SEGMENTAL FOREIGN ACCENT

University of the Basque Country

(UPV/EHU)

Facultad de Letras

Rubén Pérez Ramón

supervised by

Prof. María Luisa García Lecumberri and Prof. Martin Cooke

2019 – University of the Basque Country

Abstract

The speech of non-native speakers typically differs from that of native speakers, resulting in a foreign accent (FA) which may have important consequences for communication. Traditionally, FA has been studied from a holistic point of view, which masks the impact of distinct linguistic components on the effects of FA. Little is known about the way individual foreign accented segments are perceived.

The main goal of this thesis is to assess segmental FA and its effect on listeners with different linguistic backgrounds in order to complement previous holistic studies and to provide new knowledge that can help improve the production of English as a foreign language. In order to study individual accented segments, this thesis presents a set of techniques whose goal is to generate English words in which only one segment has a Spanish FA. Additionally, a technique is introduced that allows the generation of various degrees of foreign accent. One of the main findings drawn from perceptual studies using the gradation technique is that the degree of acoustic deviation from the native prototype does not always correlate with perceived nativeness. To normalise across different segments, the gradation technique is applied to obtain continua with equal-sized steps of perceived nativeness.

Using such 'iso-nativeness' continua, the outcome of an extensive perceptual study of segmental foreign accent is presented. Listeners with very different linguistic backgrounds in terms of first language (English, Spanish, Czech and Japanese), assessed a range of non-native-to-native continua via a batch of tasks: intelligibility, perceived degree of foreign accent, native accent categorisation, AX discrimination and segment identification. The principal finding is that accented segments have an impact in the perception of non-native speech beyond that which a holistic methodology can detect. While some segments are deemed to possess a strong foreign accent without conveying a loss in intelligibility, other misrealisations can lead to communication difficulties for both native and non-native listeners. The phonetic representation of each segment and its foreign-accented counterpart in the phonological system of the listener play a crucial role in how segmental FA continua are perceived. Non-native listeners show difficulties in discriminating foreign accented words when the native segment and its non-native counterpart are part of the same phonological category in their system, but they exhibit native-like capabilities when it comes to bottom-up processing. Our results highlight the necessity of pursuing research on segmental foreign accent.

Findings are discussed in the light of previous research, and new contributions are highlighted. On the basis of experimental findings and methodological innovations, practical applications are suggested, alongside limitations and further studies required to achieve a deeper understanding of segmental foreign accent.

Acknowledgements

First of all, thanks to my supervisors, Professor María Luisa García Lecumberri and Professor Martin Cooke. Their knowledge, experience and insightful feedback have been priceless during the elaboration of this thesis. What is more, they are excellent hosts of what has been my home for the last years, Vitoria-Gasteiz, and I could never thank them enough for helping me feel welcome from the beginning.

While writing this thesis I have met other great researchers who have helped me and whom I have had the pleasure to work with. Thanks to Dr. Barra-Chicote for teaching me the basics of HMM synthesis in Madrid; to Professor King for receiving me in his Speech Synthesis course in Edinburgh; and to Professor Yamagishi for introducing me to DNN systems in Tokyo. I want to extend this acknowledgement to their laboratories and research groups: the Grupo de Tecnología del Habla of the Universidad Politécnica de Madrid (Spain), the CSTR of the University of Edinburgh (United Kingdom) and the NII Speech Group (Japan).

Thanks to Dr. Volín, Dr. Beinhoff, Professor Kondo and their wonderful teams for allowing and arranging my experiments on their respective labs with the best will: the Institute of Phonetics of the Charles University (Czech Republic), the Anglia Ruskin Research Centre for Intercultural and Multilingual Studies (United Kingdom) and the School of International Liberal Studies of the Waseda University (Japan). And thanks to all the participants of my experiments, both speakers and listeners.

This research has been partially funded by the grant BES-2013-062618, linked to the project DIACEX (FFI2012-31597, Ministerio de Economía y Competitividad), and by the Grupo de Investigación del Sistema Universitario Vasco IT904-16, Departamento de Educación, Gobierno Vasco. The international stays have been partially funded by the grants EEBB-I-15-09611 and EEBB-I-17-12505.

To all my colleagues at the University of the Basque Country, with whom days have gone easily amidst coffee breaks, long dinners and inspiring conversations: Ander, Asier, Aurora, Esther, Guillem, Idoia, Luis, Mátë, Olina, Sergio and many others who have come and gone. A special acknowledgement to Edurne Petrirena for her continuous support with the paperwork, and to Miguel Puente for being the "brother-in-arms" I needed in the hard times. A special thank you to the two people who introduced me to speech research, Juana Gil and Joaquim Llisterri.

Thanks to my brother, Diego, who taught me to always be the best version of myself, to wake every morning with my best smile and ready to give everything. And thanks to my partner, Irantzu, for jumping with me into this path and always, always supporting me without hesitation, in the good and in the bad times. If I could learn anything, it would be how to be more like them, and I hope someday I can give them back all I have received.

Finally, and above all, to my parents for always believing in me. Not a single one of these words would be possible without them. They are the giants on whose shoulders I stand.

Table of contents

Appendices

Chapter 1

Introduction

1.1 Motivation

Non-native speakers' pronunciation frequently differs from native speakers' productions. These misrealisations are interpreted as foreign accent (FA), and can have major implications for the communication process (Munro & Derwing, 1995a). A foreign accent might stem from a number of individual divergences, at the segmental and suprasegmental level, each of which can contribute in specific ways and in varying extent to the conveyance of FA. Nevertheless, in the main FA has been assessed holistically, that is to say, evaluated using an overall judgement on utterance length speech (Bradlow & Bent, 2002; Southwood & Flege, 1999, among others). In these types of studies, individual cues are generally glossed over and their specific contribution to FA is not described. However, understanding the relative role and weight of non-native speech characteristics is an important step towards successful communication amongst interlocutors with different L1s.

The relatively small amount of research on individual cues has primarily focused on suprasegmental features, including duration (Tajima, Port, & Dalby, 1997), speech rate (Munro & Derwing, 2001), nuclear stress (Hahn, 2004) as well as intonation and pauses (Kang, Rubin, & Pickering, 2010). The main advantage of targeting specific cues is that, in contrast to holistic studies, by manipulating a certain feature in a controlled way other variables are factored out and the observed results can be more robustly associated with the target cue.

One of the few attempts at measuring segmental FA was described by the study of Munro, Derwing, and Morton (2006) which analysed consonants' functional load. However, this study does not make firm conclusions regarding the role of various segments in foreign accented speech. The specific weight of individual mispronounced segments in the perception of foreign accent remains unknown. Early studies which have considered the segmental level with respect to the effect of the incidence of foreign accent (Anderson-Hsieh, Johnson, & Koehler, 1992, e.g.) used impressionistic evaluation methods, but technological developments have made it possible to measure accurately and objectively specific target cues. For instance, Sereno, Lammers, and Jongman (2015) presented two sets of English sentences (with and without Korean accent) and the same two sets with mutually interchanged intonation patterns to native English listeners via accentedness, comprehensibility and intelligibility tasks. Their results suggest that segmental information contributes to a larger extent than intonation to the emergence of foreign accent-related problems, but, other than this wide comparison, no insights about the specifics or the extent of segmental misrealisations in the perception of foreign accent are offered. No research, to our knowledge, has focused on the role of isolated segmental foreign accent or delivered a comprehensive analysis of the weight of individual segments in non-native speech. This is, precisely, the motivation for the current thesis: to assess the saliency and effect of segmental deviations in foreign accented speech devoid of higher order influences.

1.2 What is foreign accent?

Historically, foreign accent has received attention from a number of fields, and it has come a long way since its early description as a form of stammering speech (Greene & Wells, 1927; Griffen, 1980). Nowadays foreign accent is considered a "deviation from a standard pronunciation of a certain language and whose characteristics are reminiscent from [...] the native language of the speaker" (Silva & Barbosa, 2018). For the purpose of this thesis, foreign accented segments are defined as those articulated in such a way that they are perceived as deviating from the native norm by native listeners due to the interference of a speaker's different L1.

Several models of L2 perception (Best & Tyler, 2007; Flege, 1995) imply that, to a large extent, foreign accented realisations of segments may be the outcome of speakers' misperceptions. These models argue that such misperceptions are mostly due to the interference of the L1 of the speaker, that is, that the discriminability of native and non-native segments is deeply linked to the representation of such segments in the L1 and L2 phonological system of the non-native speaker. Kuhl (1993) argues that native phonological categories function as prototypes in our L1 phonological system. The perceptual space around a prototypical category shrinks as the category becomes more robust in the process of L1 acquisition, while, simultaneously, the perceptual space between categories is expanded. This view implies that discriminability inside what has been established as a category becomes more difficult for listeners. Accordingly, L2 sounds that fall within an L1 prototype's perceptual space will be difficult to discriminate from it and from each other.

Munro (1998) and Southwood and Flege (1999) highlight the fact that foreign accent is non-pathological speech and, additionally, is partially systematic. Major (2001) reckons that, from a general point of view, everyone is subject to the influence of other dialects or non-native realisations, this is, "everyone speaks an interlanguage" and Beinhoff (2013) states that "there is no such thing as not having an accent".

From these statements we can see that foreign accent is not an unequivocal entity that can be defined with a unique set of rules, but rather it is an aggregation of features, related to both production and perception, highly dependent on the linguistic background of the interlocutors, both speakers and listeners (Cho & Harris, 2006; Polyanskaya, 2015, e.g.). The amount and saliency of these features gives rise to different degrees of foreign accent.

In this thesis both speaker and listener roles will be considered. Most of the studies dedicated to foreign accent have analysed how non-native speech affects native listeners. Nevertheless, some researchers have put the emphasis on how non-native listeners with the same or different L1 from the non-native speaker perceive foreign accent (Bent & Bradlow, 2003; Stibbard & Lee, 2006, among others). This thesis examines how listeners from very different linguistic backgrounds perceive Spanish segmental foreign accent in English speech.

1.3 About this thesis

This thesis addresses the production of foreign accented segments and their consequences in the perception of listeners differing in their linguistic background. Target segments are tested mainly in carrier words, which means that acoustic manipulations are required in order to generate English words in which only one segment is pronounced with Spanish foreign accent. In chapter 3, three techniques for segmental foreign accent generation are introduced. Three cohorts of listeners differing in their L1 assess the intelligibility and degree of foreign accent of the generated tokens.

Chapter 4 introduces a new technique to generate acoustic degrees of foreign accent in consonants. The gradation technique allows the user to generate continua between a target native segment and its foreign accented counterpart in which steps are equally distant in terms of acoustic deviation. Six English consonants in initial position and their corresponding Spanish-accented realisations are selected, and 9-step continua are generated between each pair. Native cohorts and non-native cohorts with different levels of proficiency in English are presented with nativeness categorisation and AX discrimination tasks to determine whether segmental foreign accent is perceived categorically or if, on the other hand, acoustic deviations are directly correlated to perceived degree of foreign accent.

Chapter 5 refines the gradation technique, which is extended to vowels, and addresses the correlation between perceived nativeness and signal quality. In a mean opinion score task, native listeners are asked about distortions in tokens generated from the gradation technique, and their answers are correlated to the perceived nativeness of the generated continua.

In chapter 6 the gradation technique is extended with the goal of transforming the generated continua, comprised of steps equally-sized in terms of acoustic deviation from the native prototype, into iso-perceptual continua, made up of steps equally-sized in terms of perceived nativeness. Nativeness elicited in acoustic continua generated from four bilingual speakers is evaluated by native listeners and, based on these results, one speaker is selected for the iso-acoustic→iso-perceptual transformation of the continua.

In chapter 7 four groups of listeners with different L1s (English, Spanish, Czech and Japanese) evaluate segmental foreign accent in a set of iso-perceptual continua. With that purpose, four perceptual tasks are set up: syllable-isolated segment identification, word intelligibility, nativeness categorisation and AX discrimination. The results of this experiment offer insights about the implications of segmental foreign accent in listeners of different linguistic backgrounds, as well as about the saliency of particular segments as compared with others.

Finally, in chapter 8, the main findings and contributions of this thesis are presented. The limitations are also discussed, followed by a set of suggested practical applications of the knowledge acquired during this research. This chapter ends with some suggestions for potential future studies.

Chapter 2

Foreign accent framework

Several factors, both linguistic and non-linguistic, influence a successful communication process. In this chapter, foreign accent factors directly related to speech variables and their relationship are presented, and a framework for foreign accent is proposed (figure 2.1).

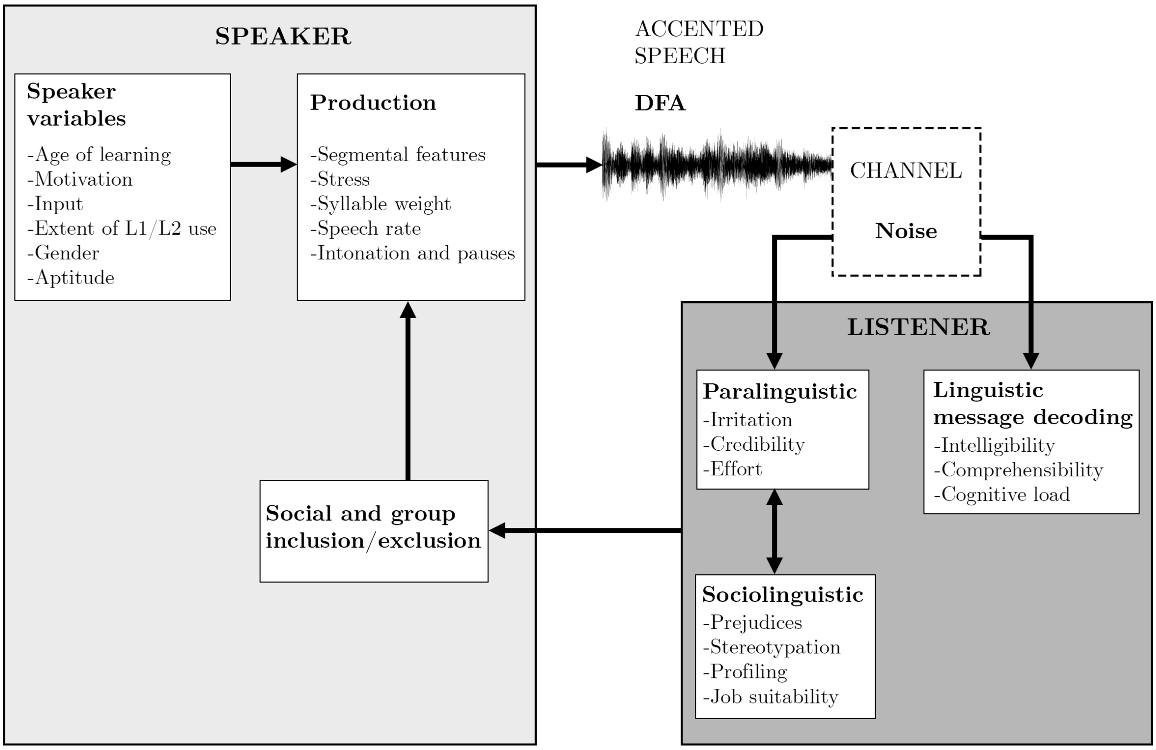


Figure 2.1: Proposed foreign accent communication framework.

There is a set of speaker variables that have been traditionally studied as correlated with the ability of the speaker to produce speech in a second language. These variables have a direct effect on the production of a second language, which, as explained in section 1.1, is formed from several traits that can individually contribute to the presence of a foreign accent, but to differing extents. The misproduction of such traits leads to a speech signal which conveys a certain degree of foreign accent (DFA), whose strength depends on the amount and saliency of the misproduced features. Other than these variables, which are related to the speaker, the signal can be further impaired by a transmission channel disturbed by background noise. Foreign accented speech can affect the ability of listeners to correctly decode the linguistic message. Even if the message is correctly decoded by listeners, they may experience paralinguistic and sociolinguistic effects. More generally, the attitude of listeners towards foreign accented speech can feed itself back to the speaker, leading to a sense of social and group inclusion/exclusion that can further affect their production. The way in which each of these factors affects communication is explored in the present chapter, along with a review of research devoted to the topic.

2.1 Speaker variables

One of the most studied variables relating to the speaker is the age of learning (AOL) of the L2, that is, the age at which learners start their instruction in a second language. According to Lenneberg (1967), there is a critical period ending around 12 years of age in which native-like language acquisition is possible. In principle, this hypothesis applies only to the learning of a first language, but it has also been widely studied with respect to L2 acquisition (Johnson & Newport, 1989; Long, 1990; Oyama, 1976; Singleton & Ryan, 2004). For instance Flege, Munro, and MacKay (1995) carried out an experiment to check the incidence of the AOL in the acquisition of a second language. As part of their experiment, they analysed 240 native Italians who had been living and learning English in Canada for 32 years on average, with AOLs spanning the range 2 to 23 years. Sentences produced by these speakers, among control sentences elicited by native English speakers, were rated by native English-speaking listeners with no training in any speech-related discipline. In their study, almost no native Italians were perceived as native when their English training started after they were 15 years old, which the researchers interpreted as a significant effect of AOL. Interestingly, in addition to AOL, other factors emerged as significant, such as length of residence in Canada and subjects' amount of use of the target language.

A more recent example of research on the effect of AOL for the acquisition of L2 pronunciation is provided by Huang and Jun (2014). In this study, Chinese speakers (children, adolescents and adults) of English varying in their AOL, along with native

English speakers, were rated by native and non-native listeners with different degrees of experience in foreign accent rating. The three groups of judges (inexperienced native English, experienced native English and advanced non-natives) agreed that non-native listeners who started learning English as children sounded significantly more native than adolescent and adult groups.

The most commonly-held view nowadays is that there is an age advantage rather than a critical period in the learning of L2s (Kuhl, 2000; Muñoz, Gallardo, & Llurda, 2000; Singleton & Ryan, 2004). On the other hand, other researchers consider that AOL *per se* is not a suitable measure because it can only be applied to immigrants who have both the opportunity and the need to use the L2 (Flege & Wayland, 2019). Instead, Flege (2018) suggests that quantity and quality of input is a much more reliable cue to predict the L2 abilities of a non-native speaker. In his study, four groups of Italians differing in their age of arrival in Canada and their amount of daily use of Italian completed discrimination and error detection tasks. Results indicate that Italians with an early age of arrival in Canada and a high use of Italian matched the discrimination capabilities of those with a late age of arrival but a low use of Italian. This was argued to be incompatible with the critical period hypothesis and implies that input plays a more important role than AOL.

