analyzed the HNR of the CS and GAE sibilants, following Boersma and Weenik (2009). CS sibilants presented a mean HNR of 1.84 dB (*SD* = 2.29), with a mean value of 6.41 dB (*SD* = 1.93) before voiced approximants, 0.43 dB (*SD* = 3.12) before voiceless stops, and -1.33 dB (*SD* = 1.82) in intervocalic position, indicating greater noise before voiceless than before voiced consonants. GAE sibilance resulted in a mean HNR of 3.67 dB (*SD* = 2.37) with very similar values in the three contexts.

5.2.1 Voiceless stops

We now report the results in terms of closure and VOT duration of stops in WAS and CS. For aspiration, VOT is understood as containing post-aspiration values. Table 13 shows that the duration of the VOT of CS [t] after sibilance is shorter than that of [p] and [k]. In WAS stops after aspiration, the shortest VOT is also for [t^h] (in Table 13, under the label *st*), followed by [k^h], and finally [p^h].

Table 13

Mean values for VOT (in ms) after aspiration and sibilance

		VOT			
		WAS		CS	
		M	*SD*	*M*	*SD*
target	sk	45	16	28	5
	sp	47	13	28	10
	st	43	9	27	5

Statistical analysis confirmed that the VOT duration of the three aspirated stops was significantly longer than that of the three unaspirated stops after sibilance (/p/: $U = 8$, $p < .05$; /t/: $U = 3$, $p < .005$; /k/: $U = 6$, $p < .01$), as we saw in Section 2.1.1. Interestingly enough, a comparison between the closure duration of stops after aspiration and sibilance (Table 14) revealed that there were no statistical differences between both sets of consonants (/p/: $U = 21$, $p = .25$; /t/: $U = 26.5$, $p = .56$; /k/: $U = 15.5$, $p = .08$), which provides further evidence of the existence of post-aspiration.

Table 14

Mean values for closure duration (in ms) after aspiration and sibilance

		closure			
		WAS		CS	
		M	*SD*	*M*	*SD*
target	sk	32	12	41	9
	sp	51	9	53	6
	st	49	14	44	13

A further analysis between the duration of closure and VOT for each stop of each L2 dialect (Figure 40) revealed that closure was significantly longer than VOT ($p < .05$) for the three CS stops after sibilance, while differences were not statistically significant for the three WAS stops after aspiration ([p^h]: Z = -.84, p = .40; [t^h]: Z = -1.68, p = .092; [k^k]: Z = -1.68, p = .092).

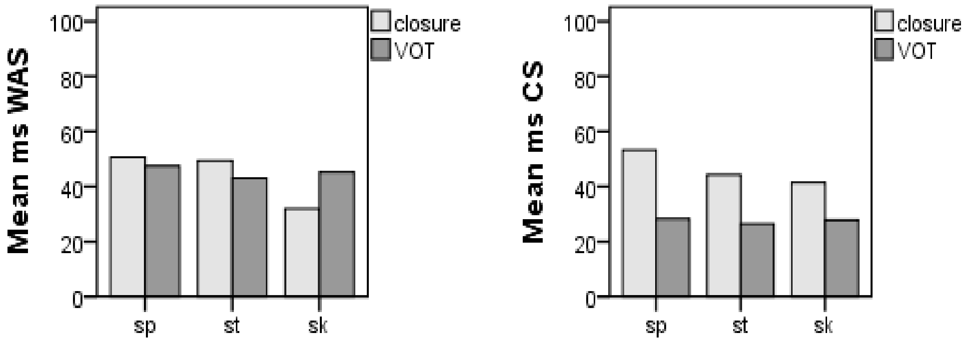
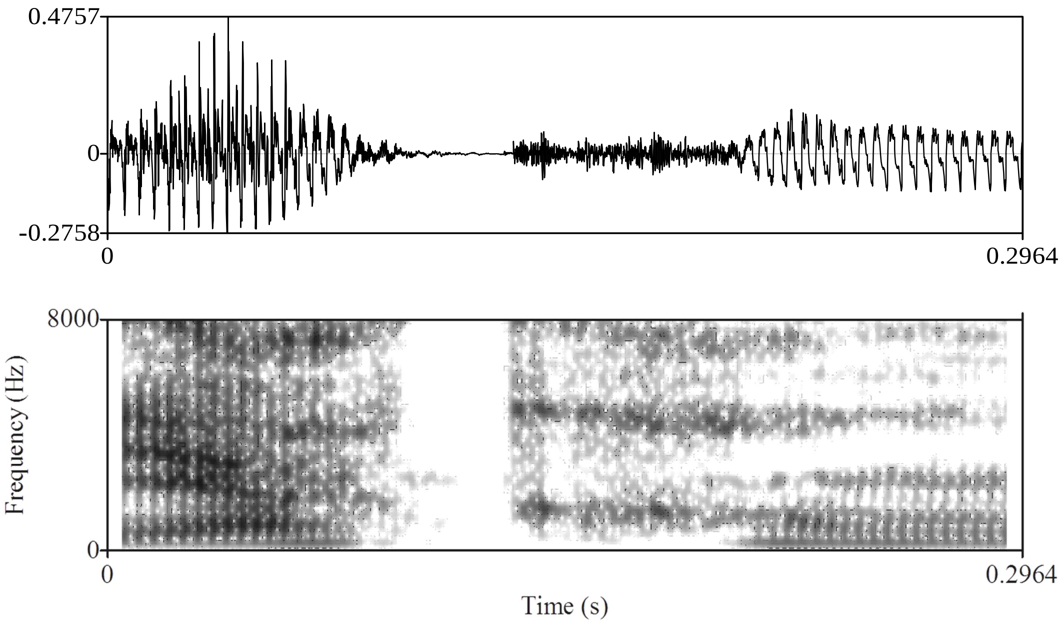


Figure 40. *Closure and VOT values of voiceless stops after aspiration and sibilance*

Additionally, Kruskal-Wallis tests revealed that the duration of VOT for the three WAS aspirated stops was not statistically different [$\chi^2(2) = -.512$, p = .77] nor were there statistical differences between the three CS unaspirated stops [$\chi^2(2) =$ $-.273$, p = .87]. Differences between the three stops in both dialects were found in terms of closure duration, with shorter closure duration for [k^h] and [k], respectively [$\chi^2(2) = 9.69$, p < .01; [$\chi^2(2) = 5.77$, p = .056[20]]., and longer for [p^h] and [p].

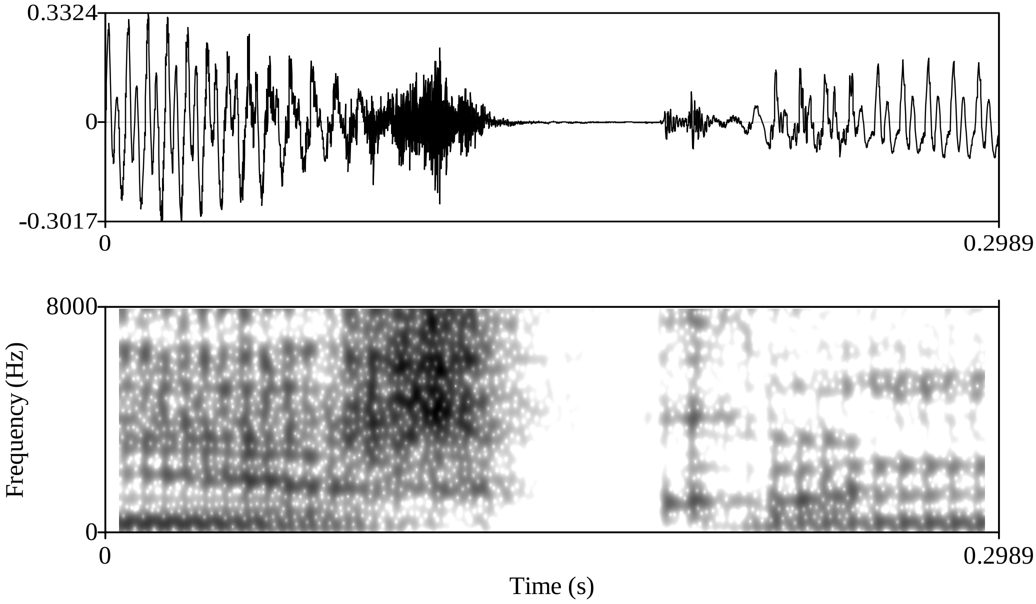
Figures 41 and 42 below show two examples of waveforms and spectrograms for /t/ after WAS aspiration and CS sibilants, respectively, from the extract *me arañas con* (you scratch me with). Notice how the VOT in WAS [k^h] is considerably longer than the VOT in CS [k]. Also, the length of the preceding vowel is longer in the first case, where the aspiration of /s/ takes place. In the second case, a sibilant [s] can be perfectly seen, with a high concentration of energy at high frequencies.

[20] On the verge of statistical significance p =.05.



VOWEL	CL	VOT	VOWEL

Figure 41. *Word-initial [k^h][21] after aspiration*



VOWEL	[s]	CLOSURE	VOT	VOWEL

Figure 42. *Word-initial [k][22] after sibilance*

[21] See Appendix G for further waveforms and spectrograms for WAS voiceless stops.

One more aspect that we analyzed was the duration of the vowels that preceded the WAS aspirated stops in 2PV and PN sentences and the vowels that preceded their unaspirated counterparts in 3PV and SN sentences (Table 15). A Kruskal-Wallis analysis showed significant differences between the duration of the vowels preceding the aspirated stops [$\chi^2(2) = 6.03$, p <. 05], which were longer for vowel before [t^h]. The same tests revealed that vowels before [t] were particularly shorter than before the other two unaspirated stops [$\chi^2(2) = 6.53$, $p < .05$].

Table 15

Vowel length with and without aspiration before voiceless stops

	vowel context					
	k		p		t	
	M	*SD*	*M*	*SD*	*M*	*SD*
yes	63	18	61	13	79	7
no	72	13	67	11	57	10

A comparison of vowels before both sets of stops revealed no significant differences for their duration before both velar sounds ($U = 20$, $p = .20$) and vowels before both bilabial sounds ($U = 19.5$, $p = .19$). However, the analysis showed significant differences between the duration of vowels before aspirated and unaspirated dental stops ($U = 3$, $p < .005$), i.e., vowels whose coda /s/ had been aspirated were significantly longer.

Thus, [p^h] presented the longest VOT and closure duration, while the preceding vowel was the shortest, [k^h] had the shortest closure duration, and [t^h] had the shortest VOT value but the preceding vowel presented the longest duration.

[22] See Appendix H for further waveforms and spectrograms for CS voiceless stops.

5.2.2 Fricatives

5.2.2.1 Intervocalic fricatives [s] and [h]

Fricatives analyzed in this study were CS [s] before voiceless stops, approximants, and in intervocalic position, WAS [h] in intervocalic position, and the fricatives which are the result of the process of fricatization that voiced stops undergo after aspiration in WAS. As we explained in Section 4.7.1, we conducted an analysis of the spectral moments for these sets of sounds.

Let us begin with the intervocalic sounds. Sibilance in intervocalic position had a mean duration of 85 ms and a mean COG of 5510 Hz. On the contrary, aspiration in intervocalic position had a mean duration of 71 ms and a mean COG of 2078 Hz. From looking at these numbers alone, we can already observe that sibilance has a much higher energy at higher frequency than aspiration. It also presents less variability than aspiration in the distribution of its energy (Table 16 below).

Table 16

Spectral moments of intervocalic [s] and [h]

	context			
	CS [s]		WAS [h]	
	M	*SD*	*M*	*SD*
intensity	66.78	2.178	72.359	4.191
duration	85.16	11.79	70.59	24.60
COG	5510.334	451.807	2078.409	910.150
dispersion	1906.343	340.036	1837.470	449.649
kurtosis	.046	.774	10.487	7.673
skewness	.292	.463	2.973	1.193

In fact, [s] presented higher COG and spectral peak (U = 0, $p < .001$) than [h], while [h] displayed higher amplitude (U = 6, $p = .005$), kurtosis (U = 1, $p < .001$), and skewness (U = 0, $p < .001$) than [s]. Nevertheless, no significant differences were found between these two sounds in terms of duration and dispersion.

5.2.2.2 CS fricatives [s] and [z]

In the context of voiceless stops, [s] before [k] presented the highest intensity and the longest duration, while the lowest dispersion values. Sibilance before [p] showed the lowest intensity, duration, and COG values, whereas the highest kurtosis and skewness. In turn, [s] before [t] presented the highest COG and dispersion values, as well as the lowest skewness and kurtosis (Table 17).

Table 17

Spectral moments of [s] before voiceless stops

	context					
	[k]		[p]		[t]	
	M	*SD*	*M*	*SD*	*M*	*SD*
intensity	68.836	1.487	66.493	3.551	67.000	2.154
duration	62.60	14.40	55.25	9.46	56.25	11.76
COG	5315.246	487.152	4959.341	532.251	5732.890	781.833
dispersion	1735.012	255.352	1908.128	496.308	1992.983	192.913
kurtosis	1.264	1.328	2.294	2.374	.366	.940
skewness	.653	.587	1.032	.716	.415	.392

In this case, [s] before [t] had significantly higher COG than before [p] ($U = 11$, $p < .05$) and higher dispersion than before [k] ($U = 13$, $p = .05$). No further differences were found as far as sibilance before the three voiceless stops is concerned.

In terms of sibilance before voiced approximants, [z] before dental [ð̞] presented the highest intensity and COG values, but also the highest dispersion. However, it had the shortest duration and the lowest kurtosis and skewness. Sibilance before bilabial [β̞] presented the longest duration and the highest kurtosis values, as well as the lowest intensity. In turn, [z] before velar [ɣ̞] had the lowest COG and the highest skewness values (Table 18).

Table 18

Spectral moments of [z] before voiced approximants

	context					
	[ß̞]		[ð̞]		[ɣ̞]	
	M	*SD*	*M*	*SD*	*M*	*SD*
intensity	66.693	3.669	69.908	2.915	69.475	3.567
duration	66.83	17.34	61.75	22.70	62.51	26.17
COG	4958.090	1059.328	5742.061	1915.487	4461.001	761.350
dispersion	1896.014	376.511	1991.298	511.699	1844.090	273.432
kurtosis	1.925	3.268	1.174	1.228	1.851	2.238
skewness	.814	.565	.443	.745	.940	.504

Upon analysis, it was found that the COG of [z] before the interdental approximant [ð̞] was significantly higher than before the velar approximant [ɣ̞] (U = 12, p < .05). However, no further significant differences were found for sibilance before the three voiced approximants.

So far, the COG values of both [s] and [z] before dental [t] and [ð̞], seem to be significantly higher than the two sibilants before [p] and [ɣ̞], respectively. A cross-analysis of [s] before voiceless stops and [z] before voiced approximants presented some statistical differences in terms of their spectral moments, according to their following sound. Sibilance before [k] and [t] had significantly higher COG than sibilance before [ɣ̞] (U = 10, p <.05; U = 8, p = .01). Additionally, sibilance before [ɣ̞] presented higher skewness than before [t] (U = 9, p < .05).

In comparison with intervocalic sibilance, [s] before the three voiceless stops was significantly shorter in duration ([p]: U = 1, p < .001; [t]: U = 3, p < .001; [k]; U = 8, p = .01). Likewise, [z] before the voiced approximants [ß̞] and [ð̞] were also shorter than intervocalic [s] (U = 10, p < .05). Furthermore, sibilance before [ð̞] had a significantly higher intensity (U = 11, p < .05) than intervocalic [s]. Finally, even when [s] before [ɣ̞] and intervocalic [s] presented no differences in duration, the COG value of intervocalic [s] was significantly higher than the COG of sibilance before [ɣ̞] (U = 7, p < .01).

Summing up, sibilance before the three voiceless stops and before the approximants [β̞] and [ð̞] were significantly shorter than intervocalic [s], but it was not the case for sibilance before [ɣ̞]. Nevertheless, the COG values rendered more varied results. Sibilance before [ð̞] and [t] showed the highest values while before [p] and [ɣ̞] COG was the lowest. In fact, the COG for sibilance before [ɣ̞] was also significantly lower than the COG for intervocalic [s].

5.2.2.3 WAS fricatives [v], [ð], and [x]

WAS interdental fricative [ð] presented the highest intensity, COG, and dispersion values, while the lowest kurtosis and skewness. In turn, labiodental [v] had the lowest intensity and duration values. Velar [x] presented the longest duration and the highest kurtosis and skewness values, while it had the lowest COG and dispersion values (Table 19).

Table 19

Spectral moments of WAS fricatized sounds

	context					
	[v]		[ð]		[x]	
	M	*SD*	*M*	*SD*	*M*	*SD*
intensity	68.654	7.303	70.582	7.382	69.667	7.665
duration	61.72	13.81	63.52	17.40	65.57	22.92
COG	2116.212	687.418	3747.926	2155.501	2068.464	952.661
dispersion	2056.158	938.815	2581.873	794.520	1817.923	778.588
kurtosis	21.123	24.047	5.840	9.249	44.342	80.933
skewness	3.592	2.039	1.769	1.651	4.449	3.806

In our analysis of WAS fricatized sounds, only kurtosis presented differences on the verge of significance between the three contexts [$\chi^2(2) = 5.81$, p = .055].

Specifically, the kurtosis of [v] was significantly higher than that of [ð] ($U = 11$, $p < .05$).

A further comparison between intervocalic aspiration and the fricatized sounds revealed that [ð] had a significantly higher COG ($U = 12$, $p < .05$) and dispersion ($U = 8$, $p < .01$) than intervocalic [h], while [h] presented higher kurtosis ($U = 9$, $p < .05$) and skewness ($U = 12$, $p < .05$) than [ð]. No differences were found between [v] or [x] and the intervocalic aspiration.

As it was the case with WAS vowels before aspirated and unaspirated stops, we also decided to analyze the duration of vowels before the fricatives in 2PV and PN sentences, and the vowels before the voiced approximants in 3PV and SN sentences (Table 20). The analysis revealed no significant differences in the duration of vowels with and without aspiration in this case (U = 502, p .89). Even when vowels before the fricative [ð] (aspiration) are longer than vowels before the approximant [ð̞] (no aspiration), the difference was not statistically significant ($Z = -1.42$, $p = .16$).

Table 20

Vowel length with and without aspiration before WAS fricatized sounds

	vowel context							
	[v]		[ð]		[x]		intervocalic	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
yes	75	18	85	29	79	19	57	15
no	79	16	64	12	82	24	74	23

To sum up, the three fricatized sounds presented similar values for intensity, duration, and the spectral moments measured. However, one difference that was found was in terms of kurtosis ([v] higher than [ð]). In comparison with intervocalic [h], [ð] presented higher dispersion and COG values than [h], while [h] displayed higher kurtosis and skewness than [ð]. Both [v] and [x] had similar values to [h].

A further cross-analysis between [s] before voiceless stops, [z] before voiced approximants, and aspiration in fricatized sounds revealed significant higher COG values for sibilance in both CS contexts [$\chi^2(2) = 29.54, p < .001$] than for the WAS fricatized sounds, while significantly higher kurtosis [$\chi^2(2) = 17.01, p < .01$] and skewness [$\chi^2(2) = 27.02, p < .001$] for the fricatized sounds. Thus, [s] and [z] present a higher mean energy than non-sibilant fricatives, while these have a higher peakedness and amount of energy concentrated at lower frequencies, as reported in Section 2.1.2 (Barreiro Bilbao, 1994; Jongman et al., 2000).

5.2.2.4 Place of articulation

In this section, we analyze the extent to which the different acoustic characteristics measured can distinguish place of articulation, with the addition of spectral peak measurements. For this reason, we also took voicing into account to separate voiceless [s] from voiced [z], although both share the same place of articulation (alveolar). Table 21 below displays the values for each variable according to place of articulation.

Alveolar [s] has the highest spectral peak and COG values, while the lowest amplitude, kurtosis, and skewness. Thus, this sound presents a great concentration of energy in the higher areas of the spectrum. On the contrary, [h] presents the highest amplitude and duration, whereas it shows the lowest spectral peak, COG, and dispersion values. In this case, this sound concentrates its energy in the lower area of the spectrum. Additionally, labiodental [v] displays the shortest duration, while interdental [ð] presents the highest dispersion of energy, and the velar [x] shows the highest kurtosis and skewness, i.e., a great concentration of energy in the lower part of the spectrum but high peakedness.