The attitude or motivation of the speaker towards the learning process can influence production. In a set of studies carried out by Gardner and Lambert (Gardner & Lambert, 1959, 1965, 1972) a group of secondary school children showed motivation both in an instrumental way (a native-like accent is good for social advancement and better jobs) and in an integrative way (native speakers will more likely accept someone with a native-like accent). In Suter (1976) and Purcell and Suter (1980), the "Strength of concern" factor was found to be highly significant, this is, "the more concerned a speaker was about his pronunciation, the more accurate his pronunciation tended to be". In a recent study (Saito, Dewaele, & Hanzawa, 2017), motivated Japanese students of English as a foreign language showed a significant improvement in comprehensibility, but not in accentedness, after six months of instruction.

The extent of L1 use has also been identified as a relevant variable in the quality of L2 production. Suter (1976) classifies it as the third most significant factor – although it is worth noting that in a revision of this work (Purcell & Suter, 1980) it was no longer found to be significant. Flege, Frieda, and Nozawa (1997) tested the native accent of two different groups of native Italians living in Canada. Both groups presented a similar age of arrival in Canada, but differed in their self-reported daily use of Italian. This study showed that the extent of use of their first language had a significant impact on the accentedness perception when judged by a group of native English-speaking listeners. Both groups were evaluated as foreign accented, but the group that reported a higher use of Italian was rated with a stronger foreign accent than the other group. A replication of this study in which the two native Italian groups were also matched for age of bilingualism found similar results Piske and MacKay (1999). The outcomes of these studies suggest that the quantity of L1 use has a strong impact on L2 production accuracy.

Regarding the importance of gender on the accuracy of L2 production, in the previously mentioned study by Flege et al. (1995), it was found that females with an AOL lower than 12 years old were rated as significantly better on L2 pronunciation than males, but this trend changed when the AOL was higher than 16 years old, when females were perceived as significantly more accented than males. Ohara (2001) found that advanced native English learners of Japanese were aware of the use of a higher pitch by native females, unlike beginner learners of Japanese. Piske, MacKay, and Flege (2001) found no significant differences in terms of gender in Italian-English bilinguals, and suggested that no strong conclusions can be drawn from previous studies regarding this factor.

The findings of Piske et al. (2001) suggest that formal instruction is another factor which has no significant impact on foreign accent, meaning that learning a second language in a classroom or in other instructional settings is not beneficial for the acquisition of L2 pronunciation. Kissling (2014), on the other hand, found that English students of Spanish who had received formal instruction in Spanish phonetics had a slight perceptual advantage in a delayed post-test paradigm. In some cases, neither instruction nor input type can explain the performance of the speakers. Ioup, Boustagui, Tigi, and Moselle (1994) identified some native English speakers whose pronunciation of Arabic excelled, with no other explanation than an extraordinary linguistic aptitude. The level of their L2 performance was so outstanding that they were rated as natives by Arabic-speaking listeners, something extremely unusual.

2.2 Speakers' production

The variables summarized in section 2.1 have consequences in the physical domain, that is, they can affect the speech production of the non-native speaker. Several experimental studies have analysed manifestations of foreign accent in terms of the acoustic parameters that are altered in non-native speech.

One of the possible acoustic manifestations of foreign accent is the duration of individual segments. In Tajima et al. (1997), short English phrases were recorded from both a native English speaker and a Chinese speaker. These utterances were later modified in order to adjust the duration of the segments elicited by the non-native to that of the native and vice-versa. Results show that the intelligibility of non-native utterances improved when the duration was modified (39% vs. 58%). On the other hand, intelligibility of native sentences modified to match the segmental duration of non-native utterances decreased (94% vs. 83%).

Speech rate can also contribute to the perception of foreign accent, as demonstrated in Munro and Derwing (2001). In their study, several English sentences from Mandarin speakers were digitally modified to adjust them to the mean English rate, the mean Mandarin rate and a 10% reduced rate. Their findings suggest that native listeners and non-native listeners with a different L1 than the speaker tend to prefer slower a speech rate in accented utterances compared to native utterances, while non-native listeners with the same L1 of the speaker would rather listen to accented speech at the same rate as native speech. This finding suggests that non-native listeners sharing the same L1 of the speaker may able to overcome FA-related traits that are imposed on the L2.

In Polyanskaya, Ordin, and Busa (2016), speech rate and speech rhythm were compared in terms of perceived foreign accent. French speakers with different proficiency levels in English were asked to record a set of sentences. Afterwards, these sentences were modified to generate three sets of stimuli: (i) a set in which speech rate was equalised along the sentences; (ii) a set in which speech rhythm was equalised; and (iii) a set in which both features were equalised. Native English listeners were asked to rate these three sets along with the unmodified set of sentences for degree of foreign accent. Their results revealed that speech rhythm had a greater effect on foreign accent perception than speech rate. A later study by Kang et al. (2010) examined the impact of non-native speech in several suprasegmental features, namely speech rate, intonation and pauses. For each of these features, a number of measures were taken instrumentally, as opposed to being rated by native judges. Analysis of sentences pronounced by TOEFL students suggested that the combination of the three suprasegmental features are the reason for up to 50% of the ratings in the measurement of comprehensibility and the variance in oral proficiency.

Hahn (2004) analysed how the misplacement or disappearance of a nuclear stress affected the perception of foreign accent. In her study, non-native speakers were rated significantly more favourably (i.e. as less foreign accented) when the nuclear stress was correctly placed, and the discourse was generally more easily processed by native listeners.

Syllable patterns have also been shown to be a reliable source for native listeners to identify the intended words of non-native speakers. Zielinski (2008) analysed transcriptions made by native English listeners of utterances in English produced by Korean, Chinese and Vietnamese L1 speakers. Her findings suggest that non-standard patterns of weak and strong syllables lead to the misperception of words.

2.3 Channel

If the environment in which speech is transmitted is not clear enough and the speech signal is to some extent masked by noise, speech communication may be affected. The appearance of a foreign accent in such a context may lead to further obstructions in the flow of communication. García Lecumberri, Cooke, and Cutler (2010) offer an extensive review of the literature on non-native perception of speech under adverse conditions, focusing on three factors: the task presented to the listeners, the effect of adverse conditions and differences among listener populations.

Munro (1998) presented a study in which the effect of cafeteria noise was tested in sentence-verification and sentence-transcription tasks. Other than native English elicitations, Mandarin-accented speech was also assessed by native English listeners. Munro found that the addition of noise resulted in significantly lower verification scores for Mandarin-accented utterances than for native English utterances, and suggested that further experimentation involving different types or levels of noise and accent types was required in order to draw stronger conclusions regarding the relationship between noise and foreign accent. Even so, this study raised the notion that noise can affect the intelligibility of foreign accented speech to a greater extent than it affects native speech.

In van Wijngaarden, Steeneken, and Houtgast (2002), Dutch-speaking listeners with English as L2 and German as L3 were presented with utterances in English and German, either with or without a Dutch foreign accent. Sentences were presented with different speech-to-noise ratios (SNR) to assess speech reception threshold (SRT) differences between native and non-native listeners. SRT measures the SNR at which 50% of words in the tested sentences are perceived correctly. Among other findings, their results indicated that listeners required a 1 − 7 dB better SNR to obtain the same SRT in sentences with a foreign accent compared to sentences without a foreign accent.

In a study by C. Rogers, Dalby, and Nishi (2004), monolingual American English-speaking listeners were presented with utterances elicited by native American English speakers and Mandarin-accented English speakers with high and low proficiency in English. These utterances were presented to listeners in quiet and in three different multi-talker babble SNR (+10 dB, 0 dB and -5 dB). Participants were asked to write down the words they heard in an intelligibility task. Adding noise to the signal was found to drastically reduce the intelligibility of non-native utterances elicited from highly proficient non-native speakers, who were perceived as almost native in quiet conditions, with intelligibility reducing to the level of low-proficiency speakers.

When the channel is not optimal, speakers usually modify their productions in order to make themselves more intelligible. This kind of modification, known as Lombard speech, was first described in Lombard (1911). Lombard speech has been compared with foreign-directed speech, a speech mode that enhances intelligibility. Sankowska, García Lecumberri, and Cooke (2011) suggest that Lombard speech may be more intelligible in noisy environments than foreigner-directed speech, but not in quiet.

2.4 Linguistic message decoding

The presence and degree of foreign accent can lead to problems in the decoding of the received linguistic message. The most direct way of measuring foreign accent is through subjective scales, typically with ranges of 7 or 9 points. In Flege et al. (1995), native speakers of English used a continuous scale ranging from 0 to 255 to represent the perceived accentedness, 0 meaning the strongest foreign accent and 255 meaning no foreign accent at all. This kind of measure has the advantage that it is methodologically quite easy to implement, but it has the deficiency that it requires an understanding of what *foreign accent* is, and this may differ from listener to listener. *Degree of foreign accent*, or DFA, is usually related to *Intelligibility*, which is defined in Nelson (1982) as the understanding of the message in the sense intended by the speaker. This definition is similar to the one provided in Munro and Derwing (1995b) (the extent to which an utterance is actually understood) and others. Intelligibility is usually measured by asking listeners to write down what they have heard in tasks where tokens are words, sentences or other utterances. It is an objective way to measure the impact of foreign accent in the sense that what is measured are not listeners' subjective impressions, but rather their ability to successfully decode utterances. Munro (2008) stresses the importance of differentiating between accentedness and intelligibility. The author mentions the judicial case of Mirek Gajecki v. Board of Trustees, School District No. 36 (Surrey) (1990), where, as stated by the court, a non-native teacher was rejected as a teacher candidate simply because of his accent, and not due to communication problems with the students. This is a clear case in which more research on the distinction between accent and intelligibility could have clarified key issues.

A dimension related to intelligibility is *comprehensibility.* As stated by Smith and Nelson (1985), comprehensibility and intelligibility have traditionally been interchangeable concepts, so they propose a distinction: Intelligibility is an objective measure of the ability to recognise utterances, while comprehensibility is the subjective judgement of the difficulty of understanding those utterances. A word or sentence, therefore, can be totally intelligible but, at the same time, hard to comprehend. Field addressed the difference between comprehensibility and intelligibility as a "division between local formal recognition and global processing effort". Comprehensibility has been found to be more correlated to intelligibility than to accentedness (Munro & Derwing, 1995a). Derwing (2018) offers an in-depth review of the concept of comprehensibility, and suggests that an utterance can be perceived with a strong foreign accent, but that does not necessarily make it harder to understand.

Reaction time has been used as an empirically-measurable variable to address comprehensibility and cognitive load under the assumption that processes that are more demanding in terms of cognitive resources are associated with longer response times (Borsky, Tuller, & Shapiro, 1998; Jensen & Thøgersen, 2017; Weil, 2003). However, Skarnitzl and Šturm (2016) found no correlation between reaction times and comprehensibility, and suggest that familiarity with the tokens and their collocations were more strongly correlated with comprehensibility than reaction time.

2.5 Paralinguistic and sociolinguistic consequences of foreign accent

Since AOL has been demonstrated to be an important variable in terms of the production of foreign accent, the age of the receiver (i.e. listener) has also been studied as a factor in the perception of foreign accent. Burda, Hageman, Scherz, and Edwards (2003) presented a pilot study in which three cohorts differing in age range (20-39, 40-59 and 60+) were exposed to words and phrases in English pronounced by native speakers of English, Taiwanese speakers and Spanish speakers. By means of intelligibility and comprehensibility tasks, Burda et al. found that the intelligibility scores of older adults were significantly lower than those of younger cohorts. The authors suggested three possible explanations: (i) First, a physiological change during aging that could lead to general poorer speech perception or an increment of the cognitive load of speech processing; (ii) Second, other health problems non-related to neurological or hearing disorders, such as discomfort (e.g. arthritis) or altered states due to medication; (iii) Finally, a problem with short-term memory, which may have had an effect on their responses due to the response modality (listeners were asked to write their answers instead of speaking them out loud). Even though the overall performance of the older group was significantly lower, the interaction between the native language of the talker and the age of the group was non-significant, meaning that the presence of a foreign accent lowered intelligibility among the three groups to roughly the same extent.

Ferguson, Jongman, Sereno, and Keum (2010) inquired into the relationship between hearing loss and foreign accent. Three cohorts were tested: young adults without hearing loss and older adults with and without hearing loss. Stimuli were words produced by a native speaker of American English and a native speaker of Spanish, presented in quiet, filtered to mimic the phone speech signal. Additionally, the signal was distorted with a background of 12-talker babble at several signal-to-babble ratios. Older adults were significantly more affected by background noise than younger adults, but, as found in Burda et al. (2003), all listeners were proportionately affected by the presence of foreign accent, regardless of their hearing capacities.

Potential reductions in intelligibility and comprehensibility brought about by a foreign accent can also lead to distraction and irritation on the part of listeners, whether natives or non-natives. Fayer and Krasinski (1987) found that English utterances produced by Puerto Rican native Spanish speakers with different degrees of proficiency were rated as linguistically worse by Puerto Rican listeners, but not by American listeners. Puerto Rican listeners also reported a higher degree of irritation, even though the scores for intelligibility were similar for both groups. These results suggest that, in some cases, non-native listeners sharing the same L1 as the speaker are less tolerant of foreign accent than native listeners. The reasons for this, the authors argue, are not completely clear. The concept of *irritation* was broken down into two other concepts: *distraction* and *annoyance*. In the distraction dimension, both groups rated speakers similarly, but Spanish listeners were more annoyed. The intuitive idea that intelligibility correlates negatively with irritation, therefore, does not apply in this case, as the Spanish group should have shown significantly less intelligibility, which was not the case. The authors hypothesize that once irritation has reached a certain level it is no longer related to intelligibility. The familiarity factor was also addressed as a possible cause for these results. Spanish listeners were more likely to feel familiar with the situation of the Puerto Rican speakers, and "familiarity with the topic of discourse greatly facilitates comprehension" (Gass & Varonis, 1984). This being the case, it may be that even when annoyance is high, familiarity with the context of fellow non-natives allows listeners to reach higher scores in intelligibility. In Chapter 3 we will analyse in detail the connections between a shared L1 and foreign accent intelligibility.

In del Puerto, Lacabex, and García Lecumberri (2009), Content and Language Integrated Learning (CLIL) instruction methods were evaluated for pronunciation in comparison with traditional classroom teaching. Three dimensions were tested: intelligibility, irritation and degree of foreign accent. One of the main findings was that degree of foreign accent was inversely correlated with intelligibility and directly correlated with irritation. Furthermore, correlation was stronger between intelligibility and irritation than between degree of foreign accent and either of the other two variables. The authors suggest that it may be that when listeners assess the communicative effects of foreign accent (intelligibility or irritation), their judgements are affected by other factors such as fluency (Anderson-Hsieh et al., 1992) and grammar (Varonis & Gass, 1982). In a follow-up study (del Puerto, García Lecumberri, & Lacabex, 2014) this correlation was further explored by comparing the judgements of naïve native listeners and experienced non-native listeners. The authors found that the better comprehensibility results presented by non-native judges did not result in lower irritation scores.

One of the most common situations in which foreign accent becomes part of everyday life is when speakers move to a new country and have to communicate in a different language. It is not surprising, therefore, that one of the main concerns of a non-native talker is job suitability. Boyd (2003) analysed the reaction of school principals, pupils and other academic evaluators towards foreign-born teachers in Sweden. More specifically, she focused on whether listeners considered that the talkers were suitable to teach in Sweden or not. One of the outcomes of this investigation was that the degree of perceived foreign accent was directly correlated with the impressions about the suitability to perform as a teacher. The author claims that, when asked, respondents considered that problems in a classroom with a native teacher were due to “too little experience”, while problems in a classroom with a non-native teacher were consistently associated with inadequate proficiency in Swedish.

A foreign accent may have consequences far beyond job suitability. Lev-Ari and Keysar (2010) investigated the impact of a mild and strong foreign accent on speaker credibility for the listener. Their results showed that non-native speech was rated as less truthful than native speech, regardless of foreign accent strength. In a second experiment, researchers informed the participants about the nature of the task they were taking part in. In this case, listeners stopped perceiving mild accent as an untruthful source, while a strong foreign accented speech was still significantly considered as not truthful. Credibility can be, therefore, one of the biggest problems a non-native speaker has to face.

This predisposition to assume certain patterns of behaviour with just the auditory input of the speaker is known as “linguistic profiling”. The term was coined by John Baugh in Baugh (2003) in comparison to the so called “racial profiling”, related to the judgements made by the looks of a certain subject. Early evidence for linguistic profiling were detected when fair housing advocates reported several rental application denials due to the voice of the interlocutor through a phone call. (“Sounds Mexican” or similar statements were reported by house owners.) Linguistic profiling is not only associated with nativeness, but also to other characteristics such as sex or age and can also be related to regional or ethnic variation. Baugh states that linguistic profiling can be both discriminatory or preferential, and focuses on the case of African-Americans in the United States of America and the legal paradox resulting from *black-like voices.* Linguistic profiling, which may be stronger or milder depending on listeners’ socio-cognitive background, makes the presence of a foreign accent a sufficient condition for a substantial impact in the everyday life of a person residing in a place with a different standard accent.

2.6 Social and group inclusion / exclusion

The attitudes of listeners towards non-native speakers, either as individuals or as a group, can have a big impact on the attitude of the speaker. Zuengler (1988) presented a study in which native Spanish speakers were asked to mimic Americans speaking Spanish. Their utterances were recorded and analysed, and some features were consistently swapped (e.g. the Spanish [r] was changed to the American [ɹ], or the differentiation between [b] and [v] was marked). The author suggests that these specific features are tied to the perceived identity of the speaker, and are socially marked more swiftly than others. A speaker presenting this kind of realisation might feel, therefore, detached from the listeners' social group. In the words of the author, "pronunciation is a domain within which one's identity is expressed". Nevertheless, some non-native speakers are reluctant to lose their foreign accent, as they feel that better pronunciation implies a detachment from their own identity (Lybeck, 2002).

Derwing (2003) conducted a study in which non-native speakers of English in Canada were interviewed and questioned about their perception of their own accent. Her results showed that students perceived their problems to be related to segmental mispronunciations in roughly 79% of cases, and almost all participants (97%) considered it crucial to have a good pronunciation in English. A majority of participants (95%) also reported their wish to pronounce English as native speakers. The interviewer asked them if they would not feel identity loss in the assumption that they would acquire a totally native accent in English, and their response was mainly that identity is linked to their L1, and therefore would not be threatened by a perfect L2 pronunciation. The yearning for a perfect accent was not academic in nature, but practical. Most of the participants expressed the view that with better English pronunciation, Canadians would treat them better. Some of the responses highlighted problems in everyday life, both from a general point of view ("They don't pay attention to you if your English isn't good") to specific contexts ("Waiters make faces in restaurants"), and including situations that are non-trivial ("A doctor got angry with me - it hurt me"). Note that the reported reactions often made reference to the accent and not to other linguistic features such as grammar ("I ask bus driver [a question]. When he hear me in a different accent, he just say "no." He doesn't even look at me or say or explain to me which bus to take. He looked at me in a miserable way and say *No!*").