Table 21

Spectral peaks and spectral moments by place of articulation and voice

	place of articulation and voice					
	alveolar [s] *M*	alveolar [z] *M*	labiodental [v] *M*	interdental [ð] *M*	velar [x] *M*	glottal [h] *M*
intensity	67.727	3.310	68.654	7.303	70.582	7.382
duration	65.90	20.41	61.72	13.81	63.52	17.40
peak	4444.69	1531.19	984.06	422.51	2774.78	2192.91
COG	5319.281	1133.423	2116.212	687.418	3747.93	2155.501
dispersion	1903.563	394.720	2056.158	938.815	2581.87	794.520
kurtosis	1.396	1.879	21.123	24.047	5.840	9.249
skewness	.626	.652	3.592	2.039	1.769	1.651

A cross-analysis of the values of these characteristics revealed significant differences in COG, kurtosis, and skewness ($p < .001$), but none in terms of intensity, and duration. Nevertheless, analyses by pairs of sounds showed no significant differences between the alveolar [s] and [z], between the velar/glottal [x] and [h], between the interdental [ð] and [x], and between the labiodental [v] and neither [x] nor [h]. Both sibilants [s] and [z] displayed higher COG and spectral peaks than [v], [x], and [h] ($p < .001$), while [v], [x] and [h] presented higher kurtosis and skewness than the sibilants ($p < .001$). Additionally, the sibilants also had higher amplitude than [h] ($p = .001$, $p < .05$). The interdental [ð] was distinguished from the labiodental [v] by lower kurtosis ($U = 11$, $p < .05$). It was differentiated from [s] and [z] by higher dispersion than both sibilants ($U = 69$, $p < .05$; $U = 46$, $p < .05$), but also lower COG than [s] ($U = 63$, $p < .05$). In comparison with [h], interdental [ð] presented higher dispersion ($U = 8$, $p <.05$) and lower kurtosis ($U = 9$, $p <.05$) and skewness ($U = 12$, $p = .01$).

Thus, in this study, COG and spectral peak served to differentiate alveolar sibilants from labiodental, velar and glottal non-sibilant fricatives. In the case of the interdental non-sibilant, only COG distinguished it from the voiceless fricative, while kurtosis differentiated it from both sibilants. Kurtosis, skewness and dispersion were useful to distinguish interdental [ð] from [h], and from [v] only in terms of kurtosis, whereas no parameter differentiated it from [x]. The labiodental, velar, and glottal fricatives displayed no significant differences among themselves for any of the cues measured.

5.2.3 Speaker Gender

As we saw in Section 2.1.2, speaker gender may (Horn, 2013; Ruch, 2012) or may not (Yao, 2007) have an effect on the characteristics of the stimuli. Therefore, we now review our stimuli according to the gender of the speakers of WAS and CS.

5.2.3.1 WAS voiceless stops [p^h], [t^h], and [k^h]

In Table 22, we can see the duration of closure and VOT of each of the three voiceless stops after aspiration, according to our WAS female and male speakers.

Table 22

Closure and VOT values of WAS aspirated stops by speaker gender

		[k^h]	[p^h]	[t^h]
		M	*M*	*M*
WASF1	closure	31	48	56
	VOT	62	50	47
WASM2	closure	33	53	43
	VOT	36	50	33

In this case, there were no significant differences between both speakers in terms of closure duration for any of the three aspirated stops. However, differences were found for the VOT of [k^h], which was significantly longer for WASF1 (female speaker) than for WASM2 (male speaker) ($U = 0$, $p < .05$). Differences between both speakers concerning the VOT of [t^h] were on the verge of significance (U = 1, p = .057); once more, longer for the female speaker.

In any case, L2 listeners' identification of [p^h], [t^h], [k^h] was higher in the stimuli from the female speaker than from the male speaker (WASF1: 12.67%, 22.31%, 11.75; WASM2: 9.28%, 12.80%, 11.64%).

5.2.3.2 WAS fricatives [v], [ð], and [x]

Table 23 shows a summary of the mean spectral values for the WAS fricatized sounds [v], [ð], and [x] according to the gender of the speaker.

Table 23

Acoustic characteristics of WAS fricatives by speaker gender

		[v]	[ð]	[x]	[h]
		M	*M*	*M*	*M*
WASF1	intensity	62.807	63.888	62.857	69.425
	duration	63.33	74.12	63.88	83.25
	peak	666.48	4561.46	1250.62	806.18
	COG	2706.321	5594.638	2552.352	1936.658
	dispersion	2907.65	3159.652	2301.045	1801.728
	kurtosis	4.046	.158	15.216	37.678
	skewness	1.968	.485	2.922	4.351
WASM2	intensity	74.502	77.275	76.477	75.292
	duration	60.10	52.93	67.26	57.93
	peak	1301.64	988.10	1086.88	1061.53
	COG	1526.102	1901.215	1584.576	1393.204
	dispersion	1204.665	2004.094	1334.801	1409.639
	kurtosis	38.2	11.522	73.469	23.796
	skewness	5.216	3.053	5.976	4.499

This time, statistical analysis revealed a few differences in terms of speaker gender. WASF1 realized [v] and [ð] with higher COG and dispersion than WASM2, while WASM2 realized these two fricatives with higher kurtosis and skewness (U = 0, p < .05). Additionally, WASF1 presented higher spectral peak for [ð] than the male speaker, whereas he displayed higher spectral peak for [v] than the female speakers.

Additionally, the only difference between both speakers for [x] was in terms of intensity (U = 0, $p < .05$), which was higher for the male speaker. No differences were found for intervocalic [h] between the two speakers.

L2 listeners presented a higher identification of [ð] and [x] in the stimuli from the female speaker than in those from the male speaker (WASF1: 39.51%, 11.74%; WASM2: 17.43%, 7.40%), while the opposite was true for [v] (WASF1: 17.02%; WASM2: 18.19%). Their identification of intervocalic [h] was the same in both cases (4.59%).

5.2.3.3 CS fricatives [s] and [z]

In this section, we show the values of the spectral moments of CS [s] (Table 24) and CS [z] (Table 25) in terms of the gender of our CS speakers.

Table 24

Acoustic characteristics of CS [s] by speaker gender

			[s]			
			[k]	[p]	[t]	[s]
			M	*M*	*M*	*M*
speaker	CSF1	intensity	69.623	66.984	68.449	66.867
		duration	57.59	57.21	59.56	83.82
		peak	4815.22	4619.37	4869.34	5453.90
		COG	5498.407	4904.591	6197.158	5174.064
		dispersion	1907.870	2238.332	2061.313	2163.914
		kurtosis	.990	1.816	-.282	.275
		skewness	.716	1.376	.203	.693
	CSM2	intensity	68.049	66.002	65.550	66,693
		duration	67.61	53.28	52.94	86,.1
		peak	4586.02	3572.89	4587.88	4196.76
		COG	5132.084	5014.091	5268.622	5623.277
		dispersion	1562.154	1577.925	1924.654	1692.992
		kurtosis	1.539	2.772	1.014	.703
		skewness	.590	.687	.626	.114

In the case of [s] before the three voiceless stops and in intervocalic position, no significant differences were found between the two CS speakers. L2 listeners identified [s] before [p] and [k] slightly better in the stimuli from CSM2 (male speaker) than those of the CSF1 (female speaker) (CSM2: 90.41%, 94.23%; CSF1: 86.95%, 92.46%). The opposite was shown for [s] before [t] (CSM2: 90.91%; CSF1: 92.10%).

Table 25

Acoustic characteristics of CS [z] by speaker gender

			[z]		
			[ß̞]	[ð̞]	[ɣ̞]
			M	*M*	*M*
speaker	CSF	intensity	67.928	71.252	71.677
		duration	55.29	50.37	58.06
		peak	3653.95	4884.33	3007.69
		COG	4536.458	6026.078	4045.290
		dispersion	2020.904	2258.552	2061.780
		kurtosis	3.496	1.552	2.245
		skewness	1.233	.189	1.141
	CSM	intensity	65.458	68.565	67.273
		duration	78.36	73.13	66.96
		peak	4109.70	5954.63	3249.48
		COG	5379.723	5458.044	4876.713
		dispersion	1771.123	1724.044	1626.399
		kurtosis	.354	.795	1.458
		skewness	.396	.698	.739

As to the sibilant [z] before the three voiced fricatives, the only differences between the two speakers were found before the bilabial [ß̞] and before the velar [ɣ̞]. In the case of the bilabial sound, CSM2 displayed higher duration, whereas CSF1 presented higher skewness ($U = 0$, $p < .05$). For the velar sound, the male speaker showed higher dispersion than the female speaker ($U = 0$, $p < .05$). No differences were found between both speakers in their realizations of [z] before the dental approximant [ð̞].

L2 listeners presented an identification of [z] before the bilabial, the dental and the velar approximants consistently higher in the stimuli from the male speaker than in those of the female speaker (CSM2: 95.46%, 86.82%, 95.54%; CSF1: 85.38%, 79.76%, 85.12%).

5.3 Summary

In this chapter we have covered the results for the various elements analyzed in this study. From the very pilot study, we have seen that the identification of aspiration as a dialect variant of Spanish /s/ is problematic for L2 learners in the first stages of language learning. The addition of noise seems not to affect identification, except for those learners in the high-intermediate developing stage of learning. It seems that elementary learners do not have enough knowledge of the L2 to be able to identify this feature, while proficiency learners know enough to identify such variation.

Our current study replicates these findings with respect to L2 learners in the first two levels of Spanish. Overall, aspiration was again significantly less identified than sibilance. However, as the pilot test already pointed at, AAE listeners' identification of sibilance was significantly less accurate than that of GAE listeners. Additionally, we measured the reaction times of the listeners, which were longer for all WAS sentences, and the effect of the years of instruction received. In this case, we have seen significant results in elementary learners with less than one year of instruction.

In our current study, we also saw that the syntactic context can influence identification; namely, aspiration was best identified in second-person verbs, whereas sibilance identification was better in plural nouns. The phonetic context also affects the identification in terms of place and manner of articulation of the target features. Aspiration was best identified in the interdental fricative and the dental stop, while intervocalic aspiration was significantly the least identified sound. Sibilance, in turn, displayed higher scores for several places and manners of articulation; however, before stops and velar sounds it received the highest identification accuracy, while sibilance before the dental approximant was the least accurately identified. One more aspect worth mentioning is that AAE listeners generally identified aspiration and sibilance before voiceless stops significantly better than GAE listeners.

Finally, we have carried out an acoustic analysis of the most significant target stimuli, i.e., stops and fricatives, to find the extent to which the characteristics of the stimuli can have played a role in these identification patterns, which will be discussed in the next chapter. In terms of speaker gender, some differences in the acoustic characteristics of the stimuli have appeared, although L2 listeners' overall identification of the target stimuli seem to be more favorable for the WAS female speaker and the CS male speaker, regardless of this differences.

In the following chapter, we discuss the findings observed in this current chapter, as well as the implications for L2 speech perception and directions for future research.

CHAPTER 6

DISCUSSION AND CONCLUSIONS

The objective of this dissertation was to investigate the perception of a dialectal phonetic variant of the Spanish morphological marker *–s*, as opposed to the mainstream variant of this marker. L2 learners in this study were also native speakers of two L1 dialects of American English, which makes this not only a cross-language but also a cross-dialect study on speech perception. To our knowledge, this is the first study that comprises two dialects of an L2 and two dialects of an L1 in a cross-language experiment. Participants completed an identification task in which they determined whether nouns and verbs from both L2 dialects embedded in sentences were second-person verbs and plural nouns (ending in *–s*) or third-person verbs and singular nouns (ending in vowel). This task was preceded by a Language Background Questionnaire and a Spoken English Questionnaire that allowed the classification of the participants and the establishment of correlations of influential factors with the results. The results obtained in the previous chapter are discussed below, in terms of the overall perception of aspiration and sibilance by dialect group, perception according to proficiency level and amount of L2 instruction, and perception related to the syntactic and the phonetic contexts in which the target stimuli were found. Subsequently, this is followed by a discussion of the acoustic characteristics of the stimuli analyzed, the implications for L2 speech perception, the limitations of this study, and the suggestions for future research.

6.1 Perception of Aspiration and Sibilance

Our first research question inquired whether the L1 dialect of the listeners in this study would affect the identification of WAS aspirated /s/.

Overall, the 206 participants in this study presented a low identification accuracy of aspiration (14%), generally identifying WAS plural nouns and second-person verbs as singular nouns and third-person verbs. These results are in agreement with those found by Schmidt (2011), whose study revealed that participants at elementary level identified aspiration in syllable-final, word-internal position in 6.4% of the cases, and intermediate level participants identified aspiration in 5.5% of the cases.

As we stated in Section 3.3.1, our hypothesis was that listeners would be unable to extract enough information from the stimuli to relate it as a possible realization of /s/. A similar sound to WAS [h] occurs in English as a contrastive sound in initial position but not as a legitimate variant of /s/ in implosive position, as is the case in WAS (and other varieties of Spanish). Even when aspirated /s/ and English [h] are acoustically and articulatorily similar to each other, as described in Section 1.3.1, listeners may not assimilate these two sounds, precisely due to the phonotactic biases of their L1: this sound is never found in syllable-final position nor is it ever a legitimate allophone of implosive /s/.

Nevertheless, we cannot say that this L2 variant was not perceived at all by these listeners; in fact, their scores were higher than those reported by Schmidt (2011) above. On the one hand, elementary and intermediate level GAE listeners in our study identified aspiration in 8.41% and 16.67% of the cases, respectively. On the other hand, AAE listeners of the two levels identified aspiration correctly in 13% and 20.3% of the cases, respectively. The fact that the lowest score for all groups was for aspiration before vowel, where aspirated /s/ is clearly uttered as full [h], indicates that this L2 variant was either understood as part of the preceding vowel or regarded as noise, but not broadly identified as a possible realization of underlying sibilance.

The second research question that we posited was whether the identification of sibilance would vary according to the L1 dialect of the groups under study.

We hypothesized that, as syllable-final /s/ is a legitimate sound in both GAE and AAE, listeners would assimilate CS [s] to English [s]. In this sense, overall identification of sibilance was above 80%; listeners were generally able to identify CS second-person verbs and plural nouns as such. This again corroborates the findings by Schmidt (2011), in which listeners of elementary and intermediate Spanish obtained accuracy percentages for sibilance of 98.9% and 99.3%, respectively.

Nevertheless, when we look at the scores in relation to the L1 dialect of the listeners, we see that it was the GAE listeners who closely resembled the perception accuracy in Schmidt's study, with 92.92% and 94.61% identification for elementary and intermediate listeners, respectively. This was not the case for AAE listeners of elementary and intermediate levels in our current study, whose scores were 82.36% and 89.94%, respectively. As also stated in Section 3.3.1, AAE speakers can regularly omit final /s/ from plural nouns and third-person verbs, and as pointed out in Johnson's (2005) and de Villiers and Johnson's (2007) studies, AAE children seemed not to understand /s/ as an agreement marker, whereas GAE children did.

Can this fact be extrapolated to our results? While we cannot state that AAE listeners did not understand CS [s] as an agreement marker, we can say that the characteristics of their L1 dialect may have played a role in their perception of sibilance, particularly for AAE1 listeners, the only group whose identification of sibilance was significantly less accurate than their perception of sentences without the morphological marker *–s*. In fact, this group presented a higher identification of WAS and CS target singular nouns and third-person verbs as the test progressed, implying that they initially identified these tokens as having the morphological marker *–s* and they began to identify these stimuli more accurately as more of these tokens appeared in the task.

6.1.1 The Effect of L2 Proficiency and Instruction

We have also seen how increased experience with a second language can generally favor perception (Bohn & Flege, 1990; Flege & Liu, 2001; Flege, Takagi, & Mann, 1996). In Schmidt's (2011) study, the perception of aspiration for GAE listeners was more accurate as level of proficiency increased, although only the level 5 (proficiency level) listeners' identification accuracy was statistically similar to that of the native speakers of aspirating Spanish dialects in her study.

For our current study, taking L1 dialect and proficiency level into account, we observed two clear effects in the perception of aspiration:

i) Level of proficiency was an influential factor for AAE listeners' perception of aspiration but not for GAE listeners, who performed similarly despite proficiency.

ii) L1 dialect and proficiency were generally favorable to AAE2 listeners in the perception of aspiration, while the rest of the groups performed similarly despite L1 dialect or proficiency level.

In this case, the performance of GAE listeners corroborated that of the participants in Schmidt's (2011) work at elementary and intermediate levels, whose performance was similar despite level of proficiency. However, for AAE listeners, a higher level of proficiency played a role in the identification of the L2 variant by rendering higher accuracy of intermediate-level listeners.

In the perception of sibilance, the findings in Schmidt's (2011) research work revealed that GAE listeners at all levels of proficiency identified sibilance in a similar manner to native speakers of Spanish. In our current study, it was only at intermediate level of Spanish that AAE listeners performed similarly to GAE listeners of elementary Spanish. Two main effects were also observed:

i) Level of proficiency was again an influential factor for AAE listeners' perception of sibilance but not for GAE listeners, who performed on the verge of similarity despite proficiency.

ii) L1 dialect, GAE in this case, positively influenced perception over proficiency level. Both groups of GAE performed significantly better than their AAE counterparts.

These results corroborate the findings in Schmidt's (2011) work. GAE listeners' perception of sibilance was comparable to that by native speakers of the L2, achieving top performance despite level of proficiency. For AAE listeners, experience with the second language also rendered higher accuracy; it seems that AAE listeners initially have a disadvantage over GAE listeners for the perception of sibilance, which is apparently overcome with increased experience with the L2.

Thus, on the one hand, level of proficiency in this case was irrelevant for GAE listeners; both elementary and intermediate groups performed similarly for aspiration and sibilance. On the other hand, level of proficiency played a role in perception for AAE listeners; intermediate level participants performed significantly better than elementary level participants for both conditions.

Taking years of instruction into account, i.e., experience with the L2, an increase in the amount of instruction was directly correlated to the perception of sibilance for both elementary groups and to the perception of aspiration for GAE1 listeners. Remarkably, there was an inverse correlation between the amount of instruction and the perception of aspiration for AAE1 listeners. For intermediate listeners of both L1 dialects, an increase in the amount of instruction was not so clearly correlated to perception, although generally listeners with 3-5 years of instruction obtained the best results for both L2 variants.

Overall, there were no significant differences in the perception of aspiration and sibilance between both GAE groups and AAE2 listeners. The most significant effects were found at the elementary level: AAE1 listeners with less than 1 year of instruction outperformed GAE1 listeners in the perception of aspiration, and closely resembled the perception of AAE2 listeners with 1-3 years of instruction. On the contrary, GAE1 listeners outperformed AAE1 listeners in the perception of sibilance with less than one year of instruction and with 1-3 years of instruction.

Therefore, we believe that the perception of aspiration and sibilance by what we can consider truly naïve listeners in this study (elementary level students with less than 1 year of instruction) reflects the effect of L1 dialect on how L2 listeners perceived the two L2 dialect variants. “The relative ease or difficulty of a given contrast varies according to the listener’s native language” (Best & Tyler, 2007, p. 16) and, particularly, to their native dialect.

6.1.2 The Effect of Syntactic Context

The syntactic context in which the target words were embedded, i.e. nouns in carrier sentences and verbs in content sentences yielded quite homogeneous results. Sibilance was better identified in plural nouns than in second-person verbs while it was the opposite for aspiration, with the exception of AAE2 listeners, who identified both sibilance and aspiration in plural nouns best. In fact, they outperformed the rest of the groups (21%) in the identification of aspiration in nouns.

Likewise, singular nouns ending in vowel where better identified than third-person verbs in CS sentences by all groups of listeners. Therefore, whether sibilance was present or not, all groups of participants identified singular nouns more accurately than third-person verbs in CS sentences. In the case of WAS sentences, both elementary-level participants identified singular nouns ending in vowel better than third-person verbs, while both intermediate-level listeners identified third-person verbs more accurately than singular nouns.

In relation to this matter, Yeni-Komshian, Robbins, and Flege (2001) state that “lexical processing in adults and L1 vocabulary acquisition in children are affected by word class distinctions. The question is whether there are word class effects in L2 learning” (p. 285). They studied how Korean-English bilinguals who immigrated to the United States from Korea judged the grammaticality of sentences with plural nouns and third-person verbs in grammatical and ungrammatical English sentences. Ungrammatical sentences consisted of target words with elimination of the morphological marker *–s* when it should be present, and presence of the morphological marker when it should be absent. They found that listeners were more accurate at identifying ungrammaticality in verbs than in nouns, as a consequence of the fact that Korean is a subject-object-verb language; thus, mothers emphasize verbs more than nouns and so children acquire verbs before they acquire nouns. For English, which is a subject-verb-object language, this situation is the opposite. More emphasis is placed on nouns than on verbs. For this reason, they argue that L2 learners "will approach their L2 learning task with a linguistic mental set established from the structure of their L1, and that the hierarchies in their lexicon will influence learning in pronunciation and morphosyntax" (p. 294).

In our study, listeners have either GAE or AAE as their L1 dialects, which come from the same language, English. If we attend to what the authors claim, they would have to identify nouns more accurately than verbs. As the morphological marker *–s* also exists in English (regardless of its uses), it seems this is the case with CS sentences in which sibilance is present; all groups identified sentences with plural nouns more accurately than sentences with second-person verbs.

However, in the case of aspiration, we saw a general trend towards the opposite; three out of the four groups were more accurate at identifying second-person verbs than nouns, which means they tended to hear plural nouns as singular nouns. If we consider that [h] is not a possible realization for final /s/ in English and that the listeners were not exposed to aspirating varieties of Spanish, it is logical that they would regard aspiration as an absence of sibilance. Thus, understanding plural nouns as singular nouns would mean they focused more on nouns. Nevertheless, one group, i.e. AAE2, very significantly identified aspiration better in plural nouns, above the rest of the groups, in comparison with their identification of aspiration in second-person verbs. With the data that we currently have, we cannot provide further explanations for this exception.

6.1.3 The Effect of Phonetic Context

Phonetic context rendered some findings worth mentioning. In the case of aspiration, GAE1 listeners perceived all places and manners of articulation significantly less accurately than the rest of the groups, except for the glottal and open contexts, whose lowest scores they shared with AAE1 listeners. In the case of sibilance, it was the AAE1 listeners who significantly identified all places and manners of articulation less accurately than the rest of participants, except for dental sounds, which all groups perceived similarly. So far, these findings further corroborate the effect of L1 dialect in the perception of aspiration and sibilance at elementary level.

For aspiration, proficiency level seemed to determine the highest accuracy in identification across groups. Both intermediate groups obtained the highest scores for places and manners of articulation, with the exception of bilabial sounds and particularly voiceless stops, for which AAE1 listeners showed the highest accuracy across groups. Interestingly enough, AAE1 listeners obtained the lowest scores for the perception of voiceless stops in sibilance. The perception of full [h] before vowel seemed to be favored by level of proficiency, i.e., intermediate level listeners.

For sibilance, intermediate level listeners also obtained the highest scores for most of the places and manners of articulation, with the exception of the glottal open sound and approximants, which GAE1 listeners identified more accurately than the rest of the groups. Also remarkable, GAE1 listeners obtained the lowest scores for fricatives after aspiration. The perception of [s] before vowel was favored by L1 dialect of the listeners, i.e., GAE.

Our first discussion is devoted to voiceless stops after aspiration and sibilance, for which AAE1 listeners obtained the highest and lowest scores, respectively, in relation to the rest of the groups. In English, these voiceless stops are aspirated in initial position of stressed syllables, whether preceded by a word ending in vowel or /s/, but unaspirated in unstressed syllables and in clusters where the first element is /s/. In this case, aspiration does not mark a morphological contrast in English; it differentiates two allophones of the same category.

In WAS, aspiration in voiceless stops marks the contrast between plural and singular nouns, and second-person and third-person verbs. In word-initial position and preceded by vowel, /p, t, k/ are unaspirated. When preceded by aspirated /s/, these sounds are also aspirated [p^h, t^h, k^h]. As Torreira (2007b) points out, “Western Andalusian voiceless stops preceded by aspirated-s have a longer VOT (postaspiration), shorter preaspiration and longer stop closure” (p.119) than other Spanish aspirating dialects, such as Porteño or Puerto Rican Spanish. In CS, voiceless stops are unaspirated, whether preceded by vowel or /s/.

In this case, AAE1 listeners seemed to be more sensitive to the role of aspiration in voiceless stops, in contrast with the absence of aspiration after [V] in the context of WAS sentences, than the rest of the groups, and less sensitive to sibilance before unaspirated voiceless stops. The closest explanation we could find for their performance in aspirated voiceless stops, although remote, was the findings by Sligh and Conners (2003). They claimed that, at least concerning the perception of initial and final stops, for AAE speakers “word-initial sounds are more important or salient than word-final sounds, compared to SAE” (p. 222). While we cannot assert that this is the reason for our results, some studies have shown that L2 learners may rely on cues that are not present in their L1 (Cebrian, 2006).

A further explanation could be found in the stopping of interdental fricatives that is characteristic of this dialect for sounds in initial position, i. e., /θ, ð/ realized as [t, d], which gives more saliency to stop consonants. However, the stopping of /ð/ is a characteristic that is spreading to GAE, as reported by and Smith (2009) and Zhao (2010), and its confusability with /θ/ can also be found (Smith, 2013b). One more aspect of AAE also concerns the devoicing of voiced stops in final-position. All in all, it seems that AAE speakers tend towards a higher use of voiceless stops when GAE speakers would not. Perhaps this fact makes these listeners more sensitive to these types of sounds.

Our second discussion concerns fricatives and approximants, derived from aspiration and sibilance before /b, d, g/, for which GAE1 listeners obtained the lowest and the highest scores, respectively, in relation to the rest of the groups. Approximant allophones of these sounds are given in initial position after vowel, in both WAS and CS, and after /s/ in CS. After aspiration in WAS, they become fricative allophones. According to Romero Gallego (1995), the main difference between these fricative and approximant allophones is their duration, not their degree of constriction. Fricatives are longer than approximants, regardless of place of articulation.

In English, voiced stops in initial position remain stops whether preceded by a word ending in vowel or /s/. In this case, GAE1 listeners seemed not to understand duration as an indication of aspiration in the fricatives. The way our experiment was designed, we cannot be certain whether they perceived this difference in duration, only that they did not link this contrast to the presence of /s/.

Finally, our third discussion focuses on the perception of full [h] and [s] before vowel contexts. In these cases, we see that the perception aspiration is closely related to the level of proficiency of the listeners, while the perception of sibilance, the mainstream L2 variant, is correlated to the L1 dialect of the participants (GAE), once more providing evidence of the influence of L1 dialect characteristics in the perception of an L2.

6.1.4 The Effect of the Acoustic Characteristics of the Stimuli

In an attempt to account for the results reported for place and manner of articulation, we carried out an acoustic analysis of the target stimuli according to the most salient findings.

In the first case, our analysis focused on aspiration before voiceless stops. The results of our acoustic analysis corroborate those in the studies reported in Section 2.1.1, in the sense that VOT was significantly longer in WAS aspirated stops than in CS unaspirated stops. Additionally, we have demonstrated that the preceding vowels are also longer, taking into account that we considered any possible pre-aspiration or breathy voice as part of the vowel. As stated by Torreira, 2007a, the vowel is actually shorter if measured separately from pre-aspiration. This effect is more acute in word-final vowels than in word-internal vowels and in unstressed vowels than in stressed vowels (Marrero, 1990). In fact, our results also point at a negative correlation between the duration of pre-aspiration and post-aspiration. The bilabial stop showed the longest VOT and the shortest duration of the preceding vowel, while the dental stop presented the shortest VOT and the longest duration of the preceding vowel.

In general, L2 listeners in our study identified aspiration in bilabial [p^h] less accurately than in the other two stops, whereas their identification of dental [t^h] was the most salient. From our analysis, we found that the VOT of aspirated stops was significantly longer than the VOT of the unaspirated stops, while their closure duration was similar (as opposed to the lengthened closures in Torreira's study, 2007a). Although [p^h] had the longest VOT (47 ms) and [t^h] had the shortest (43 ms), there were no statistical differences between the three stops. Closure duration, in turn, was again the longest for [p^h] (51 ms), but the shortest for [k^h] (32 ms). This is where we found statistically significant differences. It seems that WAS [p^h] is the sound that deviates more from English [p^h]; it shows the shortest VOT duration in English but still in the category over 50 ms, while it shows the longest in our WAS stimuli (51 ms). It may be the case that the similarities in VOT duration of both sounds made L2 listeners understand the cue as their L1 [p^h]. As for English [t^h], its VOT is between that of the bilabial and the velar sounds, but still with a duration of +50 ms, while WAS [t^h] in our stimuli has the shortest duration (43 ms). With regards to WAS [t^h], our analysis revealed that the preceding vowel showed the longest duration; thus, in this particular case, it seems that this aspirated stop has longer pre-aspiration than the other two stops. The key to the higher identification of aspiration in this case may then lie on a longer preceding vowel combined with a shorter VOT.

On the other hand, the L2 listeners in this study also identified sibilance before bilabial [p] less accurately than in the rest of contexts, while before velar [k] and dental [t] it was the most accurately identified. These unaspirated stops showed identical patterns than the aspirated stops, with the exception that their closure duration was significantly longer than their respective VOT. We will comment on these results under the following category, fricatives, since [s] is a fricative sound.

The category of fricatives comprises the fricatized sounds that resulted from the preceding aspiration and intervocalic [h], as well as CS [s] and [z] according to two sets of sounds that follow sibilance: voiceless stops, voiced approximants.

Our acoustic analysis was carried out in terms of spectral peak, amplitude, duration, COG, dispersion, kurtosis, and skewness of the fricative sounds according to their place of articulation. Most studies prove the ability of spectral moments to distinguish sibilants from non-sibilants and that they can reliably distinguish only between sibilants. Shadle and Mair (1996) had already concluded that spectral moments could not reliably discriminate between different places of articulation of the fronted fricatives. However, contrary to the studies mentioned in Section 2.1.2, Jongman et al. (2000) found that spectral moments could successfully discriminate the differences between all four places of fricative articulation in English. Nevertheless, their study differs from the previous ones in the window used to analyze the fricatives (40 ms) as opposed to the narrower windows typically employed (20 ms). To them, the most reliable cues to distinguish a greater number of places of articulation were dispersion and skewness, while dispersion was also successful in distinguishing voice.

Our results indicate that duration was not a reliable cue to differentiate fricatives in terms of place of articulation, i.e. no significant differences were found between any of the fricative sounds, in agreement with Barreiro Bilbao (1994) and Jongman et al. (2000), while in disagreement with García Santos (2002). Furthermore, there were no differences between sibilants [s] and [z] or between the non-sibilants [x] and [h], and [ð] and [x], as well as between [x], [h], and [v]. Spectral peak and COG were reliable cues to distinguish sibilant fricatives from these three non-sibilant fricatives. As mentioned before, Barreiro Bilbao (1994) pointed at spectral peak as the most reliable cue to discriminate fricatives. However, spectral peak could not distinguish the sibilants from [ð]; only the voiceless sibilant [s] was differentiated from [ð] in terms of COG values. Amplitude could only differentiate the two sibilants from [h], contrary to Jongman et al.'s (2000) findings that amplitude could distinguish sibilants from non-sibilants and [v] from [ð]. In relative agreement with their findings, kurtosis and skewness in our study were key factors to distinguish [ð] and the sibilants from [v], [x], and [h]. Finally, dispersion could only differentiate [ð] from [s-z] and [h].

Overall, L2 listeners in our study identified sibilance before velar [k] and dental [t] the most accurately, while sibilance before dental [ð̞] the least accurately. On the contrary, they presented the highest scores for aspiration in interdental [ð], while the lowest for intervocalic [h] and aspiration in velar [x]. So far, it seems that the further the place of articulation of sibilance and aspiration is from the place of articulation of the following consonant, the more accurate their identification, and that (inter)dental sounds in particular lead to special patterns of identification.

In the case of sibilance as a whole, from our analysis we cannot extract meaningful data to fully account for these results unless we separate sibilance in two groups, according to the nature of the following sounds. Within the group of [z] before the three voiced approximants, sibilance before the velar [ɣ̞] was identified best, closely followed by [β̞], while sibilance before dental [ð̞] was remarkably the least accurately identified. The duration of [z] before the velar sound was similar to that of intervocalic [s], whereas in the rest of the contexts sibilance was shorter. In this context, its COG was the lowest of the three, in particular, significantly lower than its value before [ð̞]. Within the group of [s] before the three voiceless stops, L2 listeners identified sibilance before velar [k] and dental [t] the most accurately, whereas identification was the least accurate before [p], although identification percentages were not particularly different between the three contexts. In this last case, [s] before [p] presented the lowest COG value of the three contexts; specifically lower than the COG of [s] before [t], while before [t] it displayed higher dispersion than before [k]. From these data, it appears that a lower COG in [z] in combination with longer duration and a higher COG in [s] in combination with higher dispersion worked in correlation to the identification of sibilance. However, we cannot draw definite conclusions as to which characteristics were the most prominent to favor the identification of sibilance before one context in contrast with a different context.

As to the identification of aspiration in the three fricatized sounds and intervocalic aspiration, the differences in identification percentages were acute. The most accurately identified was [ð], whereas the least accurately identified was [h]. The main differences between these two sounds were higher COG and dispersion, and lower kurtosis and skewness, of the interdental fricative than the glottal fricative. Thus, as with the case with [s] before [t], it also seems that a higher COG in [ð] in combination with higher dispersion –and together lower kurtosis and skewness- made [ð] the most distinguishable of the WAS fricatives, while it also seems to account for the lowest identification of intervocalic aspiration.

Our results also rendered some differences in terms of the gender of the speaker in WAS (as in Horn, 2013; Ruch 2012) as well as in CS. The fact that the L2 listeners tended to a higher identification of the target stimuli when uttered by the female WAS speaker and the male CS speaker makes us believe that the reason for such patterns may be due to other factors apart from the characteristics of the target sounds, not measured in this current study.

6.2 Implications for L2 Speech Perception

In this section, a discussion of the findings in light of the L2 speech perception models reviewed in Chapter 3 is carried out and suggestions for the application of new models to theory and practice are also provided.

The SLM basically proposes that L2 sounds that are similar to L1 sounds will be identified as such, while those L2 sounds that are different will be easily discriminated. PAM argues that, when two L2 sounds that are assimilated to two different L1 categories, discrimination will be excellent (TC type). To this point, predictions would be that CS [s] will be assimilated to English [s] and WAS [h] will be assimilated to English [h]. Nevertheless, according to SLM's H1, "sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level" (Flege, 1995, p. 239); thus, given that [h] is not a legitimate allophone for implosive /s/, assimilation of L2 [h] to L1 [h] was not predictable.

Therefore, in terms of PAM's types of assimilation patterns, contrast between [h] and [V] in WAS sentences was perceived as an SC type in which, generally, both were understood as [V]. On the contrary, contrast between [s] and [V] in CS sentences was generally a TC type of assimilation by which CS [s] was linked to English [s] and CS [V] was linked to English [V]. As AAE1 showed significant differences between the perception of [s] and [V], it may be that for this group assimilation was a CG type, in which [V] was a good exemplar of AAE [V] but [s] sometimes was a poor exemplar of [V].

If we understand WAS [h] and CS [s] as a contrast, a contribution of PAM-L2 is the fact that one L2 can be assimilated to one L1 category, and the second L2 sound, although uncategorized as an L1 sound, can be assimilated on a lexical-functional level (UC type) provided that listeners can "discern at least some of the phonetic differences between the L1 and L2 sounds" (Flege, 1995, p. 239). In this case, learners would need further experience with the L2 and exposure to the L2 variants to equate [h] and /s/ at a lexical-functional level. With increased experience, it can derive into a CG type, in which both L2 sounds are assimilated to the same L1 category, with differences in goodness-of-fit. Ultimate attainment, or native-like performance, would be a SC type of assimilation. The problem with this type of assimilation is that PAM and PAM-L2 relate it to two L2 contrastive phonological categories which are assimilated to one L1 category. In this study, [h] and [s] are two L2 allophones of the same L2 category. Thus, assimilating these L2 variants to the L1 category /s/ would be favorable for L2 acquisition. We claim that PAM and PAM-L2 need to revisit their assumptions to incorporate allophonic and dialectal variation in their types of assimilation.

Whalen, Best, and Irwin (1997) addressed the issue of the perception of two allophones of the same phoneme in the context of English voiceless stops in stressed and unstressed syllable-initial positions. Particularly, they studied how native speakers of English perceived these allophones in correct and incorrect positions within words. They predicted that "if speakers treat context-conditioned allophones as truly being equally good members of a phonological category … then we would expect the allophones to elicit the poor discriminability exhibited by SC non-native contrasts" (p. 505). However, they also claimed that discrimination of such type of allophones may depend on the degree of phonetic difference between them. In our case, we believe that [h] and [s], understood as fricatives, sufficiently differ at least in place of articulation to be differentiated, which they were. CS [s] was assimilated to /s/ but WAS [h] was assimilated to the incorrect category, making [h] and [V] a SC type of assimilation.

Unlike the previous L2 speech perception models, the ASP model states that the type of perception mode that listeners activate depends on the type of task and the context of the stimulus, with special emphasis on the role of attention to allophonic detail and on acoustic as well as perceptual salience. As stated in Section 4.8, this type of task allowed listeners to draw on their language-specific constrains to identify L2 sounds. If this is so, given that phonetic context rendered diverse results depending on level of L2 proficiency or L1 dialect, this study corroborates that experience does not necessarily account for higher accuracy.

This model is more in line with research on speech perception that tries to link phonological and lexical processing, from an abstract to an exemplar-based perception of L2 speech. (Bradlow, 2007; Chéreau, Gaskell, & Dumay, 2007; Cutler & Weber, 2007; Gaskell & Dumay, 2003; Goldinger, 2007; Smith & Hawkins, 2012; Weber & Cutler, 2004). The term *abstract* refers to "representations that are independent of the acoustic properties of specific instances that the perceiver has been exposed to" (Davidson, 2007, p. 59) and also independent of lexical level, while the term *exemplar* conveys that "no such abstract level is necessary for lexical storage; rather, lexical entries are composed of the episodic traces of all of the utterances that a listener has experienced" (p. 59).

As we saw in Section 3.2.1, the NLM was not clear about whether phonological prototypes were stored as an abstract summary of exemplars or as individual instances of these exemplars. Bybee (2002) suggests that "the more frequent variants dominate the category formed from the exemplars and come to be used in a wider range of contexts" (p. 220).

Maye (2007) points out that failure to acquire the attentional weights that are acoustically relevant in the L2 is derived from the listener's L1. "Exemplar representations are therefore biased to some extent by the attentional weights that listeners give to different acoustic/phonetic aspects of the item and these weights will be affected by knowledge of L1 phonology, especially in weaker bilinguals"

(Hazan, 2007, p.41). In fact, McLennan and Luce (2005) believe that less experienced listeners use abstract representations more when processing L2 speech while more experienced listeners use exemplar representations. In the perception of L2 lexical items, “listeners do not have distinct lexical representations for words distinguished minimally by novel L2 contrasts and therefore treat them as homophones” (Escudero, Hayes-Harb, Mitterer, 2008, p. 346), which is what our results vastly showed for the target words in WAS sentences.

Previous models of L2 speech perception generally make use of citation-form or nonsense words in discrimination tasks instead of making use of sentences in identification tasks. Hawkins (2011) defends the use of real words and real sentences:

> using allophones that typically mark discourse functions or speaker identity to test identification of isolated words tells us what listeners do with those allophones in that type of situation, but not what listeners do with them in their natural habitat: detail may be situation-specific. (p. 16)

The phonetic detailed signal carried by sentences is not free of meaning; it reflects phonetic content as well as phonological and grammatical information. Thus, even when the phonetic context following aspiration in our study differed from their counterparts following [V], they still represented the same phonological categories. Hawkins claims that “if the sounds differ systematically, then the structures must differ (even if the phonemes are the same) because the difference in phonetic realization must be represented in the linguistic structure” (p. 388). Precisely, these allophonic variations in our study mark the difference in WAS verb person and noun plurality.

This type of detailed information conveyed by speech can be encoded in memory in several ways and retrieved when necessary, provided that “it was initially attended to and transferred into long-term memory" (Hawkins, 2003, p. 379). Here is where we can understand why the participants in our study vastly understood second-person verbs and plural nouns in WAS sentences as third-person verbs and singular nouns, respectively. This notion reflects SLM's H1, in the sense that "a pattern that violates a powerful principle is less likely to be learned fast" (Hawkins, 2011, p.12). This powerful principle is that [h] does not legitimately represent /s/ in English, and thus, in order to be acquired, it must first be noticed. As Smith (2013a) claims "learning to associate a pronunciation variant with a co-occurring factor, such as a phonetic or phonological environment, utilizes the same cognitive processes as those required to form associations with certain words, a talker characteristic, prosodic position, or grammatical category" (p.202).

As a suggestion for L2 teaching, we see the need to provide dialectal variation in the classroom, especially given that a great number of Spanish dialects make use of aspiration. As found by Cebrian and Chambers (2001), "variable and dialect-specific tendencies in pronunciation are nonetheless used by listeners to contour on-line mappings to lexical candidates" (p. 427), independent of whether these listeners make later use of these variants in production. A good example for teaching is the study by Barden (2011), in which she combined familiarization with a different L1 accent and periodical tests to maintain attention. Familiarization was carried out by means of a story recorded in the target accent, trying to represent real-life conditions. Tests checked for the prediction of whether certain words (homophones) were verbs or nouns, attending to durational differences. Listeners adapted to these differences for the interpretation of the syntactic structure of the lexical items. As she states, "factors such as attention, prior knowledge and auditory salience can influence what is learned by affecting the perceptual salience of phonetic properties and the degree of activation of previously-learned categories" (p. 125).

6.3 Limitations and Future Directions

This section presents the limitations of this study, as well as the suggestions for improvement concerning future research in the field.

The first limitation of this study was the uneven number of participants that comprised each group, particularly with respect to AAE1 (*N* = 97) and GAE1 (*N* = 30). Despite the number of connections with Spanish instructors in the USA, some of them had a limited number of students at elementary levels. Additionally, participation was only voluntary even when extra credit was awarded. The fact that it was necessary to discard some of the participants due to influential characteristics, such as a stay in a Spanish-speaking country or reported speech/hearing disorders, further reduced the number of listeners in the three of the groups. As recruitment was carried out over two consecutive semesters, time limitations are also accountable for.

The second limitation regards the time between stimuli (ISI) presentation and the time for participants to select an option. The identification task in this study was self-paced. Although an average duration of 15 minutes was estimated, participants had no time limit to finish the task. Studies show that at higher ISI, listeners resort to their native-language perceptual patterns to categorize L2 speech rather than make use of basic auditory sensory capabilities (Werker & Tees, 1984, Werker & Logan, 1985). Even when the use of these general auditory mechanisms were not given in this case, otherwise listeners would have been highly accurate in the perception of aspiration, a more controlled and measurable pace is advised.