Summary

In the light of the reviewed literature, it can be seen that foreign accent may arise from several conditioning factors and lead to unintended consequences. Even when other linguistic components (e.g. syntax) are native-like, the presence of a foreign accent is a sufficient condition to affect the communicative experience. As mentioned in chapter 1, little is known about the specific weight of foreign accented segments on the communication framework. The remainder of the thesis explores the issue of the segmental basis of foreign accent in depth.

Chapter 3

Intelligibility and degree of foreign accent at the segmental level

3.1 Introduction

FA-related effects at the segmental level remain mostly unidentified. However, in García Lecumberri, Barra-Chicote, Pérez-Ramón, Yamagishi, and Cooke (2014) (henceforth GL14), isolated segments of English words were replaced with their Spanish-accented counterparts through a number of techniques (e.g. the segment [h] was replaced with [x] in the word *house*, while the rest of the word remained unaltered). Two cohorts of listeners, one native (English) and one non-native (Spanish) underwent an experiment divided in two parts. First, in an intelligibility task, participants were asked to type native words and their accented counterparts orthographically . Afterwards, they were asked to report the perceived degree of foreign accent (DFA) over the same set of words in a 7-point Likert scale labelled as "strength of foreign accent", echoing Munro et al. (2006).

Intelligibility tasks have been widely used in the assessment of foreign accent (Munro, 2008; Munro & Derwing, 1995a; Trofimovich & Isaacs, 2012). In Munro and Derwing (1995a) intelligibility is broadly defined as the extent to which a speaker's message is actually understood by a listener. Usually, researchers use sentences and assign scores based on the deviations of listeners' responses from what was said(Munro & Derwing, 1995a, and many others), but some authors measure deviations solely in selected key words from utterances (Pinet, Iverson, & Huckvale, 2011), in isolated words (Imai, Walley, & Flege, 2005) or, as it is the case of our study, segments (GL14).

Research has broadly focused on the relationship between the target language, the L1 of the speaker and the L1 of the listener. Some studies have pointed out that native listeners tend to better perceive native speech than foreign accented speech (Bent & Bradlow, 2003; Munro, 1998; Munro & Derwing, 1995a; van Wijngaarden et al., 2002), but many others note that non-native speech produced by high proficiency speakers may be just as intelligible (Major, Fitzmaurice, Bunta, & Balasubramanian, 2002; Stibbard & Lee, 2006). Competence in the target language, therefore, also seems to be an important factor to take into account when considering foreign accent production.

In Bent and Bradlow (2003) the so-called *matched interlanguage speech intelligibility benefit* is described as the lesser drop in intelligibility by non-native listeners when exposed to speech of speakers with their same L1 with respect to native listeners. As an extension of this effect, the authors remark that non-native listeners also suffer a smaller or equal deficit in intelligibility when exposed to non-native speech by speakers of different L1 with respect to native listeners. These researchers refer to this effect as the *mismatched interlanguage speech intelligibility benefit*.

Some studies have expressed disagreement with the findings of Bent and Bradlow (2003). In Munro et al. (2006), only marginal evidence is found that a shared L1 between speaker and listener results in better understanding. Again, in Stibbard and Lee (2006), authors find that non-sharing L1 groups do not enjoy a benefit in intelligibility and that, in the case of low-proficiency speakers, a *detriment* in intelligibility can be observed among non-native listeners (an effect also found in Shin, 2018). Furthermore, Pinet et al. (2011) point out that phonological knowledge does not directly affect intelligibility, but rather a listener's accent similarity with the perceived speech (i.e. production becomes linked, to some extent, to underlying phonological processes used in perception). Wang and van Heuven (2015) suggest that the results of Bent and Bradlow (2003) should be formulated in relative rather than in absolute terms.

Another hypothesis, referred to as the *phonological mismatch hypothesis* and introduced in Imai et al. (2005), posits that phonetic representations are critical. This hypothesis proposes that non-native listeners' phonetic representations should optimize intelligibility when the production matches such representations. This idea brings a new twist to the matched/mismatched interlanguage speech intelligibility benefit hypothesis in the sense that the language of the listeners is no longer the reference point, instead shifting the focus to their phonological system and competence. Indeed, different variants of the same language may lead to mismatched phonetic representations between speaker and listener, while some speakers with different L1 may concur in matched phonetic representations.

Summarizing previous studies, it is expected that native listeners experience a high degree of intelligibility both from other native speakers' and from highly proficient non-native speakers' production. Similarly, non-native listeners with high competence in the target language will also benefit from native or highly proficient speakers' production, regardless of the interlanguage matching. On the contrary, non-native listeners with low proficiency in the target language will suffer a detriment from highly proficient speakers' production (either native or non-native), but will find the production of a low-competent speaker more intelligible.

An important issue to bear in mind is that all these studies follow a holistic strategy, as mentioned above. Speakers are classified according to their general experience with the target language (i.e. experienced or inexperienced talkers, high/low proficiency talkers, etc.), but specific variables are not controlled by the researchers, as it would be quite hard to gather a set of speakers with similar speech rates, F0, intonation, etc., differing only in their experience in the target language. Speakers' phonetic competence is not assessed directly, so there may be intrinsic interspeaker intelligibility differences.

It is possible that these uncontrolled variables or other higher-order effects such as grammatical structures, lexical knowledge and the like led to confounds which may underlie any discrepancies found in the outcomes of the above-mentioned studies. As opposed to these holistic testing methods, the current thesis (along with GL14) adopts a segmental approach supported by methodological advances in the speech-manipulation field. By focusing on modifications at the segmental level, these confounds remain controlled, so the outcome of the experiments can be attributed to the segmental level with greater certainty, and moreover linked to individual segments.

Our main goal is, therefore, to test how the interlanguage effects outlined above apply at the segmental level using controlled generated foreign accent. Derwing and Munro (1997) point out the quasi-independence of perceived accentedness and intelligibility, so a segmental-level exploration of this relationship is also relevant. In GL14, both listener cohorts scored higher in intelligibility of non-foreign tokens than in words with a foreign accented segment. Their findings also pointed to a matched interlanguage speech intelligibility benefit in the sense that the loss of intelligibility was less pronounced among Spanish listeners. Also, the results of the DFA task suggested that some segments had more impact than others in perceived accentedness. GL14 focused on comparing different methods for generating Spanish foreign accented English segments, and evaluating the effect of segmental FA on Spanish and English listeners. Their dataset is used as the starting point for the current study. However, in order to properly gauge interlanguage effects, we augment the GL14 dataset with responses from a new listener cohort whose L1 differs from both that of the target items and the accent. Specifically, native listeners of Czech undertook the same tasks as the Spanish and English cohorts of GL14. This new set of responses will provide data about the general variability of intelligibility in accented speech and, critically, about the relation between accentedness, intelligibility and speaker/listener L1s.

The research questions explored here are: (1) Following on previous studies on the link between DFA and intelligibility conducted with a holistic methodology, is there a relationship between intelligibility and perceived accentedness at the segmental level? (2) Given the discrepancies in previous studies regarding the possible (mis)matched interlanguage intelligibility benefit, is there a matched/mismatched interlanguage intelligibility benefit at the segmental level? If so, does it apply to all segments equally?

3.2 Computational segment manipulation

In this chapter acoustic manipulation will be extensively used in order to generate segmental foreign accent. Some of the first attempts at segmental manipulation go back to mechanical devices designed with the purpose of mimicking speech sounds through keys, pedals, tubes or other apparatuses (Flanagan, 1972; Kempelen, 1791). The first electrical speech synthesizer, named Voder, was presented at the World's Fair of 1939, in New York, and it is considered to be the first research, non-recreational speech synthesis device (Dudley, Riesz, & Watkins, 1939). The Pattern Playback machine (Stella, 1985) set a milestone in the field of segmental manipulation. This device allowed the user to synthesize sounds by simply drawing formant patterns, which enabled more precise manipulations.

Technological advances have fostered techniques that allow more natural and accurate manipulation of segments. Some of the first systems that facilitated speech modeling through machines were based on concatenative generation, that is, generation of speech from the concatenation of units such as individual phonemes, diphones, etc. Concatenative generation systems such as the Festival software (A. J. Clark, Richmond, & King, 2004) generate high quality output, but it has the downside that it requires a huge amount of storage space. In order to address this problem, Tokuda, Yoshimura, Masuko, Kobayashi, and Kitamura (2000) introduced the Hidden Markov Model-based Speech Synthesis (HTS), a software tool for statistical parametric synthesis that generates speech models from large sets of annotated recordings using Hidden Markov Models (HMM). Once the model is generated, new utterances can be synthesised by calculating the properties of the required utterance from the data stored in the model. More recently, speech recognition and synthesis have improved by using deep neural networks (DNN, Hinton et al., 2012). In Lazaridis, Potard, and Garner (2015), a group of listeners showed a preference for DNN-based systems over HMM-based systems, even with a small amount of training data (1.5h).

In GL14, HTS and concatenative speech techniques were adapted in order to generate words in which only one segment was accented. The implementation of these techniques for segmental foreign accent research is detailed in section 3.3.

3.3 Methods

Participants

In addition to the two cohorts whose results were already in the GL14 dataset (English, n=9; Spanish, n=21), a new cohort of Czech listeners (n=33) was recruited with a reported similar level of competence in English as the Spanish group of GL14. Recruitment criteria included the requirement that participants were not advanced in any language other than the required level of English and the condition of not having had contact with Spanish speakers, even in other languages different than Spanish.

Corpus

The corpus of GL14 consisted of 108 words for which native and foreign accented versions were generated through three different techniques, amounting to a total of 648 tokens (108 words×3 techniques×2 accents). The 108 words were distributed as follows: 9 English segments ([h], [ɹ], [kʰ], [tʰ], [v], [ð], [w], [ʤ], [j]) were correlated with their Spanish accented realisations in initial position ([x], [r], [k], [t], [b], [d], [gw], [j], [ʤ] consecutively) and in medial position ([x], [r], [k], [t], [β̞], [ð̞], [ɣ̞w], [j], [ʤ] consecutively). In the Czech phonological system, voiceless plosives are not aspirated, so [k] and [t] are closer to the Spanish system. Their rhotics system is also closer to the Spanish, with a trill /r/ which is usually realised as [ɾ] but can also be pronounced [r]. In Czech there is no voiceless glottal fricative [h], but they do distinguish at the phonemic level the voiceless velar fricative [x] and the voiced glottal fricative [ɦ]. Both phones form a kind of a "false pair" in terms of voicing (e.g. *God* is pronounced [buːx], but [boɦa] in genitive). In Czech, [ʤ] is not a realisation of the phoneme /j/, as it can be in Spanish. That is why [ʤ] has recently started to be considered as a new phoneme, used mainly in loan-words. As in English, therefore, [ʤ] and [j] are distinguished at the phonemic level. Similarly, [b] and [v] are also different phonemes in Czech, but not in Spanish. [w] is historically related to [v] (Romportl, 1973), but it has been developing into a voiced labiodental fricative for several centuries. Finally, [ð] is also not in the Czech phonological system.

The target segment of each word was placed either in initial position (88 words) or medial position (20 words). A high degree of variability related to coarticulation effects in medial position was detected in GL14.

Items in the corpus of GL14 were generated through three different techniques: Code-switching, where the speaker changes at will the accent of only one segment in an utterance, while keeping the remaining segments and suprasegments unaltered (e.g. pronouncing [**r**eɪnbəʊ] instead of [**ɹ**eɪnbəʊ]); Splicing, in which the target segment of the unaccented English word is replaced with its Spanish counterpart by means of computational sound wave manipulation (e.g. the [ɹ] segment in [ɹeɪnbəʊ] is replaced with a [r] segment); and statistic speech synthesis, where two models, one for Spanish and the other for English, are trained from the bilingual speaker and combined in order to mix features of the two languages (e.g. the word *rainbow* is synthesized from the learnt features of Spanish [r] and combined, prior to speech generation, with those of English [eɪnbəʊ]).

According to the results of GL14, synthesis proved to be poorly fitted to segmental foreign accent generation. On the other hand, code-switching has the drawback of providing little control over the changes introduced in the word.

Tasks

For comparability, Czech listeners underwent the same experiment presented to English and Spanish cohorts in GL14. The experiment was divided into two parts, carried out in separate 45-minute sessions. The first session consisted of an intelligibility task, in which participants were asked to type the presented word in a Matlab-based interface. In the second task, listeners were asked to rate each token by selecting a number in an integer 7-point Likert scale labelled as "strength of foreign accent". The endpoints of this scale were labelled as "(1) Native-like" and "(7) Very strong" (Munro et al., 2006). During this task, participants were able to see the word orthographically written on the screen. The interface was also produced using Matlab.

Tokens from each one of the three techniques (code-switching, splicing and synthetic) were blocked and presented counterbalanced to the participants. Native and foreign tokens were mixed within each block, and stimuli were randomised. The experiment was carried out in a room with similar quiet conditions as the one described in GL14 for English and Spanish cohorts. The Czech cohort undertook the experiment at Charles University in Prague. Participants were paid after completion of the experiment.

In this study, and in the light of the outcome of GL14, only a subset of the collected results are analysed. Specifically, only the results for words in which the foreign accented segment was in the initial position and generated through the splicing technique are analysed. Therefore, the final dataset of our study is made up of responses to 176 tokens (88 words × 2 accents) from each participant in the three cohorts (English, Spanish and Czech).

3.4 Results

This section reports on results for the intelligibility and DFA tasks. First, each task will be analysed separately. Next, the results of both tasks will be compared in order to check for a possible relationship between the two factors.

As the main goal of this study is to examine segmental foreign accent, in the intelligibility task only accuracy regarding the target segment was considered. This means that, e.g., for the word *told*, responses like *tool* were counted as correct, as the target segment (i.e. [t]) was correctly identified. Correct answers were detected by comparing the phonetic transcriptions of the target word and the input of the participants. Transcriptions were extracted from the BEEP dictionary (Beep, 2017).

A detailed analysis of the results revealed that listeners attempt to make sense of the perceived segments, even if accented, over the the fact that a foreign accented segment could convey a minimal pair in English with the native word. This means that, even though some words generated minimal pairs via the native and foreign accented forms (e.g. *van-ban*), this was not always chosen as the answer by listeners. This can be seen in table 3.1, in which the most frequently reported incorrect answers for each segment are shown. In the light of this behaviour, minimal pair-generating native/foreign pairs will not be considered separately.

Segment[1]	Common mistakes		
	English	Spanish	Czech
[x–h]	hen/hand (n=1)	half/have (n=11)	hammer/hummer (n=3)
[r–ɹ]	rider/writer (n=2)	reach/rich (n=14)	rice/rise (n=12)
[k–kʰ]	kiss/this (n=2)	kitchen/teaching (n=4)	keeper/deeper (n=13)
[t–tʰ]	towel/owl (n=2)	tom/tongue (n=4)	tom/tongue (n=4)
[b–v]	valley/ballet (n=2)	view/bail (n=8)	valley/ballet (n=8)
[d–ð]	theirs/there's (n=2)	there/bear (n=8)	there/bear (n=15)
[w–gw]	wane/rain (n=4)	winter/gunther (n=7)	wane/vain (n=8)
[ʤ–j]	yawning/journey (n=3)	yawning/journing (n=4)	yes/jazz (n=10)
[j–ʤ]	gym/limb (n=3)	jam/young (n=5)	jam/yawn (n=4)

Table 3.1: Most frequently reported incorrect answers in the intelligibility task by each group, for accented items. Each word/confusion pair is followed by the total number of participants who reported it in parentheses.

Intelligibility results will be presented in terms of deficit from native accented words (NA) to foreign accented words (FA), as absolute results may vary depending on the lexical knowledge of the listener (e.g. absolute results only reflect the familiarity of the listener with given words, not necessarily segmental intelligibility). In figure 3.1, the cumulative loss of intelligibility by each group of listeners (as opposed to loss of intelligibility in individual segments) is presented.

The Czech cohort suffered the most in terms of intelligibility from the insertion of a foreign accented segment. On the contrary, Spanish listeners showed the smaller difference in intelligibility between native and foreign accented words. These results suggest that sharing an L1 with the speaker is beneficial for the listener, while not sharing the L1 of the speaker nor the target language is detrimental.

We used a linear mixed-effects model (LMM) to compare the loss of each group. Individual results by participants were treated as random effects [*gain* ~ *cohort*, (1|*individual results*)]. Statistical analysis was carried out using R (R Core Team, 2017). The function

[1] To reduce visual clutter, here and everywhere else in this thesis where there is no ambiguity, inner conventional square brackets are removed from segment symbols when presented in a bracketed list.

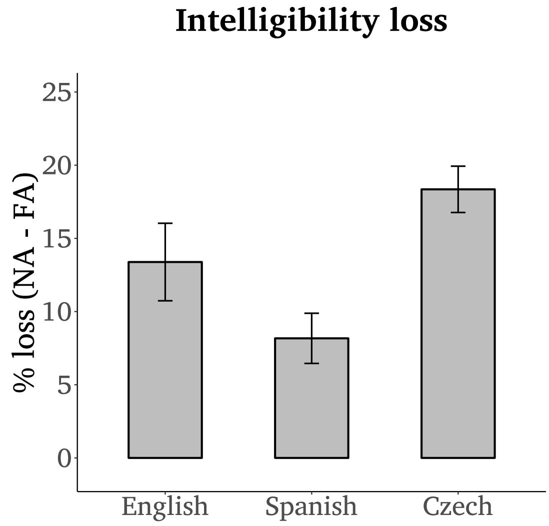


Figure 3.1: Loss in intelligibility (Native-accented segments score − Foreign-accented segments score) across groups. Vertical bars represent ±1 standard error.

employed for this analysis was *lmer*, included in the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015).

The model indicated a significant difference in the *cohort* factor at the $p < .001$ level. A Student's t test for dependent means was used to confirm if the differences displayed in figure 3.1 were significant. A significant difference in the intelligibility loss of English (μ =12.75) and Spanish (μ =8.28) was found [$p < .01$]. The loss between Czech listeners (μ =19.61) and the other two groups was also significant [$p < .001$].