A third limitation is the two-forced choice alternatives offered as possible answers. Schmidt (2011), other than a lexical identification task, employed a categorization task with nonsense words in her study, in which she provided further response alternatives to [s] and [V]. Even when she used citation-form words, the fact that she included [f] as a possible answer enabled her to see that aspiration was noticed, although not associated to /s/. At this point, we would like to explain that at least some of the AAE listeners stated that "they heard something different but they were not sure", suggesting that they noticed the presence of aspiration but were unable to identify it as /s/. On the one hand, giving them only two forced-choice options limited the possibility of finding out whether indeed they perceived more aspiration than reflected in the results. On the other hand, being real words made it difficult to insert an option that was not valid in real Spanish, at least in the content sentences. A possible solution could be the inclusion of a third option suggesting "none of the above" or "the speaker said something different".

A key point for future research was the fact that, without exposure to the L2 dialect, AAE listeners of elementary level identified L2 [h] similarly to GAE of elementary and intermediate levels but identified L2 [s] significantly lower than any of the GAE groups. Therefore, more research is needed to determine how the underlying system of this group of listeners functions and the effect it has on L2 speech perception.

A further point for future research is related to the acoustic characteristics of the stimuli. Analyses in terms of the effect of the speech rate, syllable stress and prosodic cues of the target words were not taken into account in this study, as proved not to affect the characteristics of the sounds (Torreira, 2007a, 2012; Yao, 2007). Nevertheless, these factors may play a role in the patterns observed for the identification of stimuli according to gender; thus, further research is suggested. Additional future research should also focus on i) the effect of aspiration in nasals and lateral sounds, which fell out of the scope of this current study and thus were not analyzed here, and ii) further examination of aspiration and sibilance in (inter)dental contexts, which yielded particularly interesting results.

6.4 Conclusions

This dissertation has contributed to L2 speech research in several aspects: i) it has provided evidence that dialectal variants of mainstream L2 features are more difficult for L2 learners to identify, ii) it has contributed to the claim that L1 dialect shapes the perception of L2 sounds, iii) it has corroborated that experience and amount of instruction may not predict identification of L2 sounds in some cases, iv) it has proved that the phonetic context of the target stimuli plays a role in L2 speech perception, v) it has provided evidence of the effect that the acoustic characteristics of aspirated stops and fricatives can have on speech perception, and finally, vi) it has found evidence that suggests that theories of L2 speech perception should be revisited.

In the first case, we have seen how all groups of listeners identified aspiration significantly less accurately than sibilance. These results were predictable because the L2 variant under study is not a legitimate realization in the listeners' L1 and because listeners were not exposed to this variant. Nevertheless, we observed an influence of level of proficiency by which listeners of intermediate Spanish identified aspiration more accurately than elementary level listeners, although only AAE2 participants' accuracy was significantly better.

Secondly, this study revealed that L1 dialect shapes L2 speech perception especially at elementary level and in the case of sibilance. We have observed that, at elementary level, GAE listeners identified sibilance significantly better that AAE listeners, and AAE listeners identified aspiration significantly better than GAE listeners. In CS sentences, intermediate level listeners also identified sibilance more accurately than elementary level listeners; however, across groups, GAE listeners identified sibilance significantly better than AAE listeners. This again brings the suggestion that the characteristics of AAE may have influenced the perception of this feature.

Third, we have seen how the amount of L2 instruction may not always predict higher identification. Particularly true is this statement for AAE1 listeners with less than 1 year of instruction, who significantly outperformed GAE speakers at the two levels of proficiency with the same amount of instruction. In fact, as years of instruction progressed, AAE1 listeners' identification of aspiration was less accurate. Since there were no participants in the AAE2 group with less than 1 year of instruction, we cannot determine whether AAE1 listeners would also outperform their intermediate level counterparts. Nevertheless, their performance was similar to that of AAE2 listeners with 1-3 years of instruction, so it may have been possible. In the case of GAE1 listeners, these outperformed AAE1 listeners in the identification of sibilance with less than 1 year of instruction and with 1-3 years of instruction, suggesting that AAE1 initially have more problems with sibilance, which seem to be overcome with proficiency.

Fourth, the identification of aspiration and sibilance was affected by the phonetic context which followed the two L2 variants. While the majority of places and manners of articulation were most accurately identified by either or both intermediate level listener groups, there were a few instances for which elementary level listeners were the most accurate. In particular, AAE1 listeners were especially sensitive to aspiration in voiceless stops and GAE1 listeners were more receptive to sibilance before approximants. In these cases, L1 dialect was directly linked to performance as well. By means of acoustic analyses, this study has provided further insight on the characteristics of WAS aspiration as opposed to mainstream CS. In both cases, aspiration and sibilance in interdental and dental contexts yielded the most significant data, partly providing evidence of the role that the acoustic characteristics of sounds in both dialects can play in L2 speech perception.

Finally, the findings in this dissertation have evidenced the need for current models of L2 speech perception to encompass dialectal variants in their assumptions of how non-native sounds are identified and assimilated by L2 listeners. It has also pointed at new research under the lens of exemplar theory. In light of this new line of research, we further suggest the incorporation of dialectal variants in the L2 classroom, so that more exemplars of the same category can allow a better comprehension of the L2 under study.

REFERENCES

Abramson, A. S. & Lisker, L. (1965) Voice Onset Time in stop consonants: Acoustic analysis and synthesis. *Fifth International Congress of Acoustics*, (pp. 1-18). Liège, Belgium.

Abramson, A. S. & Lisker, L. (1973) Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics, 1*, 1-8.

Alim, H. S. (2004) Hip Hop Nation Language. In E. Finegan & J. R. Rickford (Eds.), *Language in the U. S. A.: Themes for the Twenty-First Century*, (pp. 387-409). West Nyack, NY, USA: Cambridge University Press.

Alvar, M. (1990) *Norma Lingüística Sevillana y Español de América*. Madrid: Ediciones Cultura Hispánica del Instituto de Cooperación Iberoamericana

Alvar, M. (1996) *Manual de Dialectología Hispánica. El Español de España*. Barcelona: Ariel.

Alvar, M. (2006) *¿Existe el dialecto andaluz?* State University of New York, Albany

Alvar López, M. (2004) *Estudios Sobre las Hablas Meridionales.* Granada: Editorial Universidad de Granada.

Alwan, A: Jiang, J., & Chen, W. (2011) Perception of place of articulation for plosives and fricatives in noise. *Speech Communication, 53 (2),* 195-209.

Bailey, G. (2001) The relationsip between African American and White Vernaculars in the American South. In S. Lanehart (Ed.), *Sociocultural and historical contexts of African American English*, (pp. 71-110). Philadelphia: John Benjamins.

Bailey, G. & Thomas, E. (1998) Some aspects of African-American Vernacular English phonology. In S. S. Mufwene, J. R. Rickford, G. Bailey & J. Baugh (Eds.) *African-American English: Structure, History and Use*, (pp. 85-109). London [etc]: Routledge.

Barden, K.J. (2011) *Perceptual Learning of Context-Sensitive Phonetic Detail*. (Doctoral dissertation). University of Cambridge.

Barreiro Bilbao, S.C. (1999) Perception of fricative "voicing" and "place of articulation" in English sibilants. In *Actes del I Congrés de Fonètica Experimental*. Tarragona, (pp. 111-118). Universitat Rovira i Virgili - Universitat de Barcelona.

Barreiro Bilbao, S. C. (1994) *Análisis acústico descriptivo de fricativas español-inglés*. (Doctoral dissertation, Universidad de León).

Barreiro Bilbao, S. C. (2002) Speech perception in L2. *Odisea, 2,* 7-14.

Baugh, J. (2000) *Beyond Ebonics : Linguistic Pride and Racial Prejudice*. Cary, NC, USA: Oxford University Press.

Baugh, J. (2004a) Ebonics and its controversy. In E. Finegan & J. R. Rickford (Eds.), *Language in the US: Themes for the twenty-first century* (pp. 305-318). Cambridge: Cambridge University Press.

Baugh, J. (2004b) Standard English and academic English (dialect) learners in the African diaspora. *Journal of English Linguistics*, *32(3),* 197-209.

Best, C. T. (1995) A direct realist review of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-linguistic Research*, (pp. 171-204). Timonium, MD: York Press.

Best. C. T. & Strange, W. (1992) Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20,* 305-330.

Best, C. T. & Tyler, M. D. (2007) Nonnative and second-language speech perception. Commonalities and complementarities. In O-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, (pp. 31-52). Amsterdam, NLD: John Benjamins Publishing Company.

Beyer, T. & Hudson Kam, C. L. (2012) First and second graders' interpretation of Standard American English morphology across varieties of English. *First Language, 32(3),* pp. 365-384.

Billings, A. C. (2005) Beyond the Ebonics debate: Attitudes about Black and Standard American English. *Journal of Black Studies, 36(1),* 68-81.

Boersma, P. & Hamann, S. (2008) The evolution of auditory dispersion in bidirectional constrint grammars. *Phonology, 25,* 217-270.

Boersma, P. & Weenik, P. (2009) *Praat: Doing Phonetics by Computer* (Version 5.2.46). Retrieved from http://www.praat.org/.

Bohn, O.-S. & Flege, J. E. (1990) Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics,* 11, 303-328.

Boomershine, A. (2006) Perceiving and processing dialectal variation in Spanish: An exemplar theory approach. In T. Face (Ed.), *Selected Proceedings of the 8th Hispanic Linguistics Symposium,* (pp. 58-72). Massachusetts: Cascadilla Press.

Boomershine, A., Hall, K.C., Hume, E. & Johnson, K. (2008) The impact of allophony vs contrast on speech perception. In P. Avery, B. E. Dresher, & K.

Rice (Eds.), *Phonological Contrast*, (pp. 146-172). New York: Mouton de Gruyter.

Bradlow, A. (2007) Information flow and plasticity across levels of linguistic sound structure: Responses to the target papers by Cutler & Weber and by Goldinger. In J. Trouvain and W. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, (pp. 55-58). Saarbrücken, Germany.

Browman, C. & Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology, 6*, 201–251.

Bybee, J. (2002) Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition, 24*, 215-221.

Cano Aguilar, R. & González Cantos, M. D. (2000) *Las Hablas Andaluzas*. Sevilla: Consejería de Educación y Ciencia.

Cebrian, J. (2006) Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*, 372-387.

Cebrian, J. & Chambers, C. G. (2011) Subphonemic detail in lexical preception and production: The case of Canadian Raising. In W-S Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, (pp. 424-427). Hong Kong, China.

Celata, C. (2007) Allophonic variation can affect L2 speech perception: Evidence from a Tuscan neutralization process. In Rauber, A. S., Watkins, M. A., & Baptista, B. O. (Eds.), *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*, (pp. 114-124). Florianópolis, Brazil: Federal University of Santa Catarina.

Cervera, T. & González-Álvarez, J. (2010) Lists of Spanish sentences with equivalent predictability, phonetic content, length, and frequency of the last word. *Perceptual and Motor Skills, 111(2),* 517-529.

Cervera, T, & González-Álvarez, J. (2011) Test of Spanish sentences to measure speech intelligibility in noise conditions. *Behavior Research Methods*, *43(2),* 459-467.

Chen, M. & Alwan, A. (2001) On the perception of voicing for plosives in noise. *Proceedings of EUROSPEECH 2001*, (pp. 175-178). Aalborg, Denmark.

Chen, W. & Alwan, A. (2003) Perception of the place of articulation feature for plosives and fricatives in noise. *Proceedings of the 15th International Conference of Phonetic Sciences,* (pp.1497-1500). Barcelona, Spain: Universitat Autònoma de Barcelona.

Chen, W. & Alwan, A. (2006) On the perception of voicing in syllable-initial plosives in noise. *Journal of the Acoustical Society of America, 119 (2),* 1092-1105.

Chéreau, C., Gaskell, M, & Dumay, N. (2007) Reading spoken words: Orthographic effects in auditory priming. *Cognition, 102*, 341-360.

Chládková, K. & Escudero, P. (2012) Comparing vowel perception and production in Spanish and Portiguese: European versus Latin American dialects. *Journal of the Acoustical Society of America, 131,* EL119-EL125.

Chládková, K. & Podlipský, V. J. (2011) Native dialect matters: Perceptual assimilation of Dutch vowels by Czech listeners. *Journal of the Acoustical Society of America, 130(4),* 186-192.

Cho, T. & Ladefoged, P. (1999) Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27,* 207-229.

Cicres, J. (2011) Los sonidos fricativos sordos y sus implicaciones forenses. *Estudios Filológicos, 48,* 33-48.

Clopper, C.G. & Bradlow, A.R. (2006) Effects of dialect variation on speech intelligibility in noise. *Journal of the Acoustic Society of America*, *119*, 3424.

Clopper, C. G. & Bradlow, A. R. (2008) Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech*, *51(3),* 175-198.

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007) Prosodic effects on acoustic cues to stop voicing and place of articulation. Evidence from Radio News speech. *Journal of Phonetics, 35,* 180-209.

Collins, B. & Mees, I. M. (2003) *The Phonetics of English and Dutch.* Leiden; Boston: Brill.

Craig, H. K. & Washington, J. A. (2004) Grade-related changes in the production of African American English. *Journal of Speech, Language, and Hearing Research, 47,* pp. 450-463.

Cutler, A., Cooke, M., Garcia-Lecumberri, M. L., & Pasveer, D. (2007) L2 consonant identification in noise: Cross-language comparisons. In H. van Hamme, & R. van Son (Eds.), Proceedings of Interspeech 2007 (pp. 1585-1588). Adelaide: Causal productions.

Cutler, A., Smits, R. and Cooper, N. (2005) Vowel perception: Effects of non-native language versus nonnative dialect. *Speech Communication*, *47*, 32-42.

Cutler, A. & Weber, A. (2007) Listening experience and phonetic-to-lexical mapping in L2. In J. Trouvain and W. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, (pp. 43-48). Saarbrücken, Germany.

Davidson, L. (2007) The influence of non-phonetic factors on the form of L2 lexical entries: response to Cutler and Weber. In J. Trouvain and W. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, (pp. 59-62). Saarbrücken, Germany.

de Villiers, J. G. & Johnson, V. E. (2007) The information in third-person /s/: acquisition across dialects of American English. *Journal of Child Language, 34*, 133-158.

Díaz-Campos, M. & Navarro-Galisteo, I. (2009) Perceptual categorization of dialect variation in Spanish. In J. Collentine, M. García, B. Lafford, & F. Marcos Marín (Eds.), *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, (pp. 179-195). Somerville. MA: Cascadilla Proceedings Project.

Dillard, J. L. (1993) *A History of American English.* London [etc.]: Longman.

Escudero, P. & Boersma, P. (2004) Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26(4),* 551-585.

Escudero, P., Benders, T., & Lipski, S. C. (2004) Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics, 37(4)*, 452-465.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008) Novel second-language words and asymmetrical lexical access. *Journal of Phonetics, 26*, 345-360.

Escudero, P. & Williams, D. (2012) Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *Journal of the Acoustical Society of America, 131 (5)*, EL406-EL412.

Face, T. L. & Menke, M. R. (2009) Acquisition of the Spanish voiced spirants by second language learners. In Joseph Collentine et al. (Eds.), *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, (pp. 39-52). Somerville, MA: Cascadilla Proceedings Project

Fasold, R. W. (1972) *Tense Marking in Black English: A Linguistic and Social Analysis*. Arlington: Center for Applied Linguistics.

Felder, L. J (2006) *Perception of African American English word-final stop consonants by Mainstream American English and African American English listeners*. (Doctoral Dissertation). City University of New York.

Felder, L. J. & Strange, W. (2000) Perception of English linguadental fricatives by Haitian Kreyol-African American English bilinguals. *Journal of the Acoustic Society of America, 108(5),* 2653.

Flege, J.E. (1995) Second-language speech learning: theory, findings, and problems in W. Strange (Ed.), *Speech Perception and Linguistic Experience, Issues in Cross-linguistic Research*, (pp. 233-277). Timonium, MD: York Press.

Flege, J. E., Freida, E. M., & Nozawa, T. (1997) Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics (25),* 169-186.

Flege, J. E. & Liu, S. (2001) The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*, 527-552.

Flege, J. E. & MacKay, I. (2004) Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*, 1-34.

Flege, J. E., MacKay, I., & Piske, T. (2002) Assessing bilingual dominance. *Applied Psycholinguistics*, *23*, 567-598.

Flege, J. E., Schirru, C., & MacKay, I. (2003) Interaction between the native and second language subsystems. *Speech Communication, 40,* 467-491.

Flege, J. E., Takagi, N., & Mann, V. (1996) Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, 99, 1161-1173.

Forrest, K., Weismer, G., Milenlovic, P., & Dougall, R. N. (1988) Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, *84 (1),* 115-123.

Fox, R. A., Flege, J. E., & Munro, M. J. (1994) The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America, 97,* 2540-2551.

Frieda, E. M., Walley, A.C., Flege, J. E., & Sloane, M. E. (1999) Adults' perception of native and nonnative vowels: implications for the perceptual magnet effect. *Perception and Psychophysics*, *61(3),* 561-77.

Fulop, S. A. (2011) *Speech Spectrum Analysis*. Berlin, Germany. Springer Berlin Heidelberg

García Santos, J. F. (2002) *Cambio fonético y Fonética acústica.* Salamanca: Ediciones Universidad de Salamanca.

Garrido Almiñana, J. M., Machuca Ayuso, M. J., & de la Mota Gorriz, C. (1998) *Prácticas de Fonética Lengua Española I.* Universitat Autònoma de Barcelona, Servei de Publicacions: Bellaterra.

Gaskell, M. & Dumay, N. (2003) Lexical competition and the acquisition of novel words. *Cognition*, *89,* 105-132.

Gerfen, C. (2002) Andalusian Codas. *Probus, 14,(2),* 247-277.

Goldinger, S. D. (2007) A complementary-systems approach to abstract and episodic speech perception. In J. Trouvain and W. Barry (Eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, (pp. 49-54). Saarbrücken, Germany.

Gordon, M., Barthmaier, P., & Sands, K. (2002) A cross-linguistic study of voiceless fricatives. *Journal of the International Phonetic Association, 32 (2),* 141-174.

Green, L. (1998) Remote past and states in African-American English. *American Speech*, *73(2),* 115-138.

Green, L. J. (2002a) A descriptive study of African American English: Research in linguistics and education. *International Journal of Qualitative Studies in Education, 15*(6), 673-690.

Green, L. J. (2002b) *African American English: A linguistic introduction*. Cambridge: Cambridge University Press.

Green, L. J. (2004a) African American English. In E. Finegan & J. R. Rickford (Eds.), *Language in the USA: Themes for the twenty-first century*, (pp. 76-91). Cambridge: Cambridge University Press.

Green, L. J. (2004b) Research on African American English since 1998. *Journal of English Linguistics, 32*(3), 210-229.

Grosjean, F. (1998) The on-line processing of speech: Lexical access in bilinguals. In Bhatt, P. and Davis, R. (Eds.), *The Linguistic Brain*. Toronto: Canadian Scholars' Press.

Guion, S. G., Flege, J. E., Akane-Yamada, Reiko, & Pruitt, J. C. (2000) An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107(5)*, 2711-2724.

Gutier, T. (2010) *La Lengua Andaluza: Apuntes para su Gramática y Diccionario*. Córdoba: Almuzara.

Hagiwara, R. (2009, November 19) Monthly mystery spectrogram webzone [Web log post]. Retrieved from http://home.cc.umanitoba.ca/~robh/howto.html.

Hawkins, S. (2003) Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics, 31*, 373-405.

Hawkins, S. (2011) Does phonetic detail guide situation-specific speech recognition? In W-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, (pp. 9-18). Hong Kong, China.

Hazan, V. (2007) Second language acquisition and exemplar theory. In J. Trouvain and W. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, (pp. 39-42). Saarbrücken, Germany.

Hedrick, M. & Younger, M. S. (2001) Perceptual weighting of relative amplitude and formant transition cues in aided CV syllables. *Journal of Speech, Language, and Hearing Research, 44 (5), 964-974.*

Herrero Moreno, G. & Supiot Ripoll, A. (2002) Rasgod distintitvos para la percepción de f/b, θ/d, y x/g: ¿tensión, estridencia o sonoridad? In J. García (Ed.) *Actas del II Congreso de Fonética Experimental*, (PP. 220-225). Sevilla, Spain. Universidad de Sevilla.

Horn, M. (2013) *An Acoustic Investigation of Postaspirated Stops in Seville Spanish*. (Master thesis, Ohio State University).

Hughes, G. W. & Halle, M. (1956) Spectral properties of fricative consonants. *Journal of the Acoustical Society of America, 28,* 303–310.

Iverson, P. & Evans, B. G. (2009) Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *Journal of the Acoustic Society of America, 126(2),* 866-877.

Iverson, P. & Kuhl, P. K. (1996) Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America, 99(2),* 1130-40.

Jassem, W. (1979) Classification of fricative spectra using statistical discriminant functions. In B. Lindblom and S. Ohman (Eds.), *Frontiers of Speech Communication Research* (pp. 77-91). New York: Academic Press.

Jiménez Fernández, R. (1999) *El Andaluz*. Madrid: Arco.

Johnson, K. (2012) *Acoustic and Auditory Phonetics* (3rd Edition). Malden, Mass.: Blackwell Publishing Ltd.

Johnson, V. E. (2005) Comprehension of third person singular /s/ in AAE-speaking children. *Language, Speech & Hearing Services in Schools, 36(2),* 116-124.

Jongman, A. & Sereno, J.A. (1995) Acoustic properties of non-sibilant fricatives. Proceedings of the 13th *International Conference of the Phonetic Association, 4,* (pp. 432-435). Stockholm, Sweden.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America, 108,* 1252-1263.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61(5),* 1337-1352.

Kile, S. N. (2007) *The Influence of Dialect on the Perception of Final Consonant Voicing.* (Doctoral dissertation, University of South Florida).

Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R: (2013) Toward improved spectral measures of /s/: Results from adolescents. *Journal of Speech, Language, and Hearing Research, 56,* 1175-1189.

Kuhl, P. K. (1992) Infants' perception and representation of speech: Development of a new theory. In J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge, and G. E. Wiebe (Eds.), *Proceedings of the International Conference on Spoken Language Processing,* (pp. 449-456). Edmonton, Alberta: University of Alberta.

Kuhl, P. K. (1993a) Infant speech perception: A window on psycholinguistic development. *International Journal of Psycholinguistics, 9,* 33-56.

Kuhl, P. K. (1993b) Innate predispositions and the effects of experience in speech perception: The Native Language Magnet Theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, (pp. 259-274). The Netherlands: Kluwer Academic Publishers.

Kuhl, P. K. (1994) Learning and representation in speech and language. *Current Opinion in Neurobiology, 4*, 812-22.

Kuhl, P. K. & Iverson, P. (1995) Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-linguistic Research*, (pp. 121-154). Timonium, MD: York Press.

Labov, W, Ash. S, & Boberg, C. (2006) *The Atlas of North American English: phonetics, phonology, and sound change: a multimedia reference tool.* Berlin: Mouton de Gruyter.

Ladefoged, P. (1982) *A Course in Phonetics*. New York: Harcourt Brace Jovanovich College Publishers.

Ladefoged, P. & Ferrari Disner, S. (2012) *Vowels and Consonants*. Malden, MA: Wiley-Blackwell.

Ladefoged, P. & Johnson, K. (2010) *A Course in Phonetics*. New York: Wadsworth/Cengage Learning.

Ladefoged, P. & Maddieson, I. (1996) *The Sounds of the World´s Languages*. Oxford: Blackwell.

Laing, S. P. (2003) Assessment of phonology in preschool African American Vernacular English speakers using an alternate response mode. *American Journal of Speech-Language Pathology, 12(3)*, pp. 273-281.

Lerer, S. (2007) *Inventing English: A portable history of the language*. New York: Columbia University Press.

Levy, E. S (2009) Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *Journal of the Acoustical Society of America*, *125(2)*, 1138-1152.

Levy, E. S. & Law II, F. F. (2010) Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *Journal of the Acoustical Society of America*, *128(3)*, 1290-1305.

Levy, E. S. & Strange, W. (2008) Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics, 36(4)*, 141-157.

Lisker, L. (1975) Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America, 57 (6),* 1547-1551.

Lisker, L. & Abramson, A. (1964) A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20 (3), 527-565.

López-Bascuas, L. E., Fahey, R. P., García-Albea, J. E. & Rosner, B. S. (1998) Identificación del orden temporal en sonidos de habla y de no-habla, *Cognitiva*, *10*, 195-210.

Lorenz, F. (2012) *Basics of Phonetics & English Phonology*. Logos Verlag: Berlin.

Maniwa, K., Jongman, A., & Wade, T. (2009) Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America, 125 (6),* 3962-3973.

Marrero, V. (1990) Estudio acústico de la aspiración en español, *Revista de Filología Española*, *70 (3),* 345-97.

Marrero, V. (2001) *Fonética y Fonología. Addenda.* Madrid. Universidad Nacional de Educación a Distancia.

Martin, S. & Wolfram, W. (1998) The sentence in African-American vernacular English. In S. S. Mufwene, J. R. Rickford, G. Bailey, & J. Baugh (Eds.), *African-American English: Structure, History, and Use*, (pp. 11-36). New York: Routledge.

Martínez Celdrán, E. (2004) Problems in the classification of approximants. *Journal of the International Phonetic Association,* 34 (2), 201-211.

Martínez Celdrán, E. (2012) Las consonantes fricativas del español. In M.J. Albalá (Ed.), *Lengua Española. Fonética y Fonología,* (pp. 1-16). Madrid: Liceus.

Martínez Celdrán, E. & Fernández Planas, A. M. (2007) *Manual de Fonética Española: Articulaciones y sonidos del español.* Barcelona: Ariel.

Maye, J. (2007) Learning to overcome L1 phonological biases. In J. Trouvain and W. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, (pp. 63-66). Saarbrücken, Germany.

Maza, M. J. (1999) Phonetic data and functional explanation in phonology: The case of Granada Spanish. In K. Oga & G. Poole (Eds.), *Newscastle and Durham Working Papers in Linguistics, V,* (pp. 161-170). Durham, UK: University of Durham.

McLennan, C. T. & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 306-321.

Mitterer, H. & Tuinman, A. (2012) The role of native-language knowledge in the perception of casual speech in a second language. *Frontiers in Psychology, 3*, 1-13. doi: 10.3389/fpsyg.2012.00249

Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (1975) An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psycholinguistics, 18*, 331-340.

Morgan, M. (2001) "Nuthin' but a G thang": Grammar and language ideology in Hip Hop identity. In S. Lanehart (Ed.), *Sociocultural and Historical Contexts of African American English*, (pp. 205-228). Philadelphia, PA, USA: John Benjamins Publishing Company.

Moya Corral, J. A. (2007) Noticia de un sonido emergente: La africada dental procedente del grupo -st- en Andalucía. *Revista de Filología de La Universidad de La Laguna, 25*, 457–465.

Mufwene, S. S. (1998) The structure of the noun phrase in African-American vernacular English. In S. S. Mufwene, J. R. Rickford, G. Bailey, & J. Baugh (Eds.), *African-American English: Structure, History, and Use*, pp. 69-82. New York: Routledge.

Mufwene, S. S. (2001) African-American English. In J. Algeo (Ed.), *The Cambridge History of the English Language*.Vol, 6: *History of American English*, (pp. 291-324). Cambridge: Cambridge University Press.

Mufwene, S. S. (2003) The shared ancestry of African-American and American White Southern Englishes: Some speculations dictated by history. In S. J. Nagle & S. L. Sanders (Eds.), *English in the Southern United States*, (pp. 64-81). West Nyack, NY, USA: Cambridge University Press.

Narbona, A., Cano, R., & Morillo, R. (1998) *El Español Hablado en Andalucía*. Barcelona: Editorial Ariel, S. A.

Ogden, R. (2009) *An Introduction to English Phonetics.* Edinburgh: Edinburgh University Press.

O'Neill, P. (2005) Utterance final /s/ in Andalusian Spanish. The phonetic neutralization of a phonological contrast. *Language Design: Journal of Theoretical and Experimental Linguistics*, *7*, 151-166.

O'Neill, P. (2009) The effect of s-aspiration on occlusives in Andalusian Spanish. *Oxford University Working Papers in Linguistics, Philology & Phonetics* 12, 73–86.

Parrell, B. (2012) The role of gestural phasing in Western Andalusian Spanish aspiration. *Journal of Phonetics* 40(1), 37–45.

Plitcha, B. (2010) *Akustyk* (Version 1.9.3). Retrieved from http://bartus.org/akustyk/index.php.

Polka, L. & Strange, W. (1985) Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. *Journal of the Acoustical Society of America, 78,* 1187-97.

Pollock, K. E., & Hinton, L. (2001) Phonetic transcription of African American Vernacular English. *Communication Disorders Quarterly, 23,* 47-53.

Poplack, S. (2000) *The English history of African American English*. Malden, MA: Blackwell.

Quilis, A. (1966) Sobre los alófonos dentales de /s/. *Revista de Filología Española, 49 (1-4),* 335-343.

Quilis, A. (1981) *Fonética Acústica de la Lengua Española.* Madrid: Gredos.

Reetz, H. & Jongman, A. (2009) *Phonetics: Transcription, Production, Acoustics, and Perception*. Chichester, UK: Wiley-Blackwell.

Rickford, J. R. (1975) Carrying the New Wave into syntax: The case of Black English BIN. In R. Fasold and R. Shuy (Eds.), *Analyzing Variation in Language*, (pp. 162-83). Washington, D. C.: Georgetown U. Press.

Rickford, J. R. (1999) *African American Vernacular English; Features, evolution, educational implications*. Malden, MA: Blackwell Publishers.

Rickford, J. R. & Rickford, R. J. (2000) *Spoken Soul: The story of Black English*. New York: John Wiley & Sons, Inc.

Roach, P. (2000) *English Phonetics and Phonology. A Practical Course* (3rd Edition). Cambridge: Cambridge University Press.

Rojczyk, A. (2011) Perception of the English Voice Onset Time continuum by Polish learners. In J. Arabski and A. Wojtaszek (Eds.), *Second Language Acquisition: Acquisition of L2 Phonology*, (pp. 37-58). Bristol, GBR: Multilingual Matters.

Romero Gallego, J. (1995) *Gestural Organization in Spanish. An experimental study of spirantization and aspiration.* (Doctoral dissertation). University of Connecticut.

Rose, M. (2010) Differences in discriminating L2 consonants: a comparison of Spanish taps and trills. In Matthew T. Prior et al. (Eds.), *Selected Proceedings of the 2008 Second Language Research Forum*, (pp. 181-196). Somerville, MA: Cascadilla Proceedings Project.

Rosner, B., López-Bascuas, L. E., García-Albea, J. E., & Fahey, R. P. (2000) Voice-onset times for Castilian Spanish initial stops. *Journal of Phonetics, 28,* 217-224.

Ruch, H. (2008) La variante [ts] en el español de la ciudad de Sevilla: Aspectos fonético-fonológicos y sociolingüísticos de un sonido innovador. MA Thesis. Universität Zürich.

Ruch, H. (2010) Affrication of /st/-clusters in Western Andalusian Spanish: variation and change from a sociophonetic point of view. *Proceedings of the Workshop "Sociophonetics, at the crossroads of speech variation, processing and communication"*. Pisa, December 14th-15th, 2010.

Ruch, H. (2012) /s/-aspiration and sound change in two varieties of Andalusian Spanish. Abstract presented at the *Sociolinguistics Symposium 19*. Freie Universität Berlin, August 21-24, 2012.

Schmidt, L. B. (2009) The effect of dialect familiarity via a study abroad experience on L2 comprehension of Spanish. In J. Collentine et al. (Eds.) *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, (pp. 143-154). Somerville, MA: Cascadilla Proceedings Project.

Schmidt, L. B. (2011). *Acquisition of Dialectal Variation in a Second Language: L2 perception of aspiration of Spanish /s/.* (Unpublished doctoral dissertation). Indiana University, Bloomington.

Seymour, H. N., Bland-Stewart, L. & Green, L. J. (1998) Difference Versus Deficit in Child African American English. L*anguage, Speech, and Hearing Services in Schools, 29*, pp. 96-108.

Shadle, C. H. & Mair, S. J. (1996) Quantifying spectral characteristics of fricatives. In H. T. Burnell & W. Idsardi (Eds.), *Proceedings of the International Conference on Spoken Language Processing, 3*, (pp. 1521-1524). Philadephia, PA. USA.

Shadle, C. H:, Mair, S. J., & Carter, J. N. (1996) The acoustic characteristics of the front fricatives [f, v, θ, ð]. *Proceeding of ETRW, 4th Speech Production Seminar*, (pp. 193-196). Autrans, France.

Silbert, N., de Jong, K., & Park, H. (2005) Linguistic generalization in L2 consonant identification accuracy: a preliminary report. In K. de Jong and H. Rice (Eds.) *IULC Working Papers Online, Vol. 5*, (pp. 1-15). Bloomington, IN. USA.

Sligh, A. C. & Conners, F. A. (2003) Relation of dialect to phonological processing: African American Vernacular English vs. Standard American English. *Contemporary Educational Psychology, 28*, 205-228.

Smith, B. (2009) Dental fricatives and stops in Germanic: deriving diachronic processes from synchronic variation. In Dufresne, Dupois, and Vocaj (Eds.), *Historical Linguistics 2007*, (pp. 19-36). Amsterdam/Philadelphia: John Benjamins.

Smith , B. (2013a) *The interaction of speech perception and production in laboratory sound change*. Doctoral dissertation, The Ohio State University, Columbus, OH.

Smith, B. (2013b) An Acoustic analysis of voicing in American English dental fricatives. *OSU Working Papers in Linguistics, vol. 60*. (pp. 117-128). Columbus, OH: OSU Department of Linguistics.

Smith, R. & Hawkins, S. (2012) Production and perception of speaker-specific phonetic detail at word boundaries. *Journal of Phonetics, 40,* 213-233.

Smitherman, G. (1998) Word from the hood: The lexicon of African-American Vernacular English. In S. S. Mufwene, J. R. Rickford, G. Bailey, & J. Baugh (Eds.), *African-American English: Structure, History, and Use*, 203-25. New York: Routledge.

Smitherman, G. (1999) *Talkin That Talk: African American Language and Culture.* London, GBR: Routledge.

Spears, A. K. (2001) Directness in the use of African American English. In Sonja Lanehart (Ed.), *Sociocultural and Historical Contexts of African American English*, (pp. 239-259). Philadelphia, PA, USA: John Benjamins Publishing Company, 2001.

Strange, W. (1992) Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In Y. Tokura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds.), *Speech Perception, Production and Linguistic Structure*, (pp. 197-219). Tokyo, Japan.

Strange, W. (2006) Second-language speech perception: The modification of automatic selective perceptual routines, *Journal of the Acoustical Society of America*, *120*, 3137.

Strange, W. (2011) Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *36(1),* 141-157.

Strange, W., Bohn, O.-S., Trent, S. A., & Nishi, K. (2004) Acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, *115*, 1791-1807.

Strange, W., Bohn, O.-S., Nishi, K., & Trent, S. A. (2005) Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, *118*, 1751-1762.

Strange, W. & Shafer, V. L. (2008) Speech perception in second language learners in J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition*, (pp. 153-191). John Benjamins: Amsterdam.

Styler, W. (2013) *Using Praat for Linguistic Research* (Version 1.3.6). Retrieved from http://savethevowels.org/praat/.

Sumner, M. & Samuel, A. G. (2009) The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60(4),* 487-501.

Sussman, J.E. & Lauckner-Morano V.J. (1995) Further tests of the "perceptual magnet effect" in the perception of [i]: identification and change/no-change discrimination. *Journal of the Acoustical Society of America, 97(1),* 539-52.

Thomas, E. R. (2007) Phonological and phonetic characteristics of African American Vernacular English. *Language and Linguistics Compass, 1/5,* 450–475.

Thyer, N., Hickson, L., & Dodd, B (2000) The perceptual magnet effect in Australian English vowels. *Perception and Psychophysics, 62(1),* 1-20.

Torreira, F. (2007a) Pre- and post-aspirated stops in Andalusian Spanish. In Pilar Prieto, Joan Mascaro & Maria-Josep Solé (Eds.), *Prosodic and Segmental Issues in Romance,* (pp. 67–82). Amsterdam: John Benjamins.

Torreira, F. (2007b) Coarticulation between aspirated-s and voiceless stops in Spanish: An interdialectal comparison. In N. Sagarra & A. J. Toribio (Eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium,* (pp. 113–120). Somerville, MA: Cascadilla Press.

Torreira, F. (2012) Investigating the nature of aspirated stops in Western Andalusian Spanish. *Journal of the International Phonetic Association, 42,* 49-63.

Tuinman, A., Mitterer, H., & Cutler, A. (2007) Speakers differentiate English intrusive and onset /r/, but L2 listeners do not. In J. Trouvain and W. Barry (Eds.), *Proceedings of the 16th International Congress Phonetic Sciences,* (pp. 1905-1908). Saarbrücken, Germany.

Tuinman, A., Mitterer, H., & Cutler, A. (2011) Perception of intrusive /r/ in English by native, cross-language and cross-dialect listeners. *Journal of the Acoustical Society of America, 130,* 1643-1652.

Vaux, B. (1998) The laryngeal specifications of fricatives. *Linguistic Inquiry, 29,* 497-511.

Villena Ponsoda, J. A. (2002) Tipología de sistemas fonológicos y variación sociolingüística en el español de Andalucía. In A. Martínez González (Ed.), *Las Hablas Andaluzas Ante el s. XXI,* (pp. 189-213). Almería: Instituto de Estudios Almerienses.

Wagner, A., Ernestus, M., & Cutler, A. (2006) Formant transitions in fricative identification: The role of native fricative inventory. *Journal of the Acoustical Society of America, 120,* 2267-2277.

Weber, A. & Cutler, A. (2004) Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50(1),* 1-25.

Werker, J. F. & Logan, J. (1985) Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*, 35-44.

Werker, J. F. & Tees, R. C. (1984) Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75,* 1866-1878.

Whalen, D. H., Best, C. T., & Irwin, J. R. (1997) Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, *25,* 501-528.

Whalen, D. H., Levitt, A. G., & Goldstein, L. M. (2007) VOT in the babbling of French- and English-learning infants. *Journal of Phonetics, 35,* 341-352.

Widdison, K. A. (1993) Hacia los orígenes de la s aspirada en español. *Estudios de Fonética Experimental, V,* 34-59.

Williams, L. (1977) The voicing contrast in Spanish. *Journal of Phonetics, 5,* 169-184.

Wolfram, W. (2004) The grammar of urban African American Vernacular English. In B- Kortman & E. Schneider (Eds.), *Handbook of Varieties of English*, (pp. 111-132). Berlin: Mouton de Gruyter.

Wolfram, W. (2006) African American English. In B. B. Kachru, Y. Kachru, & C. L. Nelson (Eds.), *The Handbook of World Englishes*, (pp. 328-346). Malden, MA: Blackwell Publishing Ltd.

Wolfram, W. & Thomas, E. R. (2002) *The Development of African American English*. Malden/Oxford: Basil Blackwell.

Wrench, A. (1995) Analysis of fricatives using multiple centres of gravity. Proceedings of the 13th *International Conference of the Phonetic Association, 4,* (pp. 460-463). Stockholm, Sweden.

Yao, Y. (2007) Closure duration and VOT of word-initial voiceless plosives in English in spontaneous connected speech. *UC Berkeley Phonology Lab Annual Report*, 183-225.

Yeni-Komshian, G. H., Robbins, M. & Flege, J. E. (2001) Effects of word class differences on L2 pronunciation accuracy. *Applied Psycholinguistics, 22*, 283–299.

Zeigler, M. B. (2001) Something to *shout* about: AAVE as a linguistic and cultural treasure. In Lanehart, Sonja L. (Ed.), *Sociocultural and Historical Contexts of African American English*, (pp. 187-204). Philadelphia, PA, USA: John Benjamins Publishing Company.

Zhao, S. Y. (2010) Stop-like modification of the dental fricative /ð/: An acoustic analysis. *Journal of the Acoustical Society of America, 128 (4),* 2009-2020.

Zue, V. (1976) *Acoustic Characteristics of Stop Consonants: A Controlled Study.* (Doctoral dissertation). Lexington: Lincoln Laboratory. MIT.

APPENDIX A

GRAMMATICAL FEATURES OF AAE

In AAE, verb tenses are the same as those in GAE, with additional features that do not occur in mainstream English, such as:

Absence

Copula absence. This feature occurs in present tenses.

He tall

They running

Third person singular –s absence. As a consequence, the auxiliary verb employed is *do* and *don't* for all persons.

He walk

He don't sing

Absence or reduction of *will.* This is the consequence of the phonological process of final /l/ deletion or reduction ("l-lessness").

He be here tomorrow

Absence of auxiliary *have*. As a consequence, the use of past participles rather than simple past forms is the common rule for simple past tense.

She seen him yesterday

Also notice that with present perfect tenses, what remains is *untressed been*.

He been sick

APPENDIX A continued

GRAMMATICAL FEATURES OF AAE

Uses

Past participle. Usually realized with the simple past form of the verb.

He had bit

Generalization of *is* and *was.* Employed for second person and plural subjects, instead of *are* and *were*.

They is some crazy folk

We was here

Double tense marking. This implies the reduplication of the suffix in some past tense or past participle forms.

Light-skinded

(Rickford, 1999, p. 6-7)

Verb + –s. For verbs in first person form to indicate habitual actions.

I can show you some of the stuff we tesses them on (Green, 2004a, p. 84)

I gets my check on the first of the month (Smitherman, 1999, p. 24)

Aspectual markers

Other than GAE auxiliaries, AAE counts with three aspectual markers that resemble GAE words but have different meaning and use. These, as other examples we have seen and will see, are "examples of *camouflaging*, the phenomenon in which a vernacular form closely resembles a standard form while being different in structure or meaning." (Martin & Wolfram, 1998, p. 14).

APPENDIX A continued

GRAMMATICAL FEATURES OF AAE

Habitual *be*. This aspectual marker indicates a usual activity or state.

She always be a clown on Halloween (She always dresses as a clown on Halloween)

Her eyes be red (Her eyes are always/usually red)

Breakfast be cooked at 8 o'clock (Breakfast is already cooked by 8 o'clock or Someone usually cooks breakfast at 8 o'clock)

Becky be watching the basketball games (Becky usually watches basketball games)

To negate or emphasize *be*, the use of the auxiliary verb *do* is required.