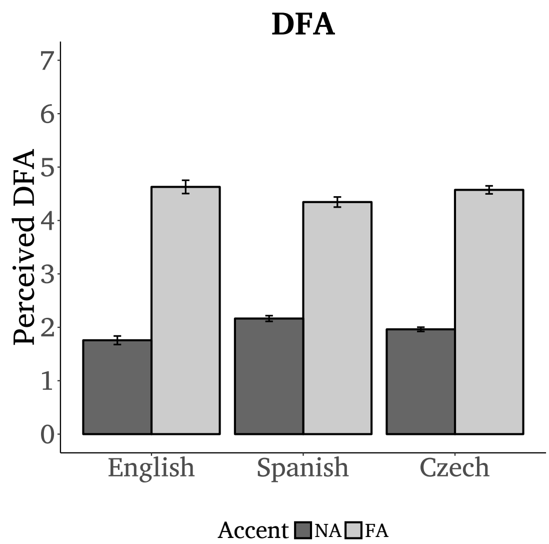


Figure 3.2: Loss in intelligibility (Native score − Accented score) across groups. Vertical bars represent ±1 standard error.

The DFA reported for native-accented words (figure 3.2) was around 2 for all groups. In foreign accented words, perceived accentedness increased to a value of around 5 for all groups. An LMM [*DFA ~ cohort × accent*, (1|*individual results*)] revealed no significant differences between cohorts. Both the *accent* effect and the interaction between *cohort* and *accent* were significant at the [$p < .001$] level. A Student's t-test revealed no significant differences between cohorts for either native or foreign items.

Individual segments were also analysed separately. In figure 3.3 each segment is plotted in terms of reported DFA for foreign accented tokens against the deficit in intelligibility. Very similar results are obtained if gain of DFA (FA-NA) is plotted instead of absolute DFA scores of foreign accented words.

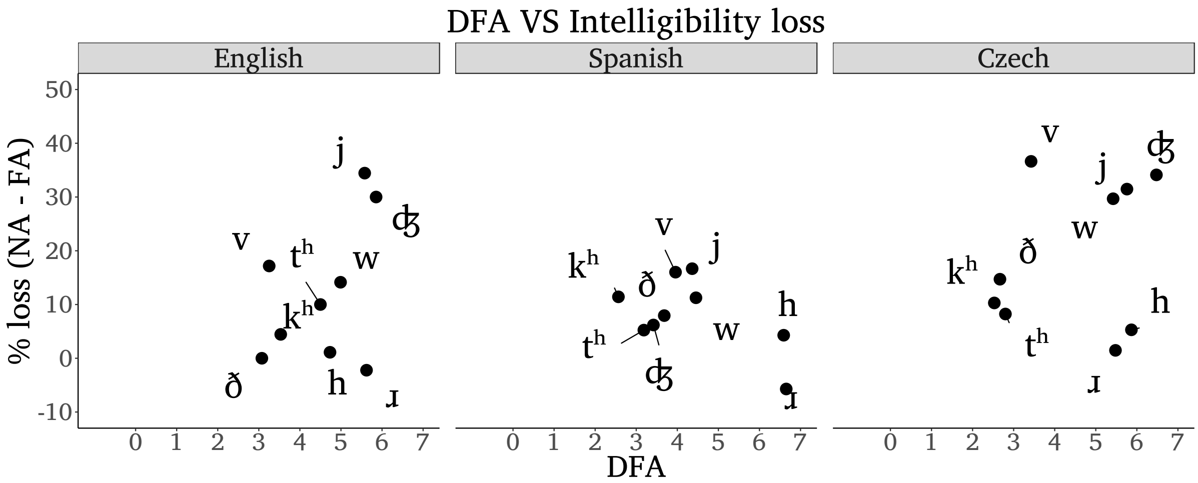


Figure 3.3: Reported DFA of accented tokens against loss in intelligibility of each segment across cohorts. The phonetic symbol used to represent each segment is the transcription of the target (i.e. the English target segment).

The foreign counterparts of [h] and [r] were perceived as highly accented by the three cohorts. However, the deficit in intelligibility was not pronounced. Furthermore, English and Spanish listeners seem to slightly gain intelligibility when the [ɹ] is accented with respect to its native realisation.

Segments [ʤ] and [j] were perceived as highly accented by English and Czech listeners, which correlated with a substantial loss in intelligibility by both groups (around 30%). However, Spanish listeners did not judge these segments as particularly accented (around 3-4 on the 7-point scale), and the loss in intelligibility was not as pronounced (especially for the affricate [ʤ]). Both English and Czech distinguish these two sounds at the phonemic level, while the Spanish phonological system includes them as allophones of the same phoneme. The fact that the drop in intelligibility was bigger in [j] for Spanish than in [ʤ] could be because they tried to accommodate to what is normal in their language, that is, the realisation of the /j/ phoneme is often closer to [ʤ] than [j] in initial position. Their responses account for this, and a higher proportion of answers with the affricate in initial position were reported regardless of the target segment.

No cohort reported a high degree of accent for the foreign tokens of [v], but the loss of intelligibility was quite noticeable, especially among Czech listeners. In Spanish, the sound [v] can appear in free variation with [b] in some accents, but in Czech and English [v] and [b] are distinguished phonemically. Similarly, Czech listeners also presented a bigger deficit in intelligibility than the other groups from native to foreign tokens in [w]. In this case, only the English system considers this sound a phoneme, while for Czech and Spanish speakers it is not phonemic. Looking at the answers of the participants, it appears that the Czechs were the only group for which a confusion [w]/[v] emerged, e.g. foreign accented *wane* was perceived as *vain* (n=9) and foreign accented *veal* was perceived as *whale* (n=14). Czech listeners have an additional competitor for these segments with respect to English and Spanish listeners, which may be the reason for the bigger intelligibility loss for the Czech cohort. This confusion is typical not only for Czechs, but for all the Slavic languages, so it is well within expectations.

Regarding [t^h], only English listeners perceived it as mildly accented. Here, perceived DFA was accompanied by a moderate loss in intelligibility. English listeners perceived the accented counterpart of [t^h] (i.e. [t]) more frequently as [d] than the other two cohorts (e.g. they confounded *touch* with *Dutch*).

Foreign counterparts of [k^h], and [ð] were not rated as particularly accented, and no big loss of intelligibility was reported by any group. These results suggest that the foreign versions of these segments are not as distant from their native counterparts as some of the other native/foreign pairs.

3.5 Discussion

In this study foreign accent has been analysed at the segmental level. By means of acoustic manipulation, we have generated English words in which the segment in initial position was replaced with its Spanish-accented counterpart. This technique contrasts directly with the holistic approach of previous research in FA (Bent & Bradlow, 2003; Munro et al., 2006; Southwood & Flege, 1999; Stibbard & Lee, 2006), providing tight control over the generated tokens and limiting the scope of FA production to isolated, selected segments. The segmental FA generation technique also prevents uncontrolled variables related to interspeaker variation, since all the tokens are generated from a single voice.

The first research question concerned the presence of a relationship between perceived DFA and intelligibility at the segmental level. Previous studies have found that both dimensions are almost independent (Derwing & Munro, 1997; Munro & Derwing, 1995a). In fact, the overall (across-segment) results (figures 3.1 and 3.2) here also suggest that there is no relationship between DFA and intelligibility, as significant differences in the loss of intelligibility are not reflected in the perceived DFA. Munro and Derwing (1995a) theorised about the possible reasons for the specific differences that lead to the independence of the two dimensions. They mention a range of possibilities, from a different interpretation of the instructions by participants, to higher order confusions (e.g., based on syntactic properties of the stimuli). Our purpose with the FA generation technique was to avoid such variability and look deeper into the effect of individual segments in the perception of non-native speech.

Results for individual segments (figure 3.3) indicate a more complex situation than suggested by the outcome of studies using holistic methods. It is especially noticeable in the case of the Spanish counterparts of [h] -[x]- and [ɹ] -[r]-, two segments that are judged as highly accented but involve almost no drop in intelligibility for any group. It is possible that the acoustic nature of the Spanish segments, which are strongly salient, makes them seem quite acoustically distanced of their less salient native counterparts. At the same time, listeners fail to assign them to any other phonological category, so is easy for them to identify [x] as [h] and [r] as [ɹ]. Using the terminology of Kuhl (1991), [x] and [r] fall inside the category of [h] and [ɹ] respectively, but are bad representatives of those categories.

The case of [v] pronounced as [b] is also worth mentioning. The loss in intelligibility is greater than what would be expected by the perceived degree of foreign accent among English and Czech listeners. In the case of native listeners much of the drop in intelligibility comes from minimal pair confusions (e.g. *van*/*ban*, *vanish*/*banish*). The case of the Czech listeners is related to the bigger drop in [w] with respect to the other two cohorts. A confusion such as this is expected in a Slavic language such as Czech. As explained in section **??**, [w] and [v] are historically related, and even though the transformation process is almost settled, both sounds can still be mistaken with each other. This phenomenon throws in an additional source of confusion for the Czech cohort: while, in the [v] accented tokens, English and Spanish listeners hesitate between [b], [v] and, to a very minor extent, [m] (e.g. Spanish: *vein*/*main* -n=6-), Czech listeners' answers also include confusions with [w] (e.g. *veil*/*whale* -n=4-). This process can also be observed in the [w] accented tokens: only Czech listeners show answers in which the target segment is mistaken with [v] (*wane*/*vane* -n=8-). Therefore, the number of phonological competitors of a native segment is also an important influence in explaining the intelligibility drop for FA tokens.

Responses to the [ʤ] and [j] tokens are particularly noteworthy. While for Czech and English listeners there is a strong relationship between perceived DFA and the deficit in intelligibility (i.e. high perceived DFA and a large drop in intelligibility), Spanish listeners reported milder values for both factors. The fact that these two segments are allophonic variations of the same phoneme in Spanish but different phonemes in English and Czech is clearly reflected in the observed perceptual behaviour. For Spanish listeners, both realisations are possible (e.g. the difference between [jestədeɪ] and *[ʤestədeɪ] seems mostly unnoticed by Spanish listeners), hence the lower values of perceived DFA. The lesser drop in intelligibility in these two sounds by Spanish listeners with respect to the other two cohorts is also an indication of this phonological particularity. English listeners try to accommodate their answers to the strange segment, even when that requires changing other segments of the word (e.g. *journey*/*yearning*) or responding with non-existing words (*jump*/**yump*), but Spanish listeners tend to adjust their answers to what is more normal in their phonological system, that is, a [ʤ]-like sound at the beginning of the word (e.g. when presented with the accented version of the word *yen*, 8 participants of the Spanish group answered *gen* and 10 answered **jen*, a word that does not exist in English). Because of this, the intelligibility drop in [ʤ] is almost zero among Spanish listeners; even in accented tokens, they still favour answers with the affricate over those containing the approximant. Therefore, when an allophonic distinction is elicited between the native and the foreign segments, listeners appear to use the rules of their L1 phonological system.

Our second research question examines from a segmental angle the discrepancies found in previous studies relating to the (mis)matched interlanguage speech intelligibility benefit. In Bent and Bradlow (2003), a matched interlanguage speech intelligibility benefit was detected, meaning that intelligibility for a non-native listener to speech from a highly proficient non-native speaker sharing their L1 was equal to intelligibility from a native speaker. Bent and Bradlow (2003) also note that non-native listeners with a different L1 from that of the speaker also experience a benefit in intelligibility with respect to the native listener. On the contrary, Stibbard and Lee (2006) failed to replicate those results, finding non-significant differences between native and high-proficient non-native speakers, regardless of their L1 matching with the listener.

In our study, we found that the three listener groups suffered a detriment in intelligibility when accented segments were introduced, but to differing extents. Spanish listeners' (i.e. the matched interlanguage group) intelligibility dropped significantly less than the native cohort, who in turn experienced a significantly lower deficit than the Czech group (i.e. the mismatched interlanguage group). In Stibbard and Lee (2006) a detriment was also found among mismatched listeners, but only when the speaker was low-proficient. In this regard, our results fail to reproduce the findings of either Bent and Bradlow (2003) or Stibbard and Lee (2006). Results from the current experiment suggest that a foreign accented segment is enough to elicit a significant detriment in the intelligibility of the mismatched interlanguage group compared to the other two cohorts.

The results of the three studies are summarized in table 3.2. For comparability, we have considered our speaker as highly competent, as only one segment was altered in each presented token. Therefore, only the results for the high proficiency speaker of Bent and Bradlow (2003) and Stibbard and Lee (2006) are compared.

			Speaker					
	Listener group			Non-native high proficiency			Native English	
BB2003	SL2006	Current	BB2003	SL2006	Current	BB2003	SL2006	Current
NN-Chinese	NN-Korean	NN-Spanish	64 (10.8)	91.69 (8.35)	81.12 (7.25)	56 (10.4)	89.5 (6.85)	92.2 (7.21)
NN-Korean	NN-Saudi	NN-Czech	60 (15.5)	89.83 (12.23)	74.52 (10.59)	60 (11.7)	91.07 (17.54)	101.02 (5.2)
N-English	N-English	N-English	77 (12.2)	110.31 (6.72)	86.66 (9.4)	109 (14.7)	112.42 (11.69)	109.84 (6.44)

Table 3.2: Mean intelligibility scores (expressed in RAU units) of the matched interlanguage group (top row), the mismatched interlanguage group (mid row) and the native group (bottom row) across two previous studies (Bent & Bradlow, 2003 -BB2003- and Stibbard & Lee, 2006 -SL2006-) and the current chapter. Values in brackets are standard deviations.

Where Stibbard and Lee (2006) found *no benefit*, as claimed in Bent and Bradlow (2003), our results suggest that, regarding segmental FA, listeners who do not share the L1 of the speaker are going to suffer a bigger *deficit* in intelligibility than listeners matching the L1 of the speaker. Additionally, we found that such deficit is highly dependent on the target segment and the phonological system of the listener.

The discrepancies between these three studies might be due to the languages selected for exploration and the simultaneity of uncontrolled effects in holistic exploration. The analysis of individual segments highlights the relevance of the relationship between the phonological systems of the languages involved in the communication process. In the case of our study, the specifics of the Czech system make listeners more prone to suffer intelligibility deficits in words starting with [v] or [w] than the other two groups, while Spanish listeners are more resistant to the [j]/[ʤ] distinction because both segments are allophones in their language. The phonetic representations of the explored segments in each group (Imai et al., 2005) have proved to be crucial to better understand where the benefit or detriment in intelligibility is coming from. It can be speculated that the differences between Bent and Bradlow (2003) and Stibbard and Lee (2006) are partly motivated by the fact that the phonetic representations of Korean and Chinese listeners, in the former, are more closely related than those of Korean and Saudi listeners, in the latter.

3.6 Conclusion

In this study, we have demonstrated that the analysis of foreign accent at the segmental level account for findings that are not detectable following the holistic method of previous studies. The expectations of the listeners regarding the segments, that is, the phonetic representation of those segments, plays a key role in the loss of intelligibility triggered by non-native speech.

This chapter introduced a method to control the accentedness of a segment without involving a change of speaker. This methodology leads to a more tightly controlled method for FA assessment. In the next chapter, a new technique that allows researchers to control the degree of foreign accent of tokens via acoustic manipulation is presented.

Chapter 4

Gradation of foreign accent

4.1 Introduction

In chapter 3 we demonstrated that segmental foreign accent can convey a drop in intelligibility and a change in perceived accentedness in listeners with different phonological systems. Phenomena such as the (mis)matched interlanguage speech intelligibility benefit (Bent & Bradlow, 2003) and the phonological mismatch hypothesis (Imai et al., 2005) were discussed in relation to those findings. Following those results, in this chapter we consider whether segmental foreign accent is perceived in a binary fashion (i.e. a segment is perceived as either foreign accented or native accented) or if listeners perceive varying amounts of deviation from the native prototype as different degrees of FA. In order to further investigate this issue, the splicing technique will be extended to generate acoustic degrees of foreign accent.

One of the first attempts at segmental continua generation (albeit not FA continua) goes back to Liberman, Harris, Hoffman, and Griffith (1957), where a continuum spanning the plosives [b], [d] and [g] was generated by an early speech synthesizer, the Pattern Playback machine. A group of native English listeners were asked to label each of the steps of the continuum as one of the three categories. Their results showed that native English listeners were able to classify the acoustic continuum in three well defined categories, i.e. that listeners perceived that continuum categorically. These results suggested for the first time the existence of what has been called categorical perception, that is, a sharp categorical boundary between phonological categories with no sensitivity to change within those categories (Tomaschek, Truckenbrodt, & Hertrich, 2011). The pioneering experiment of Liberman et al. (1957) on categorical perception gave rise to a series of new experiments carried out by a variety of authors, who also tested categorical perception for other continua such as [d]-[t] (Liberman, Harris, Eimas, Lisker, & Bastian, 1961), [b]-[p] (Liberman, Harris, Kinney, & Lane, 1961) and [ɪ]-[ɛ]-[æ] (Eimas, 1963).

Few studies have analysed foreign accent as a continuum with a purely segmental focus. Holistic studies such as Southwood and Flege (1999), in which English listeners were asked to rate the DFA of English sentences elicited by Italian speakers, concluded that FA is perceived non-categorically, that is, that the change in nativeness perceived between consecutive degrees of foreign accent is equally-sized along a $Foreign \rightarrow Native$ continuum. However, it is known that individual segments can present different degrees of deviation from the norm (Best, 1995), and the results of chapter 3 regarding perceived DFA suggest that saliency varies greatly among different foreign accented segments. In MacKain, Best, and Strange (1981) a 10-step synthetic /r/-/l/ continuum was generated. Native English and non-native Japanese listeners labelled each of the steps as either "*r*" or "*l*" in a forced-choice identification task. Their results revealed that the identification capabilities of non-native listeners with high proficiency in the target language were comparable to those of the native listeners, both reporting a clear change in the perceived category at around step 5 of the continuum. On the other hand, non-native listeners with low competence in English barely changed their responses along the continuum, which suggests that /r/ and /l/ were perceived as members of the same phonological category. In the light of the results of MacKain et al. (1981), it cannot be ruled out that each non-native speech related feature (e.g. different segments) contributes in a different way to the non-categorical perception of holistic FA.

In this chapter investigate if different native segments convey similar perceived FA given a similar acoustic deviation from the native prototype. The first research question is: do native listeners perceive segmental foreign accent categorically? As a second research question we compare native listeners with non-native listeners, and ask whether non-native listeners perceive segmental foreign accent in the same way as native listeners. Finally, our third question concerns the role of competence in the target language: does competence in a target language affect the perception of segmental foreign accent?