(Green, 1998, pp. 50 & 57)

To illustrate that the structure *be + V-ing* is not interchangeable with GAE *present continuous*, we include an example found in Dillard (1973):

You makin' sense, but you don't be makin' sense (p. 46)

which means: you are making sense right now, but you usually don't.

Remote BIN. This marker is written as GAE *been* but has a stressed pronunciation and a different meaning. It is employed to indicate that an action, an event or state happened a long time ago or has been happening for a long time. It is not compatible with time expressions such as "two hours ago", unlike *unstressed been.*

Bruce BIN running (Bruce has been running for a long time)

GRAMMATICAL FEATURES OF AAE

Sue BIN knowing he died (Sue has known for a long time that he died)

Bruce BIN in the house (Brice has been in the house for a long time)

That house BIN brown (That house has been brown for a long time)

For the negative form and for emphasis, *BIN* requires the use of the auxiliary verb *have* (usually *ain't* for the negative form).

(Green, 1998, pp. 117-8, 130)

Resultant state *dən*. Written as GAE *done* but with an unstressed pronunciation, this marker is used to indicate the completion of an action (usually with a present result).

Bruce dən lost his wallet (Bruce has just lost his wallet)

I dən saw him today/this month/this year (I have seen him today/this month/this year)

In negative and emphatic sentences, it requires the use of the auxiliary verb *have* (usually *ain't* for the negative form).

(Green, 1998, p. 48)

Future perfective *be dən*. This is similar in meaning to GAE future perfect progressive tense, sometimes just simple future tense.

They'll probably be dən growed out that by then (They will probably have grown out of that by then)

Boy I make any kind of move, this boy be dən shot me (If I move, this boy will shoot me)

(Green, 1998, p. 49)

GRAMMATICAL FEATURES OF AAE

Pseudo-markers

Come + V-ing indicates indignation on the part of the speaker.

She come telling me it was hot (She had the nerve/audacity to tell me it was hot)

Steady marks that an action has been done in a sustained manner.

She steady talking (She is talking nonstop)

(Green, 2004a, p. 84)

Sentence patterns

Double negation is common in AAE.

Nobody don't be at the library (Nobody is usually at the library)

It is also usual to find **negative inversion** in declarative sentences.

Don't nobody be at the library (Not a single person is usually at the library)

Additionally, we can also find **existential negative** sentences.

It don't be nobody at the library (Usually, there isn't anybody at the library)

(Green, 2002a, p. 686)

GRAMMATICAL FEATURES OF AAE

Finally, the common negation element in AAE is ***ain't(t).*** It can be used instead of GAE *am not, isn't, aren't, hasn't, haven't,* and *didn't.*

He ain' here

He ain' do it

(Rickford, 1999, p. 8)

To these examples we can also add the constructions ***Ain't but*** and ***Don't but.***

She ain nothin but a kid (She is only a kid)

Don't but two people know what really happen

(Only two people know what really happened)

(Smitherman, 1999, p. 33)

Double modals and quasi modals. There are several combinations of modal verbs which mean GAE *might be able to*, such as AAE *may can, might can,* and *might could*. Even more unique to AAE is the use of *must don't* for GAE *must not*. Quasi-modals are exemplified by *useta, poseta,* and *liketa.*

He might could do the work

They useta could do it

I liketa drowned (I nearly drowned)

You don't poseta do it that way (You're not supposed to do it that way)

(Martin & Wolfram, 1998, p. 32-3; Rickford, 1999, p. 7)

APPENDIX A continued

GRAMMATICAL FEATURES OF AAE

Relative clauses. As is the case with GAE, relative pronouns that function as objects can be deleted. What is particular about AAE is that relative pronouns with a subject function can also be omitted.

He got a gun sound like a bee

(Dillard, 1973, p. 68)

Questions. In direct questions, the auxiliary verb is usually omitted or not inverted, given rise to a question with declarative sentence structure.

Where the kids went?

Who that is?

In indirect questions, they generally follow a direct question structure.

They asked could she go to the show

(Martin & Wolfram, 1998, p. 29)

***If* absence.** In conditional sentences, the absence of *if* is compensated by intonation.

A man get rich, he still pay taxes

(Dillard, 1973, p. 64)

Tell say. This is a combination of verbs in which *say* is usually employed as if it were *that, what,* or *whether*.

They told me say they couldn't go

(Rickford, 1999, p. 9)

GRAMMATICAL FEATURES OF AAE

Immediate future. To talk about an immediate future, AAE usually employs the expression *fixing to* (usually reduced to *finna*).

He finna go (He's about to go)

(Rickford, 1999, p. 6)

Preterite *had*. This structure is employed when narrating events and actions in the past, and its function is similar to GAE simple past.

I had got sick when I went to the fair

(Green, 1998, p. 43)

Stative use of *here go* and *there go*. These two expressions are used with stative meaning as in GAE *here is/there is* and *here are/there are.*

There go my momma on the front row

(Smitherman, 1999, p. 23)

Existential constructions. *It's* and *they got* are used instead of *there is* and *there are.*

It's a school up there

They got some hungry women here

(Rickford, 1999, p. 9)

Nouns and Pronouns

Absence of Saxon genitive. Possession is indicated by juxtaposition.

John house (John's house)

APPENDIX A continued

GRAMMATICAL FEATURES OF AAE

Plural absence. Especially in the presence of other plural indicators.

Two boy (Two boys)

Use of *dem* instead of *those.*

She don't know bout dem shoes you bought

Associative plural. It is common to use *and them* (usually reduced to *an dem* or simply *nem*) after a person to indicate a group of people associated with them.

The boy nem lef already when I got here

For other than human beings, the usual expression is *and things,* reduced to *an thing(s*).

He don' like coffee an' thing(s)

(Mufwene, 1998, p. 79)

Pronominal apposition. This feature is used to introduce the topic or for emphasis but it is never a double subject.

Now Robert, he don't know where he going

(Smitherman, 1999, p. 24)

Pronouns. The use of *yall* to mark second person plural ("you all") and *yall's* or *y'all* for second person plural possessive adjective *your* and pronoun *yours.*

It's y'all ball (It's your ball)

Use of *they* instead of *their*.

It's they house (It's their house)

Use of object pronouns after a verb to mean "(for) myself", (for) himself, etc."

Ahma git me a gig (I'm going to get myself some support)

APPENDIX B

GRAMMATICAL FEATURES OF AS

Andalusian morphology and syntax are those of CS, with divergence found in cases affected by phonetic features, such as loss if implosive /s/, and also observed in the preferred usage of certain forms. Narbona et al. (1998) give an extensive account of the main grammatical characteristics of this dialect, while Alvar (1996, 2006) and Jiménez Fernández (1999), other than highlighting some of these points, also include AS lexical items and their origin.

There exists a preference for the use of present perfect tense instead of simple past tense in AS, in direct opposition to the preference in the northeast region of Spain.

There is also a tendency to employ simple present verbs with future meaning and to use verbal periphrasis for the same purpose, instead of simple future tense.

It is also common for many speakers to insert the preposition *de* between the conjugated verb and the following infinitive form in verbal periphrasis structures.

Lo vi de venir

In the past perfect tense of the subjunctive form, some speakers employ the verb *ser* instead of *haber* as the auxiliary verb.

Si yo lo fuera visto… (Si yo lo hubiera visto)

In genitive cases, speakers seem to use a structure similar to that of English, instead of following the standard use in Spanish.

Mi amiga, su novio, el hermano está en el paro.

Also given in other varieties of Spanish, as well as in Catalan and Portuguese, speakers can insert a definite article before a person's name in informal contexts.

Ha llamao la María

It is also usual to invert the order in certain expressions, such as

Me se ha caído.

Más nunca, más nadie, más nada.

In conditional sentences, the conditional verb is usually in imperfect past tense.

Yo que tú, me la compraba. (Si yo fuera tú, me la compraría)

In relative sentences, the use of *cuyo/cuya* and *el/la cual*, as well as the prepositions which accompany them, is generally absent.

Está saliendo con ese chico que el padre es médico.

Su padre es un hombre que siempre le ha gustado luchar por la vida.

It is not uncommon to find that the syntactic structure SVO is altered and the object is fronted.

Yo, vino, bebo sólo cuando como.

... cuando se sale uno de la habitación, las luces... se apagan.

Nouns and Pronouns

The most remarkable feature of Andalusian nouns is the formation of plural nouns affected by the aspiration or loss of implosive /s/. Nevertheless, the opposition singular-plural can be figured out by a) aspiration, b) opening or lengthening of final vowel. In any case, the context itself, as well as other clues such as determiners, articles and adjectives, can solve this ambiguity. For instance, the use of *el* would indicate singular masculine noun, while *lo[h]* would mark the plural form. The problem in this case may arise with feminine nouns (*la* vs. *la[h]*).

The use of *ustedes* in CS is limited to expressing formality for the second person plural *vosotros.* However, this pronoun can be employed without distinction of formality, as is the case in the Canary Islands and some Spanish-speaking countries in South America. In this instance, the verb is conjugated as third person plural, as in *ustedes quieren.* Nevertheless, in western Andalusia *ustedes* can be employed instead of *vosotros* with the verb still conjugated in the second person plural form, as in *ustedes queréis.* This feature also extends to pronominal pronouns and imperative forms, such as *ustedes os vais – ustedes se vais – ustedes se van - ustedes sos vais – ustedes sus vais; irse ya.*

Also in western Andalusia we find a functional use of subject pronouns, instead of a stylistic use, given the similarity among verbal forms (loss of final /s/ and /n/, use of *ustedes*):

yo vengo

tú viene

él viene

nosotros venimo

vosotros/ustedes vení

ellos vienẽ

As far as *le, la, lo* is concerned, AS maintains their etymological value against the innovations spreading in other areas of the country.

APPENDIX C

LIST OF SENTENCES

SINGULAR NOUNS

SNPA	Digo <u>cola</u> por la tarde
SNPB	Digo <u>amigo</u> por la tarde
SNTA	Digo <u>mano</u> torpemente
SNTB	Digo <u>coche</u> torpemente
SNKA	Digo <u>cama</u> con cuidado
SNKB	Digo <u>libro</u> con cuidado
SNBA	Digo <u>playa</u> vagamente
SNBB	Digo <u>dedo</u> vagamente
SNDB	Digo <u>gata</u> demasiado lento
SNDA	Digo <u>pelo</u> demasiado lento
SNGA	Digo <u>leche</u> gritando
SNGB	Digo <u>boca</u> gritando
SNMA	Digo <u>vino</u> muy rápido
SNMB	Digo <u>agua</u> muy rápido
SNNB	Digo <u>cuadro</u> nuevamente
SNNA	Digo <u>pata</u> nuevamente
SNLB	Digo <u>rata</u> lentamente
SNLA	Digo <u>noche</u> lentamente
SNVOA	Digo <u>niño</u> una vez
SNVOB	Digo <u>perro</u> una vez

APPENDIX C continued

LIST OF SENTENCES

PLURAL NOUNS

PNPA	Digo niños por la tarde
PNPB	Digo perros por la tarde
PNTA	Digo colas torpemente
PNTB	Digo amigos torpemente
PNKA	Digo manos con cuidado
PNKB	Digo coches con cuidado
PNBB	Digo camas vagamente
PNBA	Digo libros vagamente
PNDA	Digo leches demasiado lento
PNDB	Digo bocas demasiado lento
PNGA	Digo vinos gritando
PNGB	Digo primos gritando
PNMB	Digo playas muy rápido
PNMA	Digo dedos muy rápido
PNNA	Digo ratas nuevamente
PNNB	Digo noches nuevamente
PNLB	Digo aguas lentamente
PNLA	Digo patas lentamente
PNVOA	Digo gatas una vez
PNVOB	Digo pelos una vez

APPENDIX C continued

LIST OF SENTENCES

SECOND-PERSON SINGULAR VERBS

2PVPA	Espero que <u>vengas</u> por la tarde
2PVPB	Nunca <u>pides</u> permiso a nadie
2PVTA	<u>Deberías</u> tener más cuidado
2PVTB	<u>Necesitas</u> tiempo para pensar
2PVKA	No <u>creas</u> que voy a venir luego
2PVKB	Me <u>arañas</u> con esas uñas
2PVBB	<u>Tienes</u> bebidas en la nevera
2PVBA	Nunca <u>quieres</u> venir conmigo
2PVDA	<u>Comes</u> demasiado deprisa
2PVDB	<u>Estás</u> decorando nuestra casa
2PVGA	No <u>comes</u> galleta de postre
2PVGB	No <u>sabes</u> guardar un secreto
2PVMB	<u>Necesitas</u> más horas de sueño
2PVMA	<u>Comes</u> mucha carne roja
2PVNA	No <u>sabes</u> ninguna respuesta
2PVNB	Nunca <u>comes</u> nada dulce
2PVLB	<u>Deberías</u> limpiar la cocina
2PVLA	Aún no <u>has</u> lavado la vajilla
2PVVOA	Siempre <u>bebes</u> agua mineral
2PVVOB	Hoy <u>comes</u> hamburguesa de cerdo

APPENDIX C continued

LIST OF SENTENCES

THIRD-PERSON SINGULAR VERBS

3PVPA	Es mejor que pida permiso
3PVPB	Adora pintar en acuarela
3PVTA	Tiene terreno en el campo
3PVTB	Está tomando mucha verdura
3PVKA	Iba cargado como un mulo
3PVKB	Lleva cantando todo el día
3PVBA	No quiere bailar con mi amiga
3PVBB	Ha batido el récord mundial
3PVDB	Nunca discute de política
3PVDA	Tiene demasiado trabajo
3PVGA	Necesita guardar silencio
3PVGB	Seguro que compra granadas
3PVMA	Compra melón en el campo
3PVMB	Quiere meditar las opciones
3PVNB	Siempre come naranja de postre
3PVNA	No sabe nadar en el mar
3PVLB	Es probable que tenga la gripe
3PVLA	Siempre toma leche caliente
3PVVOA	Nunca bebe agua muy fría
3PVVOB	Debería actuar con cuidado

APPENDIX D

INFORMED CONSENT FORM

You are invited to participate in a research study on the perception of Spanish by speakers of American English.

You will first be asked to answer a few questions to create your linguistic profile. You will then listen to a set of Spanish sentences and choose the one you hear from the two options given. Play each sentence twice. The first five sentences will illustrate how to proceed.

- There is no scientific evidence to suggest there should be any health risks derived from this activity.
- Data will be treated confidentially and your name will not be used anywhere. Results may be used for publication and/or conference participation.
- Participation in this study is voluntary. You may stop at any time, in which case, data obtained up to that point may be used for analysis.
- Should you have any questions, please address María del Saz (msdelsaz@gmail.com).

I have read the above information and I voluntarily agree to participate in this experiment.

Participant's initials ______________________

Participant's email ______________________

Date ______________________

1. Download Google Chrome if not installed
2. Go to: http://personal.us.es/mdelsaz/Test/ (the word Test should be capitalized)

APPENDIX E

LANGUAGE BACKGROUND QUESTIONNAIRE

1. Age:

2. Gender:

    - Male
    - Female

3. Birthplace (town/province/country):

4. Birthplace of parents/guardians (town/province/country):

5. As a child, what languages were spoken at home? (by parents, guardians, relatives, etc.):

6. As a child, what languages were spoken outside of home? (school, etc.):

7. What is your English accent/dialect? (check all that apply):

    - General American
    - Southern
    - Western
    - Midland
    - Northern
    - African-American
    - Chicano

8. Have you lived in a Spanish-speaking country for over 3 months?:

    - No
    - Yes (please, specify where, when, how long)

9. What dialect/accent of Spanish are you learning/do you speak?:

APPENDIX E continued

LANGUAGE BACKGROUND QUESTIONNAIRE

10. Years of Spanish instruction/learning:

- Less than 1 year
- 1-3 years
- 3-5 years
- More than 5 years

11. What is your level of Spanish?:

- Beginner (level 1)
- Intermediate (Level 2)
- Upper-intermediate (Level 3)
- Advanced (Level 4)
- Proficiency (Level 5)

12. What other languages do you speak?

13. Do you or anyone in your family have or have had any type of hearing/speech impairment?

- No
- Yes (please, specify)

APPENDIX F

SPOKEN ENGLISH QUESTIONNAIRE

Choose the option that best represents your **spoken** language in informal situations, regardless of how you would write it.

1. Where is your friend?
   - She working right now
   - She's working right now

2. What do you do on the weekends?
   - I always be playing ball
   - I usually play ball

3. Does he like soup?
   - No, he never eat that
   - No, he never eats that

4. What do you think of her?
   - She nice
   - She's nice

5. Who wants some tea?
   - Nobody likes that
   - Don't nobody like that

6. Who's that?
   - It's they brother
   - It's their brother

7. I'm really hungry.
   - There's food in the kitchen
   - They got food in the kitchen

8. Who's out there?
   - Two boy playing
   - Two boys playing

APPENDIX F continued

SPOKEN ENGLISH QUESTIONNAIRE

9. I need to pick her up at 6pm
   - She be done finish by then
   - She'll have finished by then

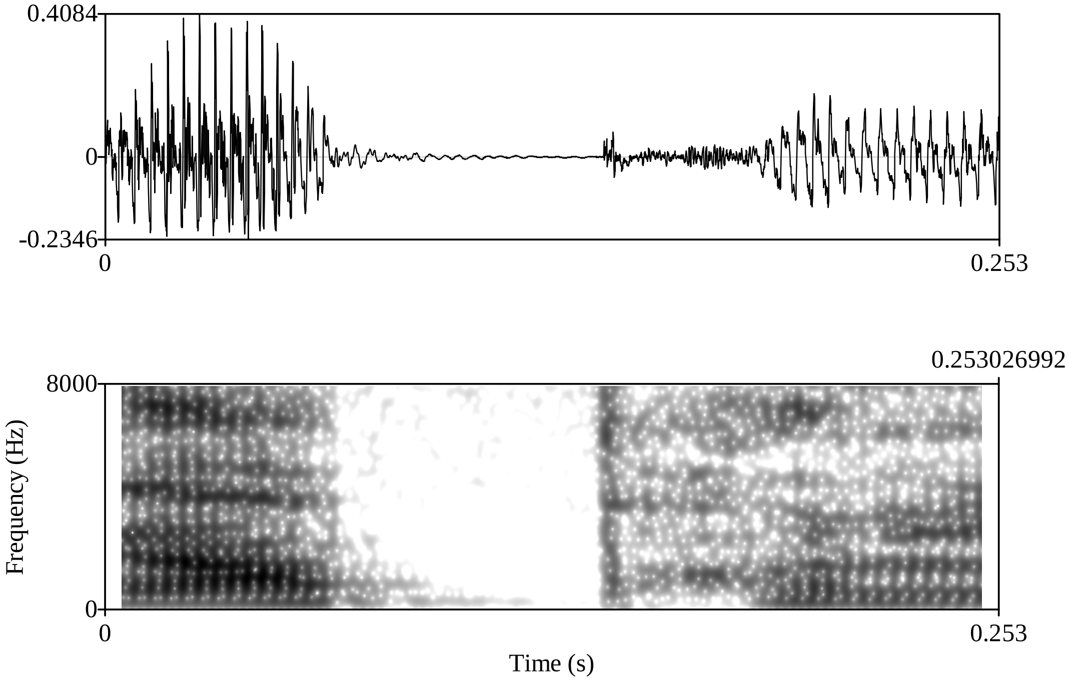
10. How would you normally say (not write) "cold air"?
    - cold air
    - col' air