4.2 Categorical perception of foreign accent

In a [non-native]→[native] continuum, different perceptual patterns can be expected depending on two factors: (1) the existence or not of a sudden change in the perceived category and (2) the location along the nativeness continuum of such change. These patterns are summarized in figure 4.1. The functional form of each pattern will be fitted using a 4-parameter logistic regression function (sigmoid or S-shaped) as expressed in equation 4.1.

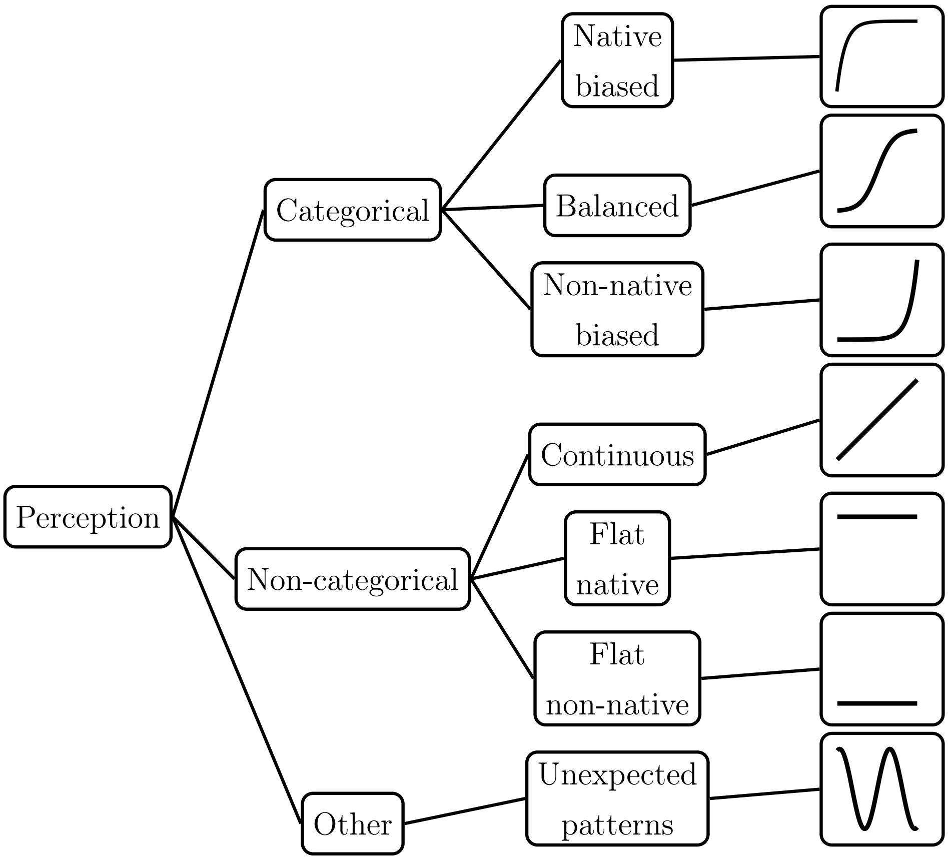


Figure 4.1: Schematic view of the expected categorisation patterns for the different continua tested in the current chapter. The x-axis represents steps in a [non-native]→[native] continuum and the y-axis represents perceived nativeness.

$$f(x) = d + \frac{a - d}{1 + \left(\frac{x}{c}\right)^{b}} \tag{4.1}$$

The four parameters of the sigmoid function are:

a: Lower asymptote

d: Upper asymptote

c: Inflection point

b: The steepness of the curve at c.

According to these parameters, participants' responses over each continuum will be considered as belonging to one or another perceptual pattern. Each pattern is defined next, and the decision procedure for the fitting regarding the four parameters of the sigmoid is described in figure 4.2.

- Categorical perception
  - Native biased: The continuum is initially perceived as non-native, but the category perceived shifts abruptly to native during the first third of the continuum.
  - Balanced: The perception is balanced between the native and non-native segments and the change in category perception is located in the central steps of the continuum.
  - Non-native biased: The perception of the whole continuum remains biased towards the non-native category until the very last steps, where there is a sudden change in perception towards the native category.
- Non-categorical perception
  - Continuous: The perception of accentedness between contiguous steps of foreign accent is equally-sized throughout the whole continuum.
  - Flat native: Every step in the continuum is perceived as native. A perceptual pattern will be considered flat native if listeners perceive every step of the continuum on the upper third of the nativeness range.
  - Flat non-native: Every step in the continuum is perceived as non-native. A perceptual pattern will be considered flat non-native if listeners perceive every step of the continuum in the lower third of the nativeness range.
- Other patterns: Unexpected patterns that do not fit any of the previous perceptual behaviours, such as reverse patterns in which the non-native end of the continuum is perceived as more native than the native end.

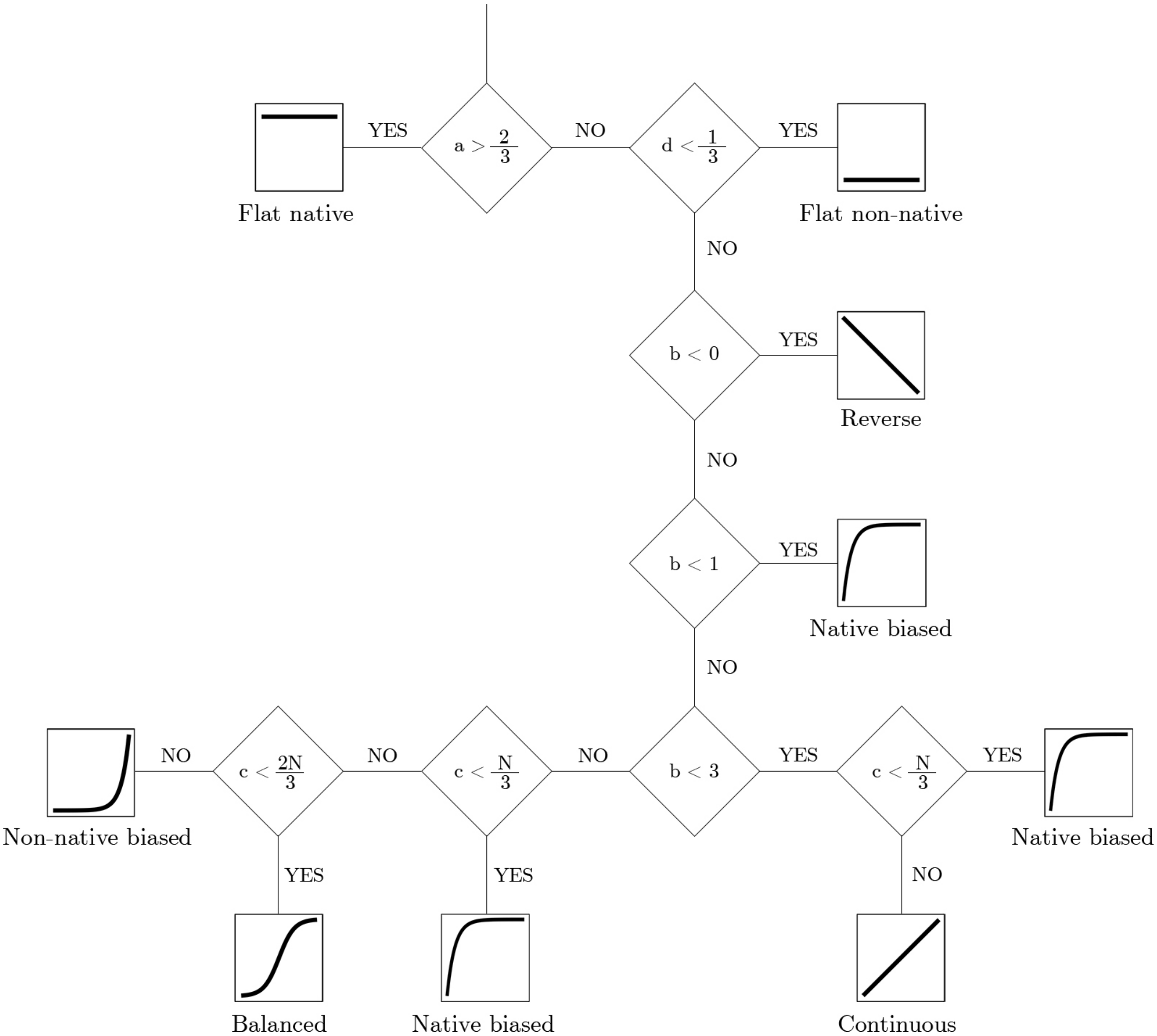


Figure 4.2: Decision procedure for categorisation curve fitting. N is the total number of steps of the continuum and a, b, c and d are the parameters of the sigmoid function (equation 4.1).

Regarding b, the steepness of the curve at c is independent of the number of steps of the continuum, i.e. two continua with different numbers of steps but equal b will have the same gradient at c. Based on visual inspection of the function, the decision was taken to set a value of $b = 3$ as the limit between a categorical pattern and a continuous pattern.

In order to determine the perceptual pattern corresponding to the categorisation results based on listeners' responses, Akaike's Information Criterion (AIC, Sakamoto, Ishiguro, & Kitagawa, 1986) will be calculated by fitting the results to the presented functional forms. Following the goodness of fit of each one of the fittings, the function that produces the lowest AIC score will be selected as the shape of the perception pattern.

4.3 The gradation technique

The goal of the gradation technique is to generate different degrees of foreign accent, in order to elicit a set of English words in which the pronunciation of a single segment is gradually modified from foreign accent to native accent through acoustic manipulations, while the rest of the word remains unchanged. The specific weight of segmental FA can be assessed by measuring (perceptually) the degree of deviation from the norm.

The technique consists of the isolation of native and non-native segments and their weighted combination in order to generate new segments made partially of features from one and partially from the other. The segments are later re-attached to the unaltered rest of the word in the same way as the spliced tokens of chapter 3, resulting in a word containing the desired degree of acoustic foreign accent in only one segment. The process, summarized in figure 4.3, is detailed below for a given step p of an n-step continuum ($p = 1..n$). The change rate for each step is $\lambda = \frac{p-1}{n-1}$, so λ varies from 0 to 1 in n equal steps.

1. **Word segmentation**

   Before generating a new segment with a degree of foreign accent, the native and non-native segments need to be isolated. For a given English word, a corresponding Spanish non-word is also selected based on the same criteria used in chapter 3 (i.e., both segments are in the same position in the word and in similar phonetic context in order to include as most coarticulation traits as possible). Subsequently, the native and non-native sounds are annotated and spliced in such a way that the native sound (N) is isolated from the rest of the native word (native frame F) and the non-native sound (NN) is also isolated from the rest of the non-word. As output of the segmentation process, N and NN serve as the basis for gradation generation, while F remains unchanged and reattached to each one of the generated steps of the continuum. In the case of this experiment, the segmentation and isolation of N and NN was performed following the criteria of Skarnitzl and Machač (2011), who indicate how to accurately place the boundaries between segments.

2. **Equating duration of N and NN**

   Since duration might be a cue for discriminating between segments, each step of the continuum is generated with a modified duration d_λ between the duration of the native segment d_N and the duration of the non-native segment d_{NN} where $d_\lambda = \lambda d_N + (1 - \lambda) d_{NN}$. Once the duration of the new segment is calculated, N and

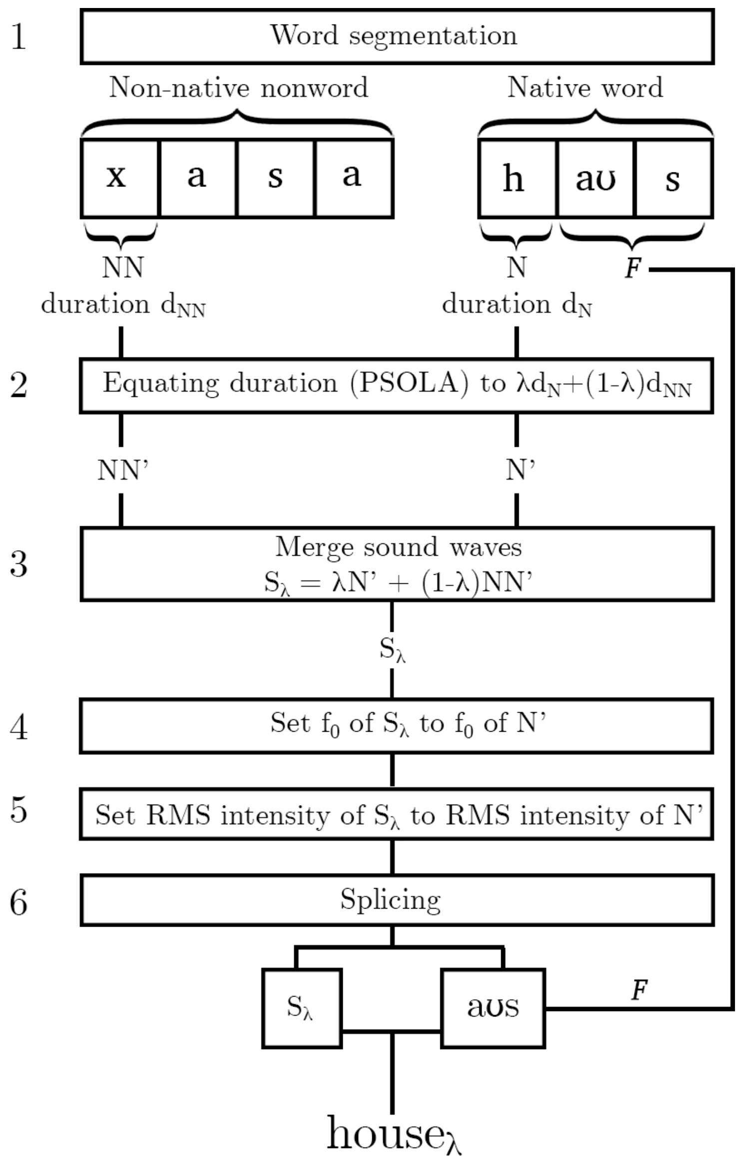


Figure 4.3: Schematic representation of the gradation process for a given step of a continuum.

NN are scaled so their duration matches d_λ, resulting in two new segments N' and NN'. The duration is modified using the PSOLA algorithm (Pitch-Synchronized OverLap and Add, Moulines & Charpentier, 1990) as implemented in Praat (Boersma & Weenink, 2018).

3. **Sound wave merging**

The sound waves of N' and NN' are combined in order to generate accented segments. The merging of N' and NN' as a function of t is performed using the *Create new sound from formula* command in Praat, applying the formula $\lambda N'(t) + (1-\lambda)NN'(t)$.

Out of this merging process, a new segment S_λ with the required duration and acoustic deviation from both native and non-native segments is generated.

Even though the native and non-native ends of the continuum can be generated with a simple overlap of the original segments (mimicking the stimuli generation process for the splicing technique depicted on chapter 3), here we explicitly generate both ends of the continuum by following the gradation technique by setting $NN'(t)$ and $N'(t)$ to 0 respectively. This decision was taken to avoid any bias that might be due to artifacts, ensuring that the ends of the continuum are treated in the same was as the intermediate points.

4. **Set f_0 of the new segment to f_0 of N'**

   Before attaching S_λ to F, it is necessary to ensure that the f_0 transition between both sounds is as smooth as possible in order to prevent artifacts that could interfere with the listeners' judgements of foreign accent. The original f_0 curve of N' is analysed, and the f_0 curve of S_λ is modified to match it using the PSOLA algorithm via Praat.

5. **Set RMS intensity of the new segment to intensity of N'**

   As stated in chapter 3, the recordings of the English words and the Spanish non-words took place in different sessions, so the intra-speaker conditions were not similar. This led to differences in intensity between the S' and F. Subsequently, the intensity of S_λ is scaled to match the intensity of N' using the *Scale intensity* command in Praat.

6. **Splicing**

   The final stage of the gradation process is the concatenation of the generated stimuli S_λ to the frame F. As with the splicing technique employed in chapter 3, an overlap between the generated segment and the unaltered frame of the word is required in order to smooth the transition between both sounds. The duration of the overlap varies depending on the available duration for it, e.g., short segments such as plosives or schwas allow for a shorter overlap than fricatives. In order to minimise the impact of the overlap on the final duration of the token, the duration of the overlap varies between 10 ms and 50 ms.

At the end of the process, a new segment *step* (i.e. the sound corresponding to the position p of the n steps continuum) is generated.

Manual adjustments

After all the stimuli were generated, one expert in English and another in Spanish listened to all continua to check for inconsistencies, artifacts or any other anomalies. In one case only, that of the continuum involving [ʤ] and [j], the f_0 curve needed to be adjusted manually after the stimuli were generated. It is possible that the acoustic differences between these two segments, more pronounced than the other cases, disrupted the automatic estimation of a correct f_0 curve. One example of a problematic f_0 curve can be seen in figure 4.4, where the f_0 curve resulting from the segment generation process (top) required smoothing (bottom) so that the transition between the generated segment and the frame remained as natural as possible, i.e. the transition between S_λ and F was as close as possible to the transition between N and F.

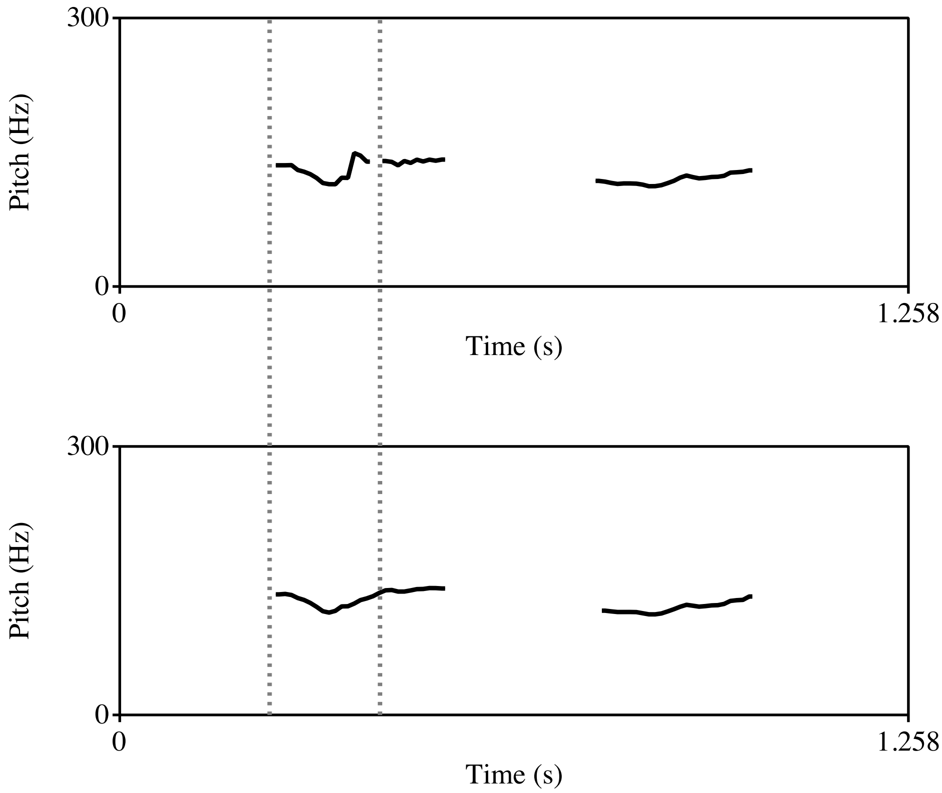


Figure 4.4: Original (top) and smoothed (bottom) f_0 contours for the sixth step of a 9-step [j]→[ʤ] continuum. The smoothed area is delimited by the grey dotted lines.

4.4 Listening experiment

As explained in section 4.1, the main purpose of this chapter is to explore the importance of listeners' competence in the target language in the perception of foreign accent. In order to evaluate this, several cohorts with different linguistic background were recruited to undergo the same experiment, consisting of two tasks: a nativeness categorisation task and an AX discrimination task.

4.4.1 Participants

Native English listeners, native Spanish listeners with high proficiency in English and native Spanish listeners with low proficiency in English were recruited for this experiment. Once the linguistic background of the participants was analysed, three different profiles were detected among the native English listeners: southern British speakers (n=16), American English speakers (n=7) and a mixture of other English backgrounds, namely: Australian (n=2), Taiwanese (n=2), Chinese (n=4) and Hindi bilinguals (n=2). The decision was taken to discard the latter mixed group and include the groups of American English and British English native speakers as different cohorts in order to check for differences across different native groups. It is important to remember here that the speaker that recorded the stimuli had a British accent.

The final set of cohorts was as follow:

1. Native British English listeners (n=16)
2. Native American English listeners (n=7)
3. Native Spanish listeners with high proficiency in English (n=13)
4. Native Spanish listeners with low proficiency in English (n=20)

Native English listeners were selected based on three requirements: English as their first language, not bilinguals in any language and no knowledge of Spanish. As for the non-native populations, the requirements were similar: Spanish as first language, two different proficiency degrees in English: B2-C1 for the high proficiency group and A1-A2 for the low proficiency group and no major knowledge of other languages.

The experiment for the native British and the native American populations was carried out at the Centre for Speech Technology Research (CSTR) of the University of Edinburgh

(UK), while the experiment for the two non-native Spanish populations took place at the Facultad de Letras of the University of the Basque Country, in Vitoria, Spain. Listening took place in a sound-attenuating booth using the same headphone model (Sennheiser HD-380 Pro) at each site.

4.4.2 Materials

A subset of six target English consonants in word initial position and two words per consonant were selected (table 4.1). The selection of these six pairs was based on the outcome of chapter 3. Two of the non-native segments, namely [x] and [r], conveyed a great degree of foreign accent among all cohorts, while the continua formed by [ʤ] and [j] entailed differences in this respect between English and Spanish listeners. The remaining two segments, [t^h] and [v], led to similar accentedness ratings for both cohorts. Such a divergence in the way of classifying these six pairs makes it worthwhile to explore them further. For each of the 12 words, a 9-step continuum was generated between the non-native segment and the native segment. The first step of each continuum was the generated non-native sound and the ninth the generated native sound.

Native sound	**Foreign sound**	**Word 1**	**Word 2**
ɹ	r	rainbow	reason
ʤ	j	gipsy	gender
j	ʤ	user	yours
h	x	hammer	happen
t^h	t	tea	type
v	b	veil	vanish

Table 4.1: Word selection for the experiment.

4.4.3 Tasks

The experiment consisted of a categorisation task and a discrimination task. These two tasks provide objective information on (a) the perceptual distance between the putative categories (namely *Native* or *Foreign*) and (b) the point of maximum discrimination between steps of the continuum.

The experiment was run with Praat (Boersma & Weenink, 2018), using the standard experimental interface that the tool provides (see the Appendix C for the experimental interfaces).

Categorisation task

The categorisation task was a two alternative forced choice test in which listeners had to decide whether the word heard was pronounced with a native or a foreign accent. Stimuli were randomly selected among all the continua and all the steps under the constraint that a stimulus from the same continuum did not appear twice in a row. For this task all the steps of the continuum for each word were presented once, except for step 9, which appeared three times. Step 9 is supposed to be perceived as fully native, and it is the only stimulus whose perceptual answer can be predicted (that is, step 9 should always be perceived as *Native*), so the decision to include it three times was taken to check consistency in the creation of the stimuli and to reduce any numerical bias towards foreign accented stimuli.

Listeners were able to see the word they were listening to spelled out on the screen, in order to avoid a Ganong effect (Ganong, 1980) or confusion with a possible minimal pair. An example of this would be the word *banish*, which can be rated as *Native* as it is an existing word in English. By showing the word *vanish* as the target, the user can disambiguate and recognize [baniʃ] as a bad representative of the word *vanish*. The main foci of the interface were two large, symmetrically placed buttons labelled as "NATIVE" and "FOREIGN" for users to choose their answers.

Listeners underwent a five-stimulus practice session with stimuli from a [d]→[ð] continuum generated following the same procedure as the main experiment, and were encouraged to choose the answer that best fitted their perception over a 50%/50% rate of answers per category. They were also informed that stimuli would play only once, and could not be repeated. The full set of stimuli was presented four times, leading to a total of 528 tokens (6 continua × 6 words × 11 steps × 4 repetitions). Participants had a 60s break every 132 trials. The completion of the task took, on average, 9.75 minutes.

Discrimination task

The discrimination task had an AX format in which listeners had to decide if two stimuli were the same or different. Both stimuli were words in the categorical perception dataset at a randomly selected step along the continuum and the same word either two steps up or two steps down the continuum (*Different* pairs). Control *Same* pairs were also introduced. The order of the stimuli in the experiment followed the same pseudo-random strategy as in the categorisation task. A summary of how the steps were paired in this task is shown in table 4.2. Each pair appeared three times for a total of 828 trials (12 continua × 23 pairs × 3 repetitions). The interval between each one of the two steps presented per trial was 0.5 seconds and there was a 60 seconds pause every 276 trials. On average, it took 30 minutes to complete this task.

Same step	1-1	2-2	3-3	4-4	5-5	6-6	7-7	8-8	9-9
Two steps up		1-3	2-4	3-5	4-6	5-7	6-8	7-9	
Two steps down		3-1	4-2	5-3	6-4	7-5	8-6	9-7	

Table 4.2: Steps presented for each continuum in the discrimination task. Step 1 refers to the fully accented version of the segment, while step 9 is the fully native version of the segment.

The interface displayed two buttons labelled as "SAME" and "DIFFERENT" which the participant had to click to choose the answer. After clicking, the next trial was automatically presented. Unlike in the categorisation task, listeners were not able to see the word they were listening to spelled on-screen. As we wanted to avoid familiarisation with the orthographic forms prior to the categorisation task, the discrimination task was presented first.

A practice session was included at the beginning of the task consisting of five trials from a [d]→[ð] continuum generated following the same procedure as the main experiment. Listeners were also instructed not to think about this task as a 50/50 task, where half of the answers fit into a category and the other half into the other one; they were encouraged to choose the answer that better suited their impression.

Statistical analysis

All processing of the data was carried out using R (R Core Team, 2017). Following J. Rogers and Davis (2009), answers whose reaction time was below 300ms or above 5000ms were removed from the analysis (<1% each task). A generalized linear mixed model (GLMM, included in the *lme4* package Bates et al., 2015 using the function *lmer*) *cohort* × *segment* × *direction* was used for comparison of upwards and downwards trials, with the individual answers as a random effect. A significant effect of the *direction* factor was found [$p < .01$]. A pairwise least-squares means test was then performed over this model, and the comparison revealed that the *direction* factor was significantly different only for English listeners in the [x]→[h] continuum [$p < .05$]. As this was the only continuum in which a significant effect was found for the *direction* factor, the decision was made to unify the upwards and downwards results in subsequent analyses.

Pairs in which the same step of the continuum was presented twice were also analysed. The English and American groups noticed no differences 93% and 91% of the time respectively; the high proficient Spanish group consider pairs as identical 88% of the times, while the equivalent figure was 87% for the low proficient Spanish cohort. A 4-sample test for equality of proportions without continuity correction revealed no significant differences between the four groups regarding the correct discrimination of the *Same* trials. As no differences were found between groups, the results of these pairs will not be further analysed. Finally, differences in the answers between words in same-segment continua were also analysed. A significant difference was found only in the answers of English listeners between the words *veil* and *vanish* (i.e. the [b]→[v] continuum), specifically in steps 5, 6 and 9 [$p < .001$].

4.5 Results

Mean responses of the four cohorts across the six segments for the six segments (figure 4.5) suggest differences in nativeness identification judgement among the four groups, especially noticeable in the low proficient Spanish cohort. On the other hand, the results for the discrimination task barely differed from one group to another.

A GLMM with *step* and *cohort* as fixed factors was generated with the results of the nativeness categorisation task. The individual results were included in the model as a random factor. A significant effect [$p < .001$] was found for the fixed factors and their interaction. A post-hoc pairwise least-squares means test revealed significant differences between the Spanish listeners with low proficiency in English and the native groups in steps 1 [English and American: $p < .001$], 2 [English and American: $p < .001$], 3 [English: $p < .001$; American: $p < .05$], 4 [English: $p < .001$], 5 [English: $p < .01$], 6 [English: $p < .05$] and 9 [American: $p < .05$]. On the contrary, highly competent Spanish listeners only differed from the English group in steps 2 [$p < .001$] and 3 [$p < .05$]. Only marginal differences were found when comparing English and American listeners [Step 9: $p < .05$], as well as when comparing Spanish listeners with low and high proficiency in English [Step 1: $p < .05$].

The responses of each group were fitted to sigmoid functions (equation 4.1) whose four parameters were extracted using the *nlsLM* command included in the package *minpack.lm* (Elzhov, Mullen, Spiess, & Bolker, 2016) for R. According to the decision procedure for categorisation curve fitting (figure 4.1), the perceptual patterns of the four cohorts were classified as non-categorical continuous.

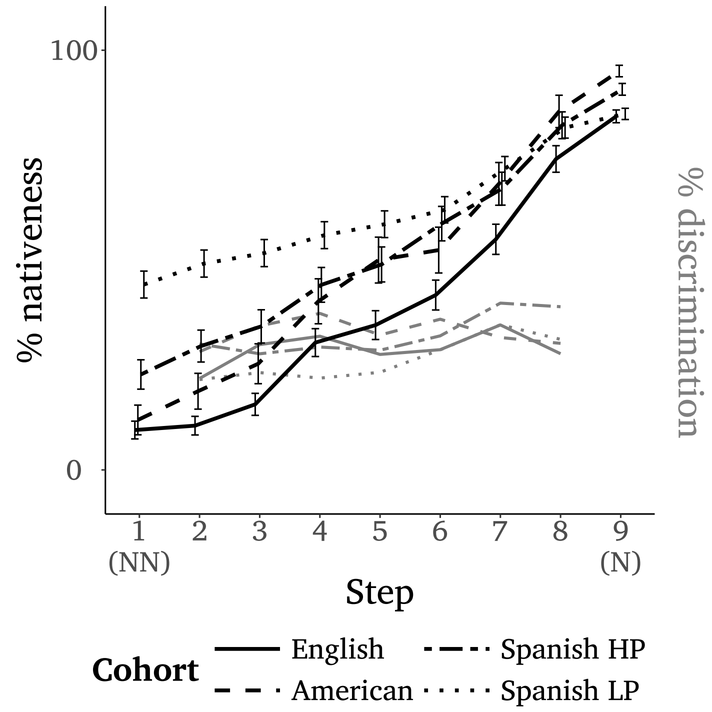


Figure 4.5: Categorisation (black lines) and discrimination (grey lines) mean responses across the six segments for the four experimental cohorts. Vertical bars represent ±1 standard error. Results have been slightly jittered to prevent over-plotting.

A similar analysis was performed over the discrimination task results. The GLMM revealed a significant effect of the *step* factor and the interaction between the *step* and the *cohort* factors [$p < .001$], but no significant effect was found for the *cohort* factor. Furthermore, no significant differences were found in the post-hoc pairwise least-squares means test, meaning that the four groups performed similarly regarding discrimination.

The responses of the four groups were also analysed for each segment individually (figure 4.6). Regarding nativeness categorisation, the four groups concurred in their answers along the [x]→[h], [r]→[ɹ] and [t]→[tʰ] continua. Listeners judged [x]→[h] and [r]→[ɹ] (figures 4.6(a) and 4.6(b) respectively) as non-native during the initial steps of the continuum, and a sudden change in category was perceived starting from step 6/7. On the other hand, nativeness perception shifted gradually during the first 6 steps of [t]→[tʰ], and the last three steps were perceived as fully native (figure 4.6(c)).

Using the nomenclature of the perceptual patterns proposed in figure 4.1, [x]→[h] and [r]→[ɹ] were perceived as non-native biased, meaning that small acoustic changes in the native segment were perceived as large deviations from the native prototype. Contrarily, the results of the [t]→[tʰ] suit the description of native biased or even balanced categorisation,

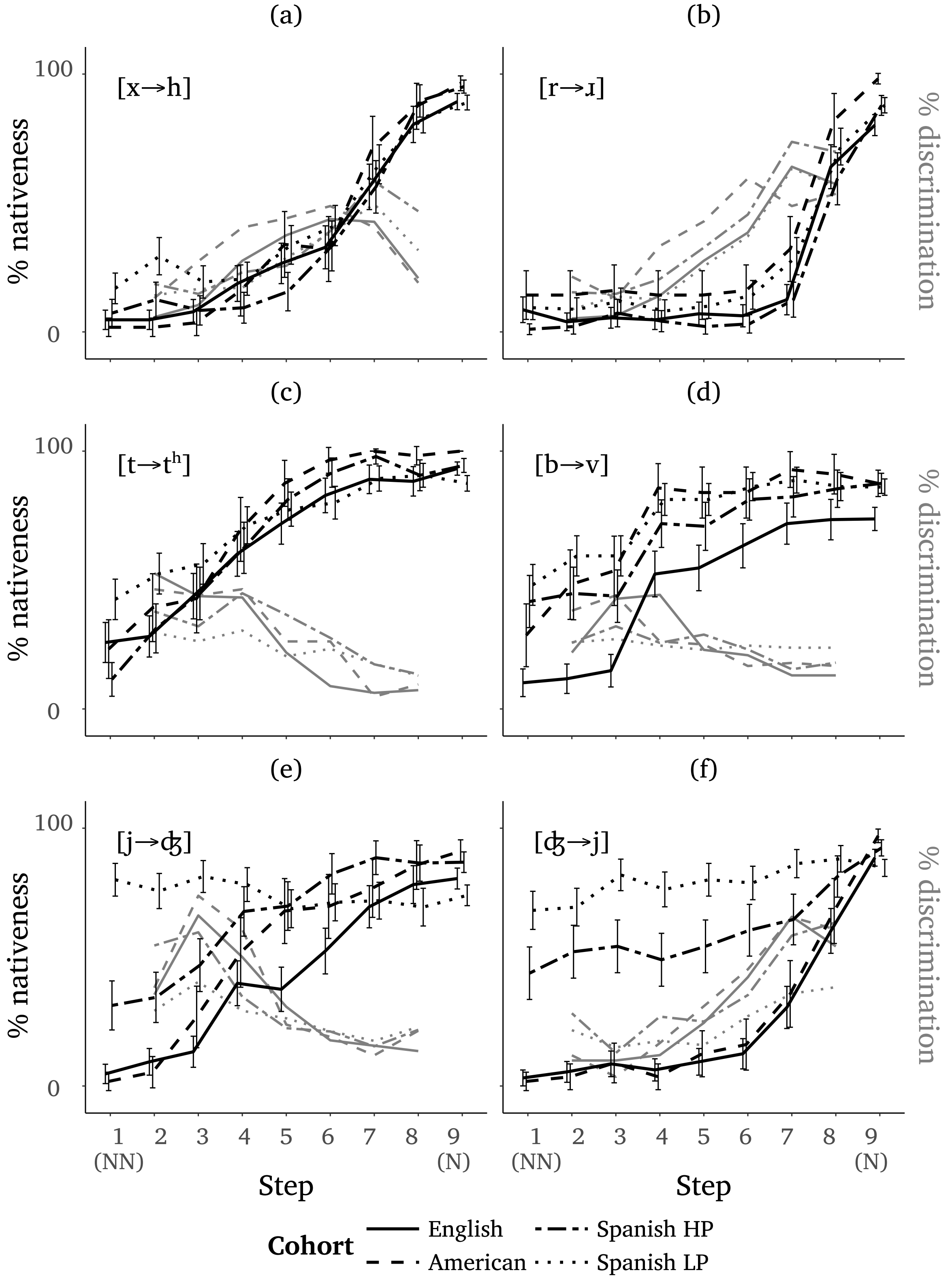


Figure 4.6: Categorisation (black lines) and discrimination (grey lines) results of the four experimental cohorts for each of the six experimental segments. Vertical bars represent ±1 standard error. Results have been slightly jittered to prevent over-plotting.

as the change in category perceived was not as sudden as in [x]→[h] and [r]→[ɹ]. This means that a segment such as [t^h] required more acoustic modification in order to be perceived as foreign accented.

English listeners judged the whole [b]→[v] continuum (figure 4.6(d)) as less native than the other three cohorts. There were also slight differences among cohorts in the initial steps. While highly proficient Spanish and English listeners perceived a sudden change in category between steps 3 and 4, American and low-competence Spanish listeners perceived a more gradual transition between steps 1 and 4. Regardless of this, the four groups perceived the last six steps with a similar percentage of nativeness.