11. How would you normally say (not write) "She picked us up"?
    - She picked us up
    - She pick us up

12. How would you normally say (not write) "Stop for a minute"?
    - Stop for a minute
    - Stop fo' a minute

APPENDIX G

WAVEFORMS AND SPECTROGRAMS OF WAS ASPIRATED STOPS



VOWEL	CLOSURE	VOT	VOWEL

Figure 43. *Waveform and spectrogram of [p^h]*

WAVEFORMS AND SPECTROGRAMS OF WAS ASPIRATED STOPS

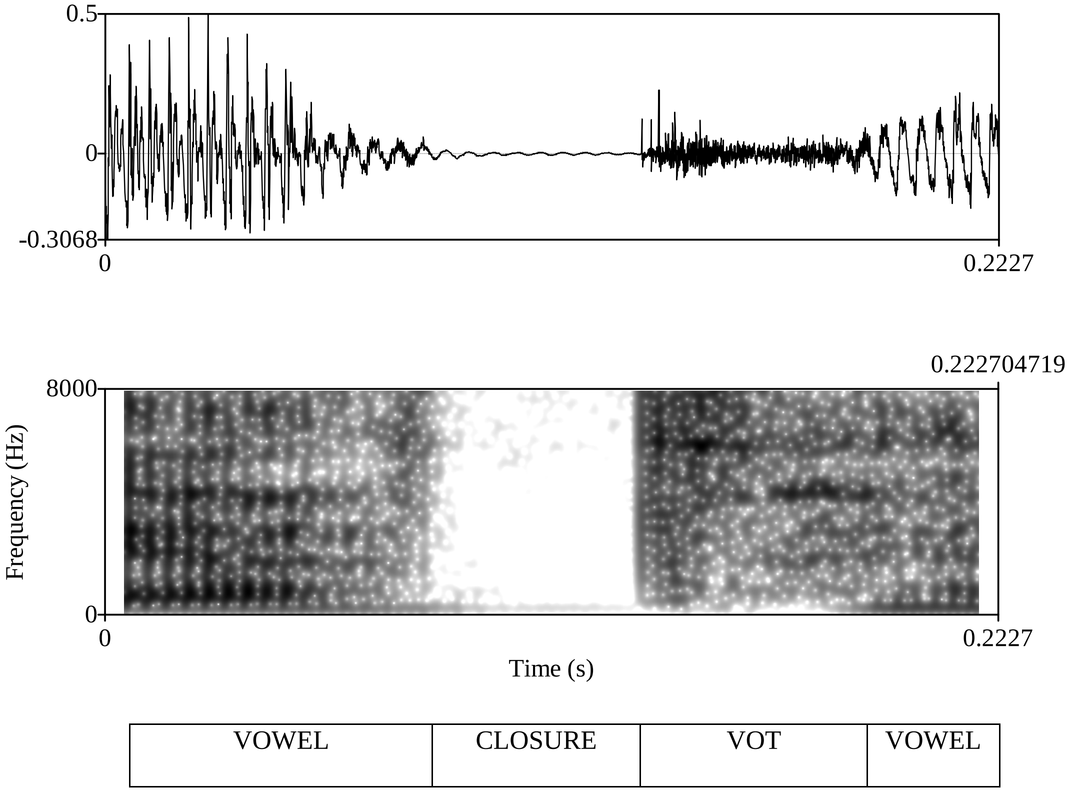
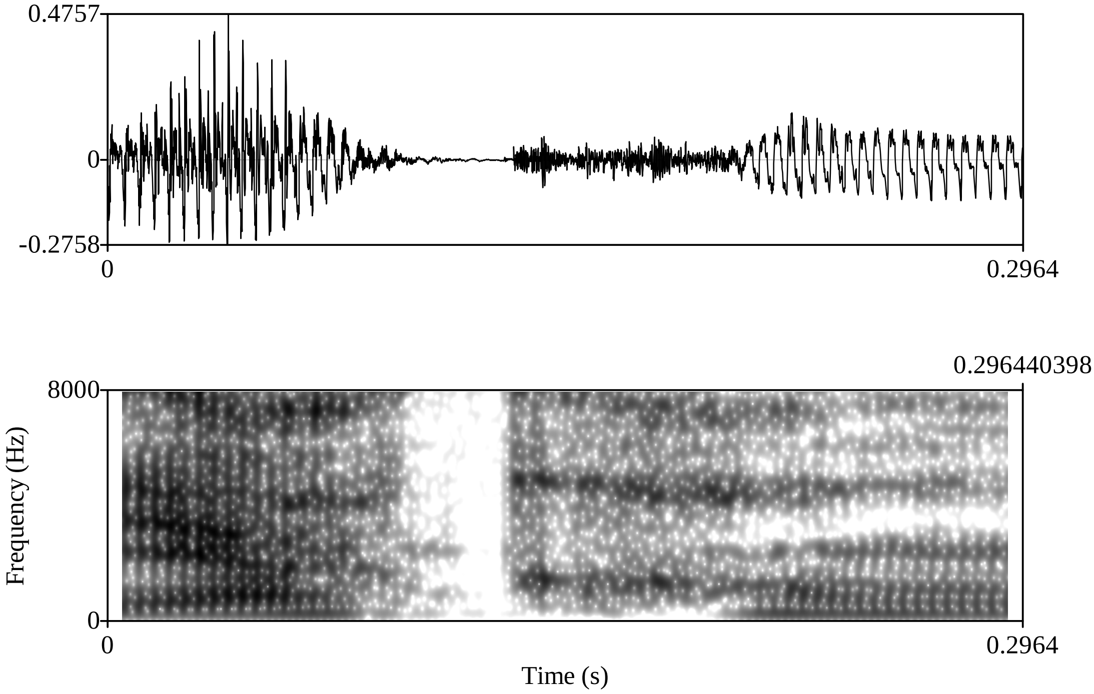


Figure X. *Waveform and spectrogram of* $[t^h]$

WAVEFORMS AND SPECTROGRAMS OF WAS ASPIRATED STOPS

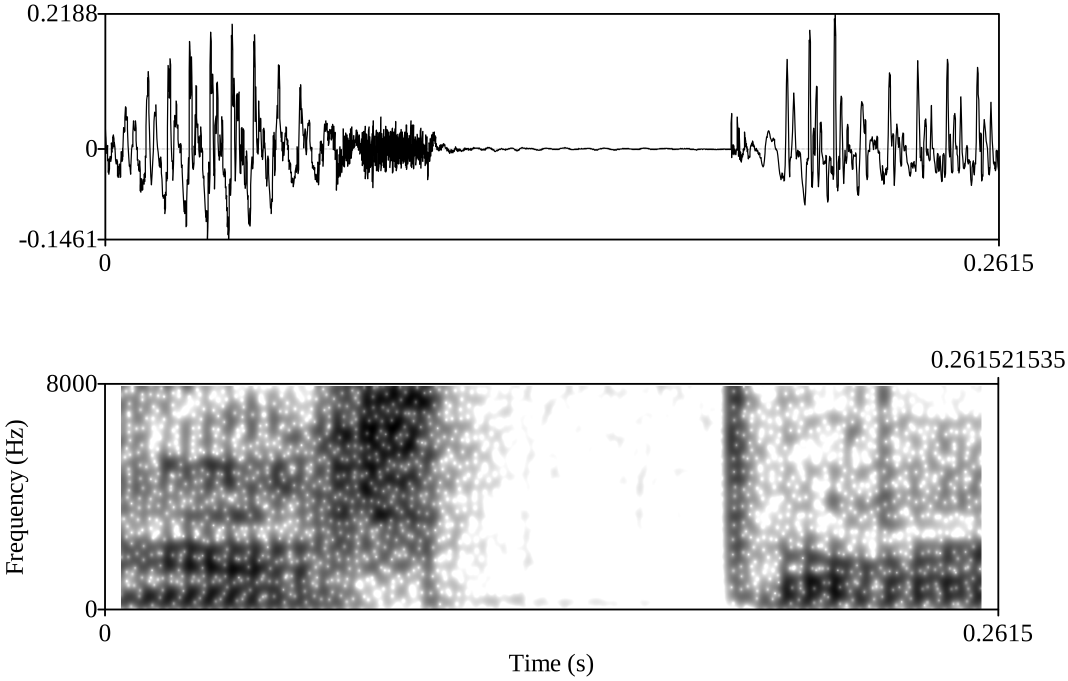


VOWEL	CL	VOT	VOWEL

Figure 45. *Waveform and spectrogram of [k^h]*

APPENDIX H

WAVEFORMS AND SPECTROGRAMS OF CS UNASPIRATED STOPS

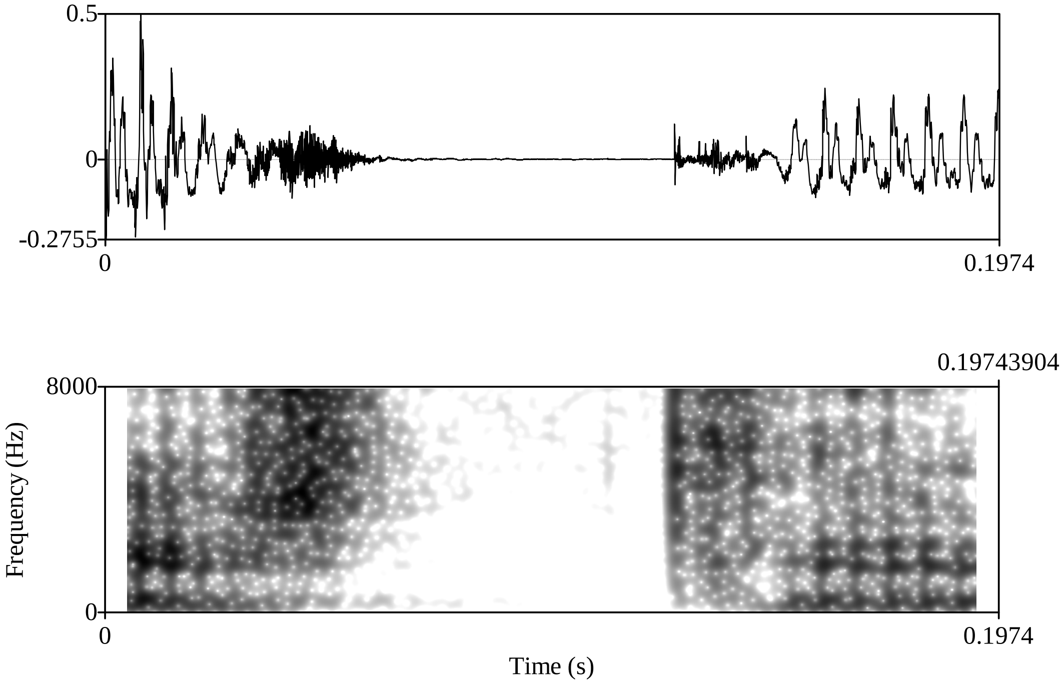


VOWEL	[s]	CLOSURE	VOT	VOWEL

Figure 46. *Waveform and spectrogram of [sp]*

APPENDIX H (continued)

WAVEFORMS AND SPECTROGRAMS OF CS UNASPIRATED STOPS



VOWEL	[s]	CLOSURE	VOT	VOWEL

Figure 47. *Waveform and spectrogram of [st]*

APPENDIX H (continued)

WAVEFORMS AND SPECTROGRAMS OF CS UNASPIRATED STOPS

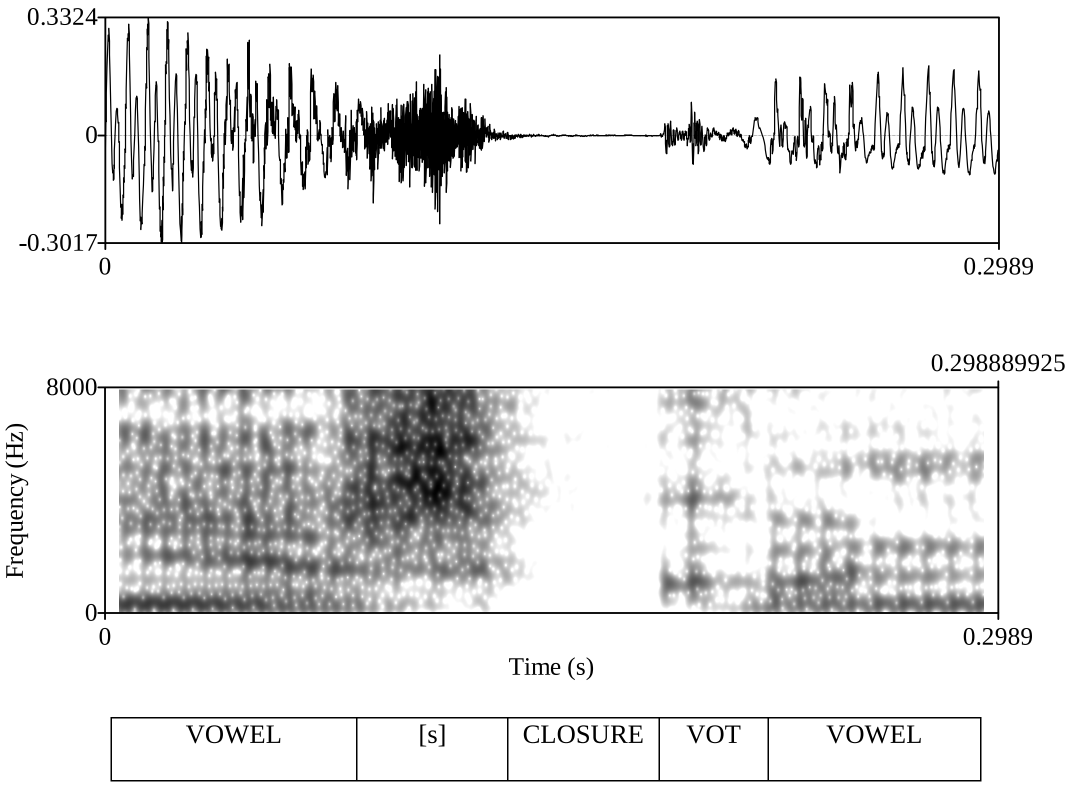
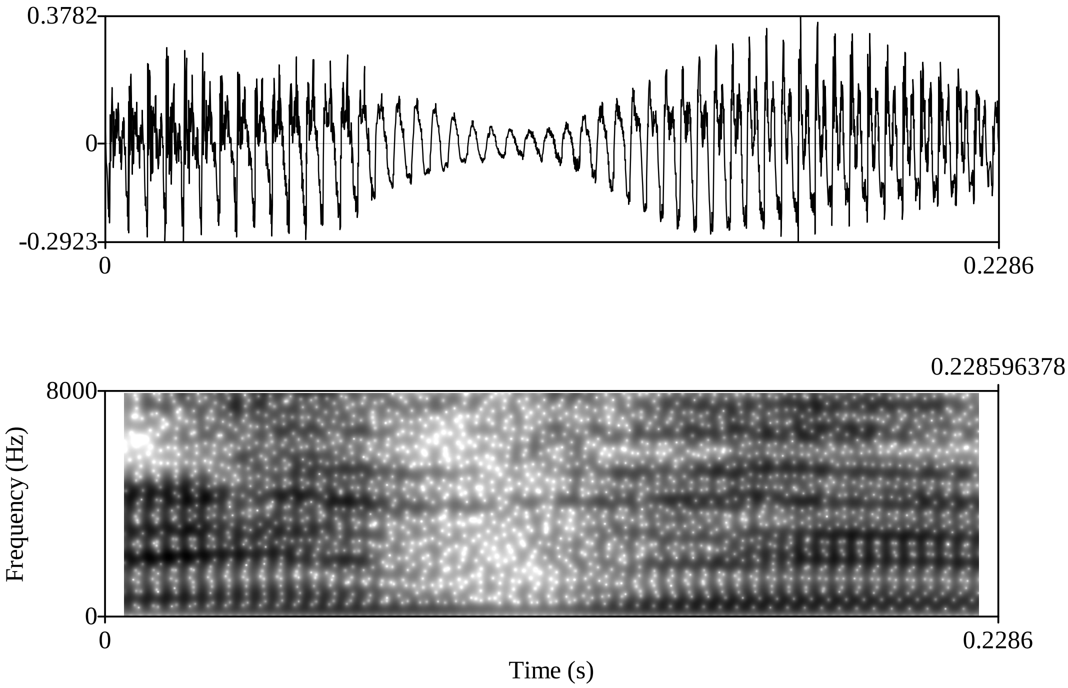


Figure 48. *Waveform and spectrogram of [sk]*

APPENDIX I

WAVEFORMS AND SPECTROGRAMS OF WAS FRICATIZED STOPS

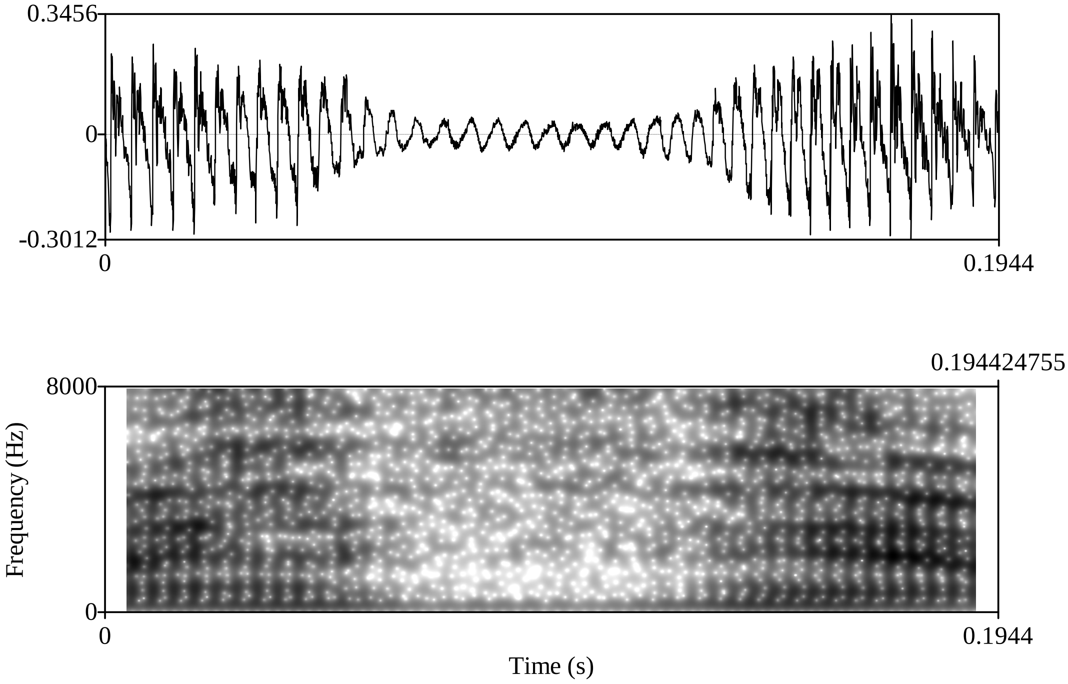


VOWEL	[v]	VOWEL

Figure 49. *Waveform and spectrogram of [v]*

APPENDIX I (continued)

WAVEFORMS AND SPECTROGRAMS OF WAS FRICATIZED STOPS



VOWEL	[ð]	VOWEL

Figure 50. *Waveform and spectrogram of [ð]*

APPENDIX I (continued)

WAVEFORMS AND SPECTROGRAMS OF WAS FRICATIZED STOPS

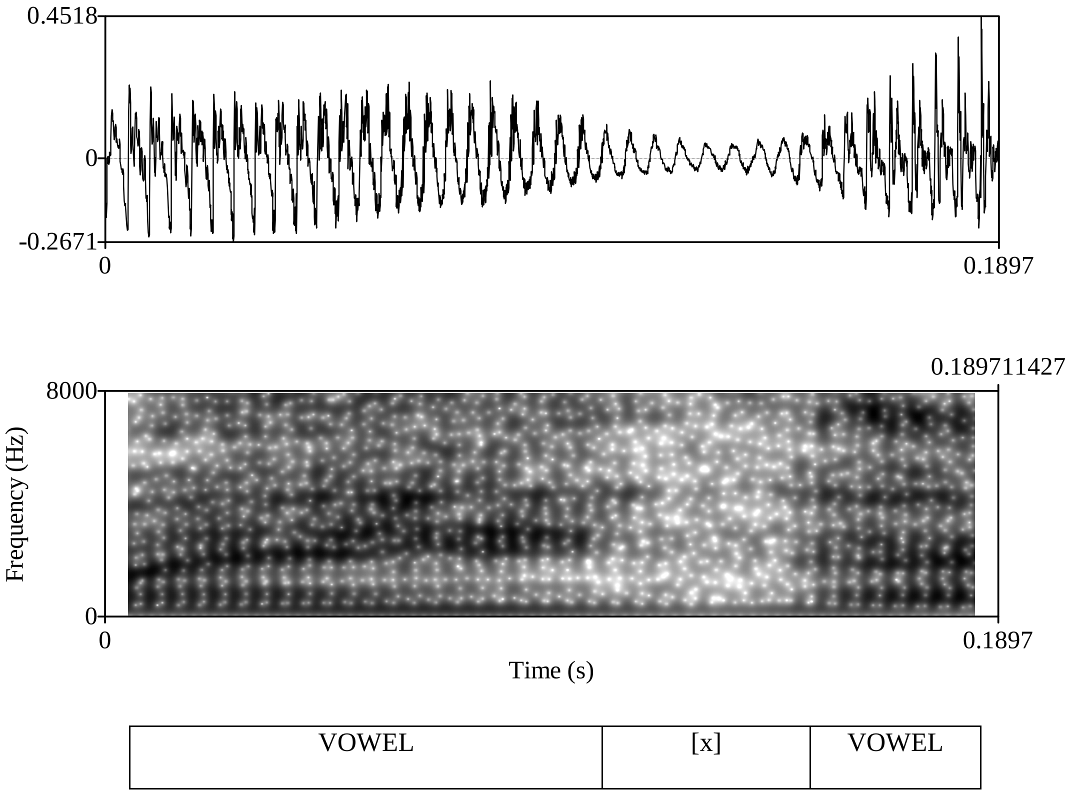
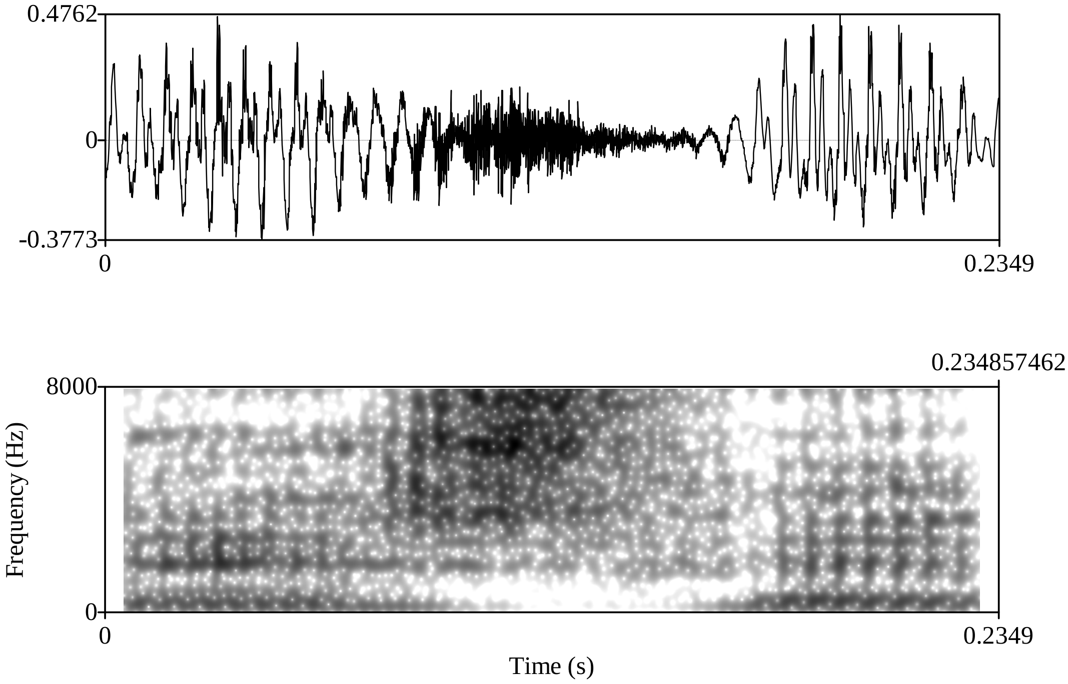