In [j]→[ʤ] and [ʤ]→[j] (figures 4.6(e) and 4.6(f) respectively) the perceptual pattern varied greatly among cohorts. Spanish listeners with low proficiency in English judged these two continua as native regardless of the step. As mentioned earlier, these two segments do not contrast phonemically in their language. In [ʤ]→[j], a somewhat continuous trend suggested a subtle change in the perception of the first and final steps, but this change is not enough to be judged as a change in category. More likely, this pattern reflects the preference of the listeners towards a [ʤ]-like sound at the beginning of a word. In fact, the pattern in [j]→[ʤ] is opposite to what would be expected, i.e. the initial steps are perceived as more native than the final steps, which accounts for their preference of [ʤ] with respect to [j] in word initial position. However, the perception of these two continua by native listeners spanned the whole perceptual space, meaning that the initial steps were perceived as completely foreign accented and the final steps were perceived as native. The perceptual pattern of these two continua, nevertheless, differed. In [ʤ]→[j], the change in the perceived category was more pronounced between steps 7 and 9 for both cohorts, suggesting a non-native biased perceptual pattern. In [j]→[ʤ], on the other hand, the change in the perceived category is located earlier in the continuum (around steps 3 to 5). The perceived differences between steps were generally more equally-sized among English listeners. Finally, the responses of Spanish listeners with a high proficiency in English were located midway between the low-proficient Spanish listeners and the native listeners. This suggests that they were able to overcome their phonological system, even though the new categories for these two segments were still not completely fixed in their phonological system, hence the lower rate of perceived nativeness at the beginning of the continuum with respect to the native cohorts.

The results of the categorisation task were fitted to each of the expected perceptual pattern canonical functions (see section 4.1). The fitting of the results to the perceptual patterns was performed by using the command *nlsLM* of the package *minpack.lm* (Elzhov et al., 2016) included in R (R Core Team, 2017). The AIC was calculated using the command *AIC* included in the same software. The best fitting patterns according to the AIC results are summarized in table 4.3.

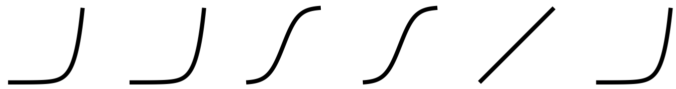
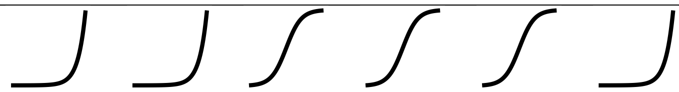
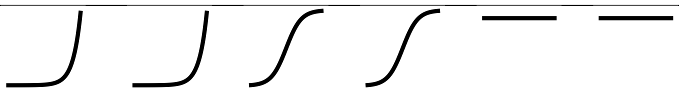
Cohort	[x→h]	[r→ɹ]	[t→tʰ]	[b→v]	[j→ʤ]	[ʤ→j]
English						
American						
Spanish HP						
Spanish LP						

Table 4.3: Summary of the perceptual patterns for the four listener groups in the categorisation task. For the meaning of each pattern, see figure 4.1.

As expected, the four cohorts agreed in the perceptual pattern of [r]→[ɹ], [x]→[h] and [t]→[tʰ]. All four cohorts perceived [b]→[v] categorically, but the transition from one category to the other started earlier for the American listeners. The disagreement between cohorts in [j]→[ʤ] and [ʤ]→[j], especially for the Spanish low proficiency cohort, addresses the differences at the phonological level of these two segments, as detailed above. It is worth mentioning that the non-categorical pattern detected for English listeners in the [j]→[ʤ] is highly conditioned by the answers of the participants in the fourth step, which are unexpectedly native-biased. Otherwise, by removing this step of the AIC fitting, the pattern would be balanced.

To confirm the significant differences between cohorts regarding the nativeness perceived in each continuum, a GLMM was set up with a *cohort* factor (4 levels) and a *step* factor (9 levels) as dependent variables. The individual responses were included as a random-effect. A significant interaction between both factors was found for all continua [$p < .001$ except [r]→[ɹ], $p < .05$]. The *cohort* factor was significant only for [b]→[v], [j]→[ʤ] [$p < .01$] and [ʤ]→[j] [$p < .001$]. Every continuum showed a significant effect of the *step* factor [$p < .001$].

A post-hoc pairwise least-squares means test was carried out to check for pairwise differences between cohorts in each step of the continua. The results of this pairwise comparison (table 4.4) confirmed no significant differences in the way the four groups perceived the [x]→[h], [r]→[ɹ] and [t]→[t^h] continua. It also confirmed that English listeners perceived the [b]→[v] continuum as slightly less native that the Spanish listeners with low-proficiency, mainly in the first three steps of the continuum. Most of the differences across cohorts were located in the [j]→[ʤ] and [ʤ]→[j] continua. The fact that most of the statistical differences involve the relationship between native listeners and non-natives with low competence in English but not the high-proficiency group acknowledges the fact that new categories can arise in the phonological system of a learner.

Step	1	2	3	4	5	6	7	8	9
[x→h]									
ENG-SLP		*							
[b→v]									
ENG-SLP	*	***	**						
[j→ʤ]									
ENG-SHP	*		*						
ENG-SLP	***	***	***	*					
AME-SLP	***	***	***						
SHP-SLP	***	**	*						
[ʤ→j]									
ENG-SHP	**	***	**	**	**	**			
ENG-SLP	***	***	***	***	***	***	***		
AME-SHP		*		*					
AME-SLP	**		***	***	***	***	*		

Table 4.4: Summary of the least-squares pairwise comparisons for the *cohort* × *step* interaction in the categorisation task. This summary entails significant differences between cohorts in a same step, from 1 (non-native end) to 9 (native end). The acronyms for the cohorts are English (ENG), American (AME), Spanish with high proficiency (SHP) and Spanish with low proficiency (SLP). The significant values are $p < .05$ (*), $p < .01$ (**) and $p < .001$ (***). Only the significantly different pairwise comparisons are displayed.

As for the discrimination results, and as predicted in section 4.1, the point of maximum discrimination correlated with the point in the continuum in which perception shifted from one category to another. Nevertheless, it is worth noting that this relationship between both tasks was stronger among native listeners and high-proficiency non-natives. In [t]→[t^h] and [b]→[v] low-proficiency non-native listeners showed a pronounced change in perceived category while the discrimination pattern remained considerable flat in comparison with the other groups. A look at the individual responses revealed that participants disagreed more frequently in the step of the continuum in which to place the discrimination peak, thereby neutralising each other and resulting in a flat discrimination pattern. Hence, the main difference of the Spanish cohort with low proficiency was not that they were not able to discriminate, but rather that the listeners did not agree on the boundary between both categories.

4.6 Discussion

In chapter 3 we concluded that listeners with a different L1 background did not perceive segmental FA similarly, either in terms of intelligibility or DFA. Moreover, significant differences were found in the way listeners perceived FA across individual segments. Following this finding, in this chapter we addressed the issue of whether listeners with different backgrounds in terms of L1 and competence in the target language categorise and discriminate segmental foreign accent similarly.

We generated continua spanning from a Spanish-accented segment to a native English segment consisting of equally-sized acoustic deviations. Unlike previous studies, these continua were not generated using stimuli from speakers with different degrees of proficiency in English. Rather, we generated the steps of the continua through acoustic manipulations of the voice of a bilingual speaker, who produced native and non-native segments. This technique provided a way to ensure that the steps of the continua were consistently, measurably equally-sized, as opposed to reliant on human evaluation, with the added incentive that our continua were comprised of tokens elicited by the same speaker rather than by different speakers.

Our first research question inquired into the (non) categorical perception of [non-native]→[native] segmental continua among native listeners. As revealed by the AIC fitting (table 4.3), both native cohorts almost always perceived the six continua in a categorical way, which means that the change in the perceived category took place within a short range of steps (generally no more than three). Notably, the area of the continuum in which such change was perceived varied from one continuum to another, i.e. each segment required a different amount of acoustic deviation from the native prototype in order to be perceived as foreign-like. We found that [h] and [ɹ] were judged *Foreign* with relatively small acoustic deviations, while other segments such as [t^h] or [v] required more acoustic deviation to convey a foreign accent. This finding suggests that, even though the perceived

DFA of the most accented token (i.e. the first step of our continua) is similar across segments, some non-native segments are more salient than others, that is, some foreign accented realisations need to be closer acoustically to the native prototype in order to stop being perceived as non-native.

Our second research question concerned the comparison between the perception of native listeners and non-native listeners. We found few significant differences between native and non-native listeners in four of the six continua, namely [x]→[h], [r]→[ɹ], [t]→[tʰ] and [b]→[v]. On the other hand, in [j]→[ʤ] and [ʤ]→[j], the perception of non-native listeners differed significantly that from the natives. As explained in chapter 3, these two segments are possible realisations of the /j/ phoneme in Spanish, and appear mostly in free distribution. Such a phonological distribution is reflected in the fact that non-native listeners perceived the whole continuum, especially the first steps, more native-like than native listeners, which implies that the difference between both ends was not as salient for them as it was for native listeners. Hence, an answer to the second research question is that native and non-native listeners perceive segmental foreign accent in a similar way as long as the relationship between the native segment and the foreign accented segment is not allophonic in the L1 of the non-native listener (Flege, 1995).

The final research question concerns the perception of segmental continua by non-native listeners with different competences in the target language. MacKain et al. (1981) demonstrated that non-native Japanese listeners with low proficiency in English did not perceive differences across the steps of an /r/-/l/ continuum. Similarly, our Spanish cohort with low proficiency in English perceived every step of the [j]→[ʤ] and [ʤ]→[j] continua as native, which suggests that they were accessing their L1 phonological system at the moment of categorising these tokens. Regarding high-proficiency listeners, Nielsen, Horn, Sørensen, McGregor, and Wallentin (2015) showed in a longitudinal study that Danish learners of Dari perceptually adapted to a [ʒ]/[ʃ] contrast after three weeks of intensive training. Likewise, the significant differences between our two cohorts of Spanish listeners in the first steps of the [j]→[ʤ] continuum lead us to think that the [ʤ]/[j] contrast can be acquired through training, even though the perceptual difference between both segments was still not as salient for the high-proficiency listeners as it was for native listeners (i.e. the non-native end of the continuum was judged as significantly less native by English listeners than by Spanish listeners with high proficiency). Therefore, and despite the fact that both cohorts of non-native listeners differed regarding the native listeners in [j]→[ʤ] and [ʤ]→[j], it is clear that, through training, listeners can retrieve information from their

L2 phonological system in order to disambiguate non-native phonemic contrasts such as [ʤ/j].

Further evidence that new phonemic boundaries can arise in the phonological system of a non-native learner are found in the results of the discrimination task. Even though the discrimination performance of low proficient non-native listeners was usually slightly lower than the other three groups, it is worth noting that the four cohorts perceived a discrimination peak in approximately the same region of the continua, including for the continua [j]→[ʤ] and [ʤ]→[j]. The discrimination task suggests that, in the absence of phonemic information concerning the contrast (i.e. top-down processing), non-native listeners made use of auditory information (i.e. bottom-up processing) to establish the boundary between non-native categories (Gerrits & Schouten, 2004; Pisoni & Lazarus, 1974). It is expected that as learners adapt to the contrast through training, the phonemic information of the L2 adds to the auditory information, resulting in better discrimination.

As a final consideration, the results of our experiment revealed that the perception of foreign accent as a whole, depicted in holistic studies as non-categorical, do not reflect the effect of individual segments. As demonstrated by the results of each continuum and the mean results across the six consonants (figures 4.6 and 4.5 respectively), each segment has a specific weight and different saliency. This finding seems to imply that the combination of different segments in larger utterances tends to result in a non-categorical continuous perceptual pattern, i.e. the nativeness perceived in sentences or larger utterances may be the result of an overall compensation across mispronounced segments with different saliency. This finding motivates the pursuit of foreign accent research from a segmental point of view.

Chapter 5

Signal quality and nativeness categorisation of continua

5.1 Introduction

In the previous chapter, a technique to generate continua of foreign accented consonants was introduced. Two groups of native English listeners and two groups of non-native Spanish listeners with different degrees of proficiency in English judged the generated continua in terms of nativeness categorisation and discrimination. Results revealed that perception of different degrees of foreign accent is highly dependent on the phonological system of the listeners and the representation of native and foreign accented segments in it.

In this chapter, we enquire further into the perception of [non-native]→[native] continua by native English listeners. In the first instance the aim is to assess the signal quality of the generated tokens. The inventory of continua is broadened to encompass target segments other than consonants in word initial position, namely consonants in word final position and vowels, and participants are asked about quality in a Mean Opinion Score (MOS) task.

A nativeness categorisation task is also be presented to the participants in order to complement the results of the MOS task. The goal is to study the impact of extremely small acoustic deviations in the perception of foreign accent, so the continua generated for this experiment consist of 21 steps. In continua generated with smaller steps, the differences along each step become less perceivable (Schneider, 2013), so nativeness differences between contiguous steps might be less pronounced than in chapter 4.

It is possible that participants judge accentedness traits as equivalent to a decrease in quality, which would imply that foreign-accented segments are so far from the canonical English phonological system that native listeners consider such realisations distorted. By comparing MOS and nativeness categorisation scores, we explore the relationship between these two dimensions. The first research question is: do native listeners perceive acoustic deviations from a given native segment as good tokens in terms of signal quality, regardless of their accentedness? The second research question reflects the choice of additional continua: do native listeners perceive segmental foreign accent categorically for vowel continua and for continua involved final consonants?

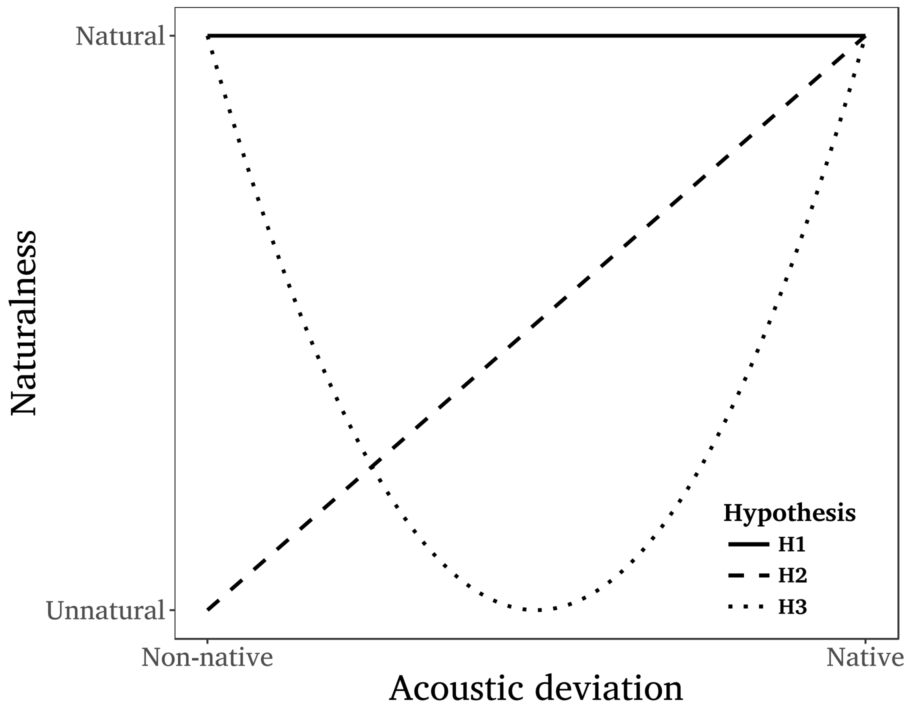


Figure 5.1: Schematic representation of the three predicted patterns of signal quality across a [non-native]→[native] continuum.

Regarding the first research question, three alternative outcomes might be anticipated for each [non-native]→[native] continuum, schematically represented in figure 5.1. These three hypotheses are:

- H1: All stimuli have good overall quality for native listeners.

- H2: Only the native end of the continuum is perceived as a good quality token, but the signal gets more distorted as the continuum approaches the foreign accented token (i.e. quality is directly correlated with nativeness).
- H3: The quality of both ends of the continuum is good, but the signal gets more distorted as the token moves away from the ends towards the middle area of the continuum.

Signal quality is generally evaluated through a MOS task in which participants rate the stimuli presented on a 5-point scale with the values 1-Bad, 2-Poor, 3-Fair, 4-Good and 5-Excellent (Dall, Yamagishi, & King, 2014; Rosenberg & Ramabhadran, 2017). Amongst other applications, the MOS scale has used on multiple occasions in international speech technology contests such as the Blizzard Challenge, which evaluates properties of synthetic speech. In the Blizzard 2008 evaluation the 5-point scale was assumed to be treated by listeners an interval scale rather than an ordinal scale (Dall et al., 2014).

As a complementary measure of the quality of the signal generated by the gradation technique, listeners will also judge the original, non-manipulated recording of the native tokens for nativeness categorisation. We will compare the nativeness perceived for this item, which is expected to be close to 100%, with the nativeness perceived in the native-like generated token of each continuum. This measure will help determine whether the gradation technique is conveying the same degree of accentedness that would be expected from a non-native human speaker at each step of the continuum.

5.2 Generation of vowel continua

For the generation of acoustic degrees of foreign accent, the acoustic characterization of the vowels required adjustments to the gradation technique introduced in chapter 4. Specifically, the third stage of the gradation process as illustrated in figure 4.3, in which the sound waves of the native and the non-native segments are merged, is replaced by two new steps specific to vowel continua generation. This stage of the gradation process is a practical application of the source-filter model, described in Fant (1960). The source-filter model proposes that every sound can be defined by two elements:

- Source: The activity of the vocal cords during phonation. The source can be periodic (for voiced sounds) or aperiodic (for voiceless sounds).

- Filter: The position of the articulators during phonation.

The sound wave merging stage is replaced here stages that accomplish source extraction and filter modification (figure 5.2).

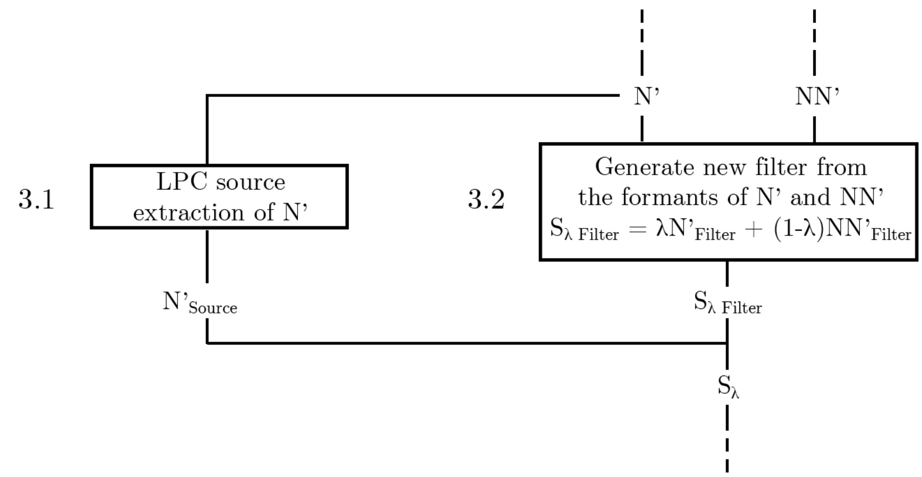


Figure 5.2: Source/filter stage of the gradation process for vowels. This process replaces the third stage of the consonant gradation process of figure 4.3. The dashed lines connect with steps 2 (top) and 4 (bottom) of the consonant gradation process.