Figure 51. *Waveform and spectrogram of [x]*

APPENDIX J

WAVEFORMS AND SPECTROGRAMS OF CS SIBILANTS BEFORE APPROXIMANTS



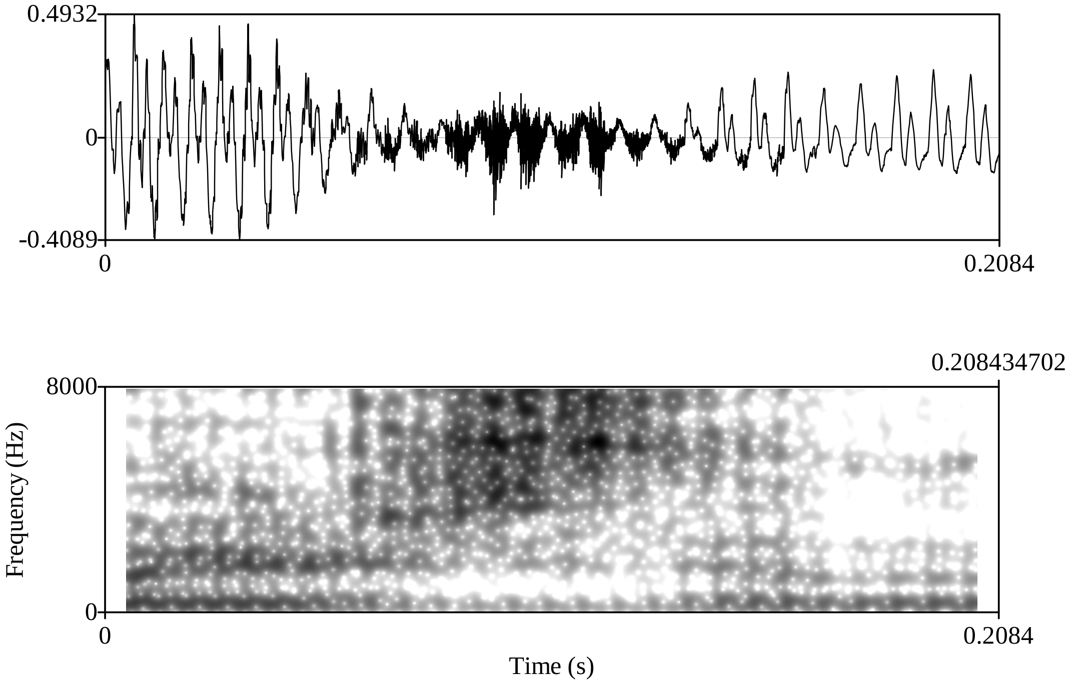
VOWEL	[z]	[ß̦]	VOWEL

Figure 52. *Waveform and spectrogram of [zß̦]*

APPENDIX J (continued)

WAVEFORMS AND SPECTROGRAMS OF CS SIBILANTS BEFORE

APPROXIMANTS

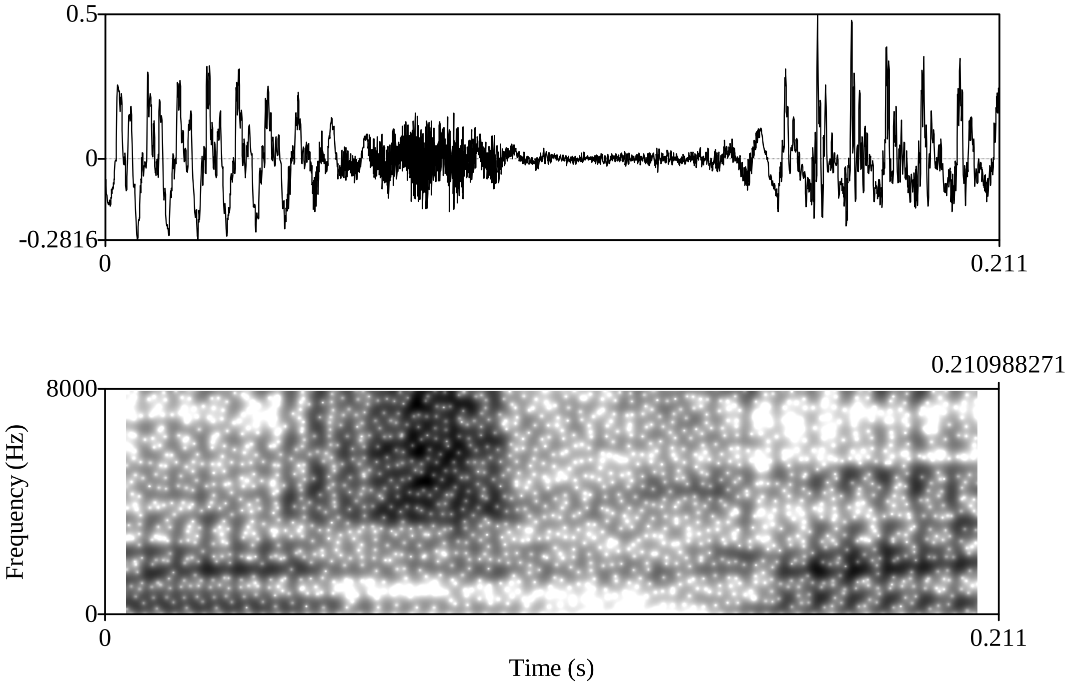


VOWEL	[z]	[ð̩]	VOWEL

Figure 53. *Waveform and spectrogram of [zð̩]*

WAVEFORMS AND SPECTROGRAMS OF CS SIBILANTS BEFORE

APPROXIMANTS



VOWEL	[z]	[ɣ̥]	VOWEL

Figure 54. *Waveform and spectrogram of [zɣ̥]*

APPENDIX K

SPECTRAL SLICES OF WAS FRICATIVES

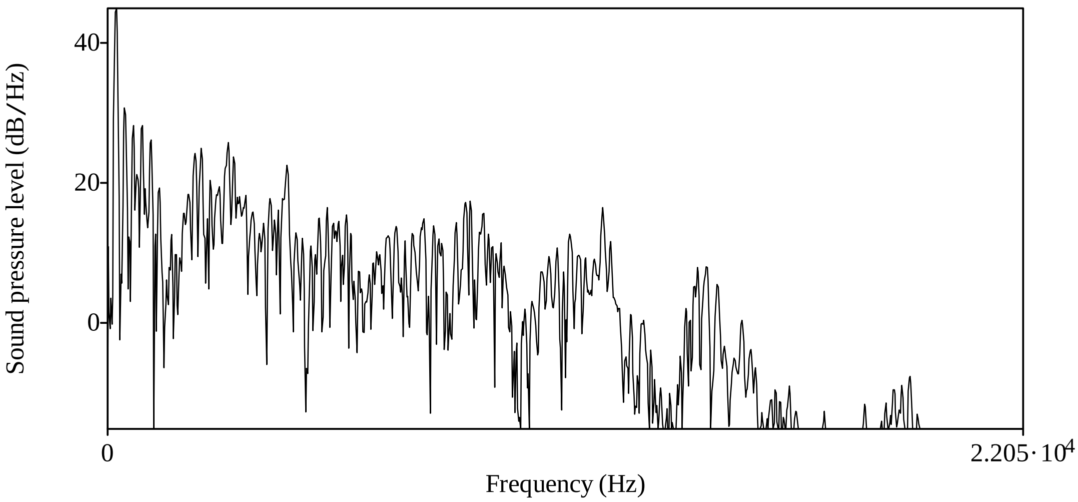


Figure 55. *Spectral slice of intervocalic[h]*

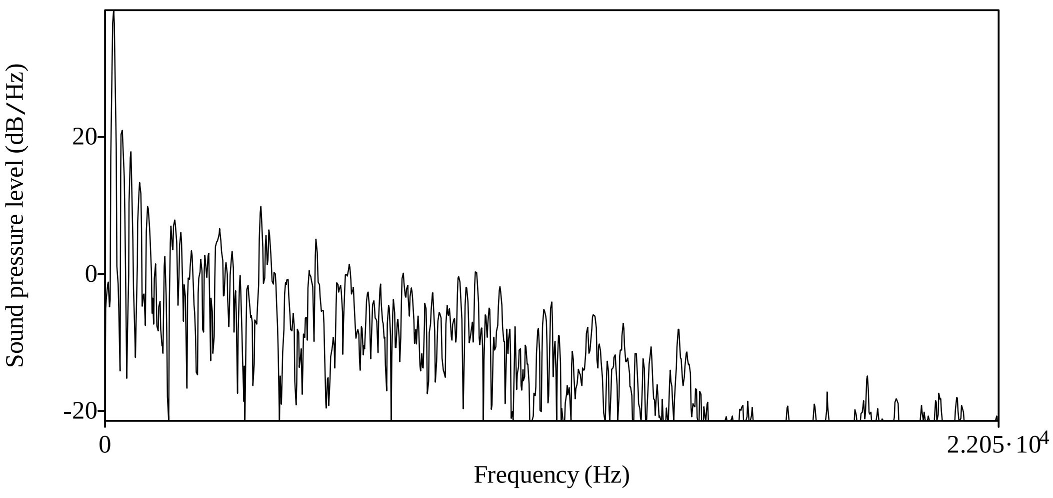


Figure 56. *Spectral slice of [v]*

APPENDIX K (continued)

SPECTRAL SLICES OF WAS FRICATIVES

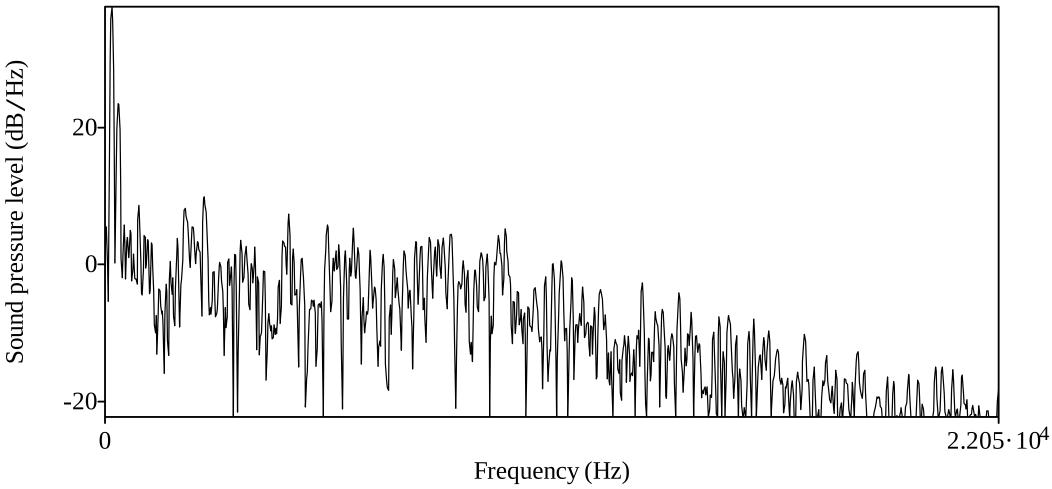


Figure 57. *Spectral slice of [ð]*

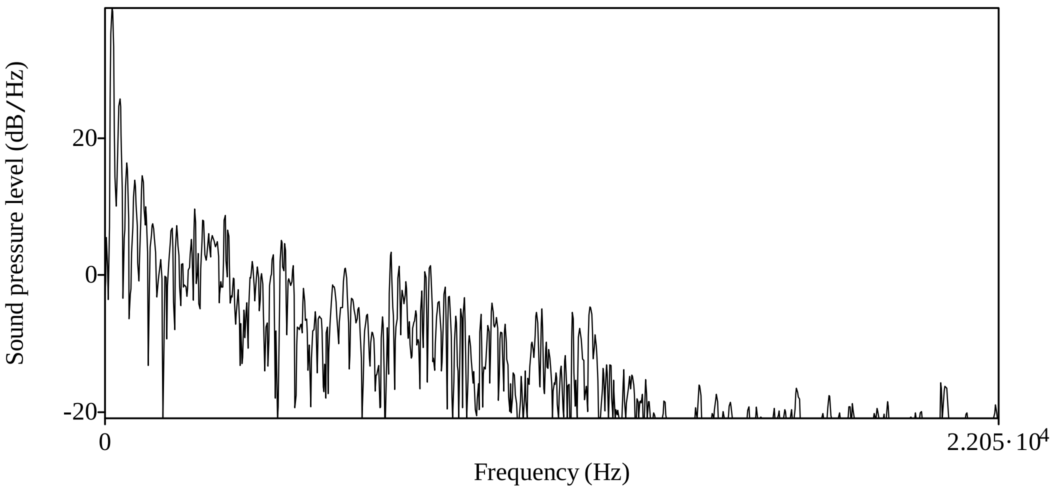


Figure 58. *Spectral slice of [x]*

APPENDIX L

SPECTRAL SLICES OF CS SIBILANTS

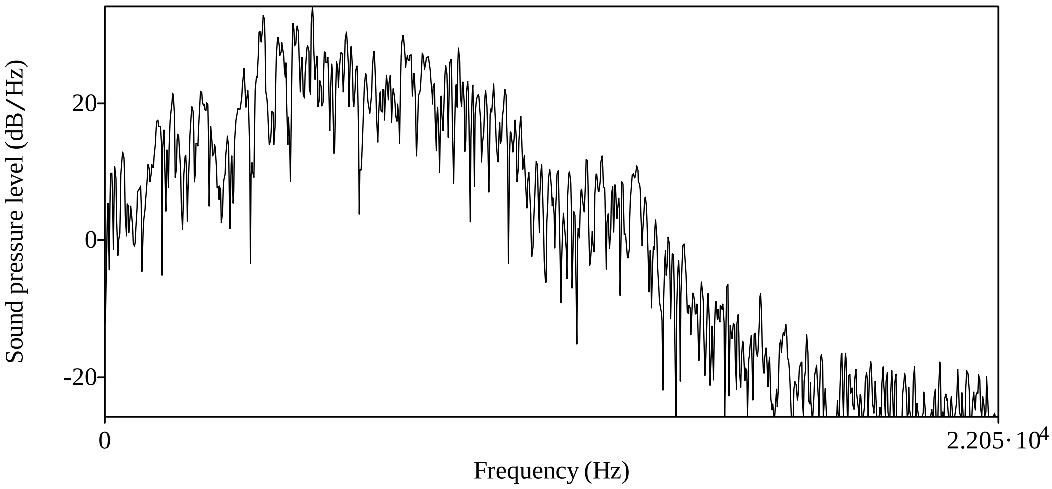


Figure 59. *Spectral slice of intervocalic [s]*

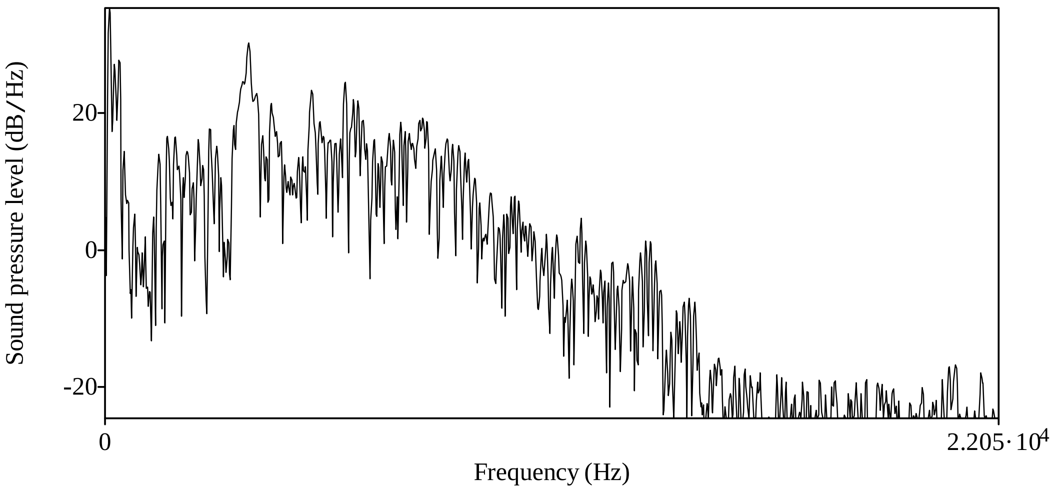


Figure 60. *Spectral slice of [z] before [ʂ̩]*

SPECTRAL SLICES OF CS SIBILANTS

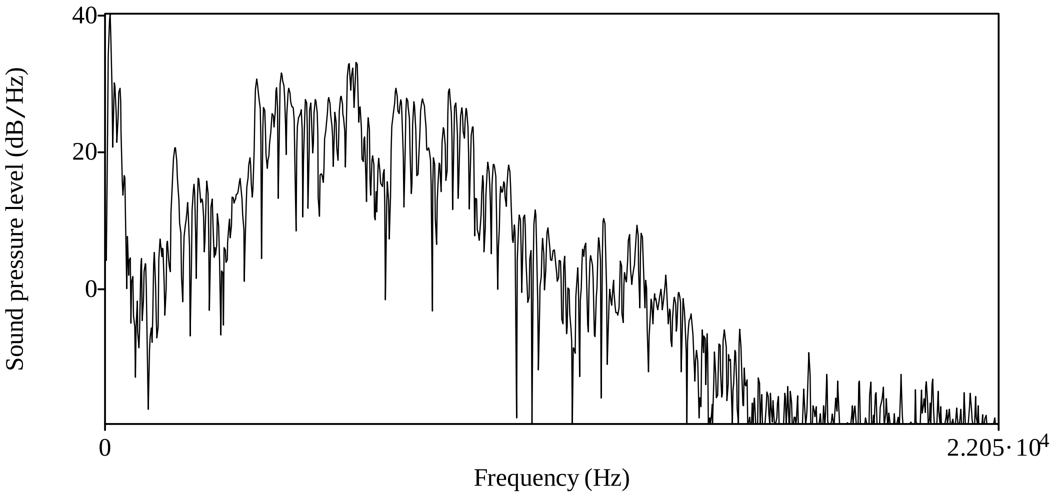


Figure 61. *Spectral slice of [z] before [ð̥]*

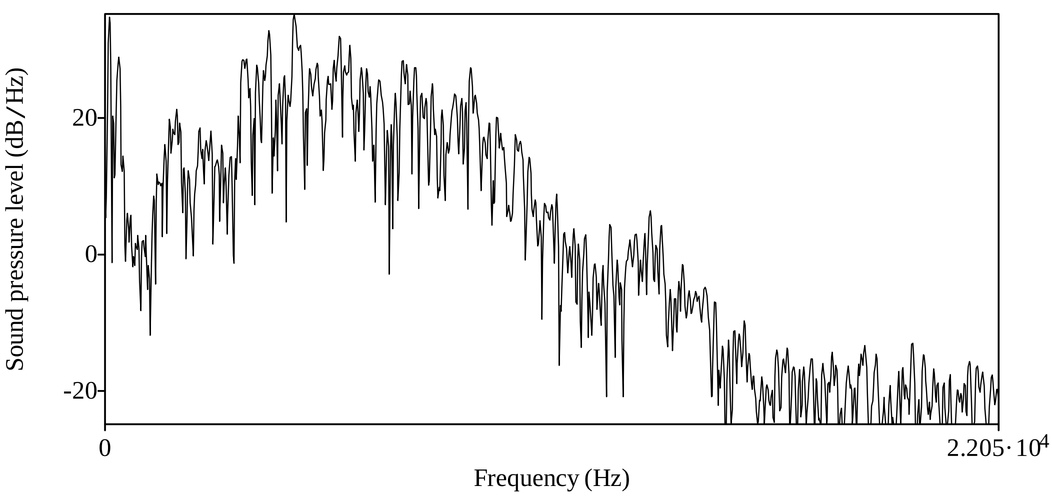


Figure 62. *Spectral slice of [s] before [ɣ̥]*