Consonants are acoustically defined by place of articulation, manner of articulation and voicing, while the acoustic parameters typically employed to define the vowels are formants, especially the first (F1) and second (F2) formants. The third formant (F3) is also involved in particular acoustic cues, such as roundness (Van Dommelen, 2018), so these three formants are employed in order to generate the filter of the new segment. The source/filter stage is described next.

3.1 **Source extraction**

The chosen approach for source extraction is based on linear predictive coding (LPC). Figure 5.3 depicts the LPC process (adapted from Gupta & Gupta, 2016 and Yusnita, Paulraj, Yaacob, Bakar, & Saidatul, 2011). Once the linear prediction coefficients (which represent the filter) are extracted from the original sound, an inverse filter is applied and the estimated source signal is generated.

3.2 **Filter modification**

For the purpose of the present study 11 continua between English vowels and their Spanish-accented realisations were selected. A schematic representation of the English vowels studied in the present chapter and the transition to their Spanish

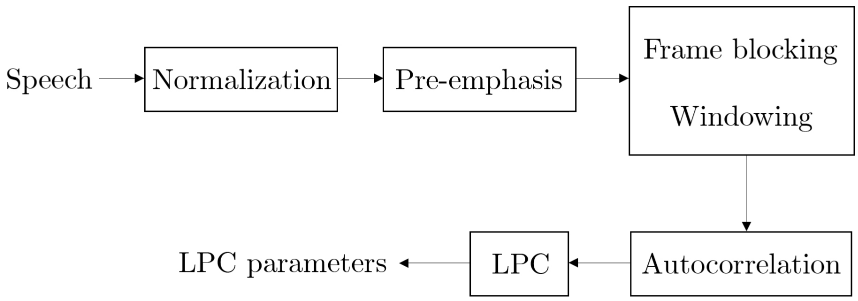


Figure 5.3: Block diagram of LPC speech analysis

accented counterparts can be seen in figure 5.4. The frequencies in Hz of the first two formants of our speaker are compiled in table 5.1.

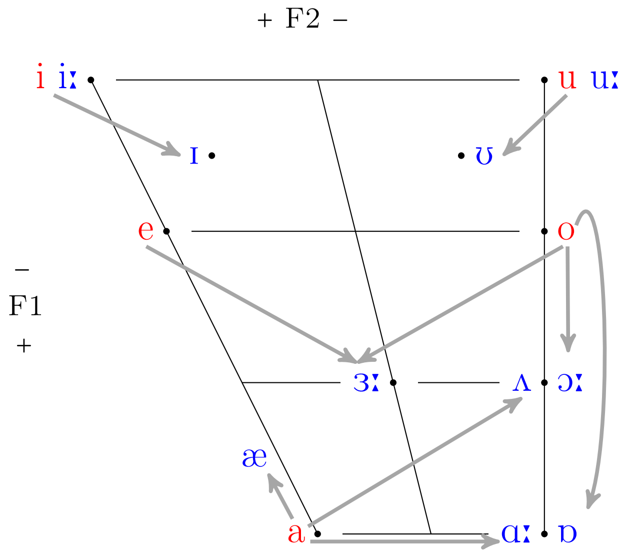


Figure 5.4: Schematic representation of vowel continua. The grey arrows mark the transitions between the Spanish vowels (in red) and the English vowels (in blue) in an F1-F2 acoustic space. Two continua ([i]→[iː] and [u]→[uː]) do not involve spectral changes, so no arrows are plotted for these continua.

Out of the 11 continua, the only expected difference between the two ends of [i]→[iː] and [u]→[uː] is vowel quantity (i.e. duration). Even so, and for consistency, these two continua are subjected to the same gradation process as the other continua. The transformations of [e]→[ɜː] and [o]→[ɜː] are special, as they represent the only case in which one English vowel can be mispronounced as two different vowels, namely [e] or [o], by Spanish speakers of English. The decision on which one to use is strongly linked to the spelling: Spanish speakers will choose [o] in cases in which the sound [ɜː] is orthographically represented by the letter 'o' (e.g. *word*, *worm*), and [e] in the rest of the cases (e.g. *bird*, *firm*).

English vowels			Spanish vowels		
Vowel	F1	F2	Vowel	F1	F2
æ	1032	1577	a	790	1597
ɑː	854	1320			
ʌ	882	1486			
ɜː	538	1653	e	577	2329
ɪ	473	2420	i	423	2742
iː	396	2865			
ɒ	508	1082	o	599	1054
ɔː	492	968			
ʊ	497	1078	u	256	726
uː	293	1054			

Table 5.1: Formant values in Hz for the vowels elicited by our speaker for this experiment, measured in the centre of the stable part of each vowel. The given values are an average derived from the four words in each continuum.

For each continuum, the values of the three formants of N' and NN' are extracted with a short-term analysis. Each sound is analysed in 25 ms windows and a Gaussian-like window is applied to each frames, and the LPC coefficients are computed using the Burg algorithm (Childers & Kesler, 1978; Press, Flannery, Teukolsky, & Vetterling, 1992). Even though only three formants are modified, this methodology requires the extraction of five formants in order to ensure accurate spectral values. If only three formants are computed, it may happen that low-frequency formants such as F1 and F2 of /o/ or /u/ get combined into one. As the speaker recruited for this experiment was female, the upper value of the formant search range was set at 5500 Hz (Quilis & Esgueva, 1983). The values for the new filter are calculated following the same formula employed in the sound wave merging: $S_{\lambda Filter} = \lambda N'_{Filter} + (1 - \lambda) NN'_{Filter}$. As explained at the beginning of this section, N'_{Filter} and NN'_{Filter} represent the values in Hertz of the first three formants of the native and the non-native vowels respectively (after duration normalisation). The formula is applied over each 25 ms window. The new filter is created with the *Create FormantGrid* command in Praat (Boersma & Weenink, 2018).

Once the source is extracted and the filter is manipulated to fit the desired transformation, both are combined to generate the new sound S_λ. In this way, the source remains the same at each step of the continuum while the filter is gradually modified from one step to another.

5.3 Listening experiment

For the purpose of signal quality rating, two tasks were designed, one to measure MOS scores of the native participants, the other a categorisation task to measure perceived nativeness. The stimuli presented in these tasks consisted of both vowel continua and consonant continua.

5.3.1 Participants

A group of 22 native English listeners with ages between 18 and 35 was recruited for this task. Recruitment took place at the Anglia Ruskin University in Cambridge, and all participants had a similar linguistic profile: native speakers of Southern British English, non-bilingual and with neither knowledge nor regular contact with Spanish. No hearing problems were detected among the participants in a pure-tone audiometry screening.

5.3.2 Materials

The set used for analysis consisted of 24 English segments and their corresponding Spanish-accented realisations. For each of these pairs, four words were selected and 21-step continua were generated. Out of the 24 segments, 10 corresponded to consonants in word initial position, 3 to consonants in word final position and 11 to vowels. For the remainder of the thesis, the continua in which the target segment appears in word final position will be marked with the _ symbol (e.g. [_x]→[_g]) as opposed to the word initial continua (e.g. [x]→[h]). All continua were generated in monosyllabic words in order to avoid syllable stress being a variable.

For some of the selected English segments, the corresponding Spanish-accented segment is also a good exemplar of a different phonological category in English (e.g., in the [b]→[v] continuum, both ends are good exemplars of different phonemes in English). For this kind of continuum, the decision was taken to include at least one word that would generate a minimal pair (e.g. *ban*→*van*) and at least one that would not (e.g. **biew*→*view*). As none of the Spanish canonical vowels are prototypical in English, no minimal pairs were expected for the vowel continua. This was also the case for the [x]→[h], [_x]→[_g] and [r]→[ɹ] continua, in which the foreign accented segments ([x] and [r]) are not good exemplars of any English phoneme. Because of the acoustic characteristics of the plosives, as explained in Flege and Eefting (1987) and Zampini and Green (2001) among others, Spanish [p], [t], [k] (voiceless plosives) might be perceived as [b], [d], [g] (voiced plosives) respectively by an English listener (figure 5.5). Therefore, Spanish segments [t] and [k] were considered as possible realisations of English [d] and [g] respectively for the purpose of minimal pair selection. The final set of words chosen for consonants and vowels can be seen in tables 5.2 and 5.3 respectively.

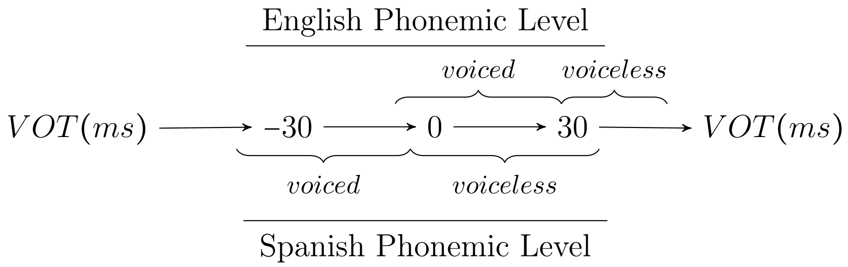


Figure 5.5: Perception of English and Spanish stops (adapted from Zampini & Green, 2001)

Continuum	Words			
[x→h]	help	hide	hole	house
[r→ɹ]	red	rent	rhyme	risk
[k→kʰ]	cast	code	cold	kiss
[t→tʰ]	town	tile	toad	tone
[b→v]	van	veil	valve	view
[d→ð]	than	that	this	thus
[s→ʃ]	shoe	short	shape	sharp
[s→z]	zap	zone	zoo	zoom
[j→ʤ]	jam	jaw	june	jest
[ʤ→j]	yak	yoke	years	youth
[_f→_b]	cab	nib	rib	crab
[_θ→_d]	god	load	dad	food
[_x→_g]	dog	frog	leg	smog

Table 5.2: Generated continua for consonants in word onset and word coda. Words shaded in grey are those for which both ends of the continuum are a minimal pair in English.

Continuum	Words			
[e→ɜː]	bird	burn	firm	learn
[o→ɜː]	word	world	worm	worse
[a→æ]	back	cat	clap	pact
[a→ɑː]	fast	raft	shark	stark
[a→ʌ]	cut	drum	gun	nut
[i→ɪ]	clip	mist	pick	sin
[i→iː]	beam	seem	steam	team
[o→ɒ]	cost	dot	pot	spot
[o→ɔː]	clause	fall	orb	storm
[u→ʊ]	look	nook	put	should
[u→uː]	choose	mood	moon	spoon

Table 5.3: Generated continua for vowels.

5.3.3 Tasks

Mean Opinion Score task

In the first task, participants were asked to rate the quality of the stimuli on a scale from 1 to 5. Due to the large number of stimuli (21 steps × 4 words × 24 continua = 2016 stimuli), out of the 21 steps generated for each continuum, only five equidistant steps were presented for evaluation (steps 1, 6, 11, 16 and 21) for a total of 480 trials (24 continua × 4 words × 5 steps = 480). Being equally-spaced, it was assumed that this sample was representative of the whole continuum. The trials were presented in semi-random order (no two steps of the same continuum appeared consecutively). This task took 13.2 minutes on average to complete.

The instructions provided to the participants appeared on-screen at the beginning of the task. These instructions were: "In this task we would like you to rate the sounds you hear for any distortions. You will have a five-point scale to rate each sound from 1 (BAD) to 5 (EXCELLENT)". The scale provided was:

1. The sound is very distorted, with BAD overall quality

2. The sound is fairly distorted, with POOR overall quality

3. The sound is somewhat distorted, with FAIR overall quality

4. The sound is slightly distorted, with GOOD overall quality

5. The sound is not distorted, with EXCELLENT overall quality

During the task, the five buttons were labelled with the numbers from 1 to 5 along with the five descriptive terms (BAD, POOR, FAIR, GOOD and EXCELLENT respectively). The word evaluated was also presented orthographically on-screen (experimental interfaces can be seen in Appendix C). The researcher asked participants to use the whole scale of five points according to their estimation, and not to limit their answers to the end points.

In order for participants to get accustomed to the task, a practice session was presented with five stimuli from a [p]→[pʰ] continuum. Participants were encouraged to ask any questions at the end of the practice session before commencing the main task.

Categorisation task

The second task was a classic two-alternative forced choice categorisation task, in which participants were asked to rate every step of each continuum as either foreign or native. Since we wanted to check the impact of small acoustic changes in the perception of foreign accent, no tokens were removed from this task. In addition to the 21 generated steps, the original recording of the native word was also included to test whether the 21st step (i.e. the native end of the continuum) was perceived as native as the non-manipulated token. The total number of trials was therefore 2112 (24 continua × 4 words × [21 steps + 1 original recording = 22 steps]), which were presented in four equal-length blocks, each separated by a 1-minute break. Each trial was presented once following the same pseudo-randomisation strategy of the MOS task, and participants were informed in advance that each sound would only be played once. Participants were encouraged to answer according to their judgement, and not to look for a 50% rate for each category. The target word was also presented on the task screen. On average, this task took nearly 46 minutes to complete.

Postprocessing

The same criteria using in previous chapters regarding reaction time was employed for the results of this experiment, i.e. all responses whose reaction time was below 300 ms or above 5000 ms were removed (< 1%). As previously stated, some of the continua generated for this experiment consisted of phonetic minimal pairs at their ends (e.g. [bæn]→[væn]).

To confirm the impact of minimal pair generation, a generalized linear mixed effects model was calculated for each task with the package *lme4* (Elzhov et al., 2016), included in R (R Core Team, 2017). The fixed factors were the analysed continua (*continuum* = 10 levels), the steps of the continua (Naturalness task: *step* = 5 levels; Categorisation task: *step* = 21 levels) and the presence or not of a minimal pair (*mp* = 2 levels). The model was used to conduct a pairwise least-squares means post-hoc test. For the MOS task, no significant differences were found for any step of any continuum between those continua that elicited a minimal pair and those that did not. For the categorisation task, steps 9 and 13 of the [j]→[ʤ] continuum were perceived significantly more native [$p < .001$] in continua that did not trigger a minimal pair. Similarly, steps 18 and 20 of the [_f]→[_b] continuum were also perceived more native in words that did not trigger a minimal pair. Finally, step 6 of the [t]→[t^h] continuum was perceived significantly more native in the same situation. Given that these were the only cases in which such a difference was found, the *mp* factor was no longer separated out in subsequent analyses.

Differences among the four words of each continuum were analysed. No differences were found for words in the MOS task. However, in the categorisation task, words *nib* and *rib* of the [_f]→[_b] continuum were judged as significantly less native between steps 18 and 20 than *cab* and *crab* [$p < .01$]. Tokens *word* and *worse* of [o]→[ɜː] and *stark* of [a]→[ɑː] were perceived significantly less native towards the end of the continuum. These segments are included in the following analysis of the results, but these specific cases are discussed.

5.4 Results

In this section, the results for the two tasks (MOS signal quality and nativeness categorisation) are presented and analysed together. Consonant and vowel continua are analysed separately.

Consonants

The overall results for consonants revealed a MOS between 3 and 4 for every step (figure 5.6), indicating that consonant continua were generally perceived as composed of good signal-quality tokens. It is worth noting that perceived signal quality showed a slight increase as stimuli became more native-like. Regarding nativeness categorisation, as in the

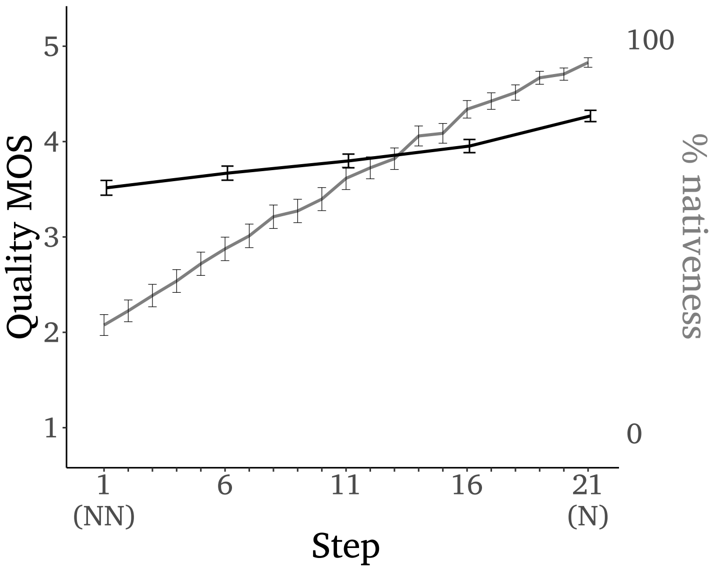


Figure 5.6: Across consonant continua means of MOS quality (black lines; left hand scale) and nativeness categorisation (grey lines, right hand scale) responses as a function of continuum step. Vertical bars here and elsewhere represent ±1 standard error. Results have been slightly jittered to prevent over-plotting

previous chapter, the mean responses across consonants revealed a clear non-categorical pattern. Nativeness and MOS responses were significantly correlated ($r = 0.95$, $p < .05$).

The MOS quality reported by the participants for individual consonant continua (figure 5.7) shows that most of the steps were perceived with GOOD overall quality, around 4 in the MOS scale, and only steps 1, 6, 11 and 16 of [j]→[ʤ] and [_f]→[_b] were judged as FAIR. A GLMM was generated with each continuum and its steps as fixed factors ($continuum = 13$ levels; $steps = 5$ levels) and the individual results as random factor. Both fixed factors and their interaction were statistically significant [$p < .001$]. A post-hoc FLSD test revealed that the only consecutive steps that were perceived significantly different were steps 16 and 21 of the [j]→[ʤ] [$p < .01$] and [_f]→[_b] continua [$p < .001$].

The categorisation results show a high degree of diversity for the different consonant continua. The non-native end of [d]→[ð] (i.e. step 1) was perceived as native more than 75% of the time, meaning that both ends of the continuum were perceived as acceptable exemplars of the /ð/ category by native listeners. In [k]→[kʰ] and [s]→[z], the first step was also perceived as fairly native-like, with nativeness ratings slightly below the upper tercile. The other continua spanned an almost complete range of perceived nativeness, from very low at the first step (e.g. [_θ]→[_d]) to almost 100% at the last step.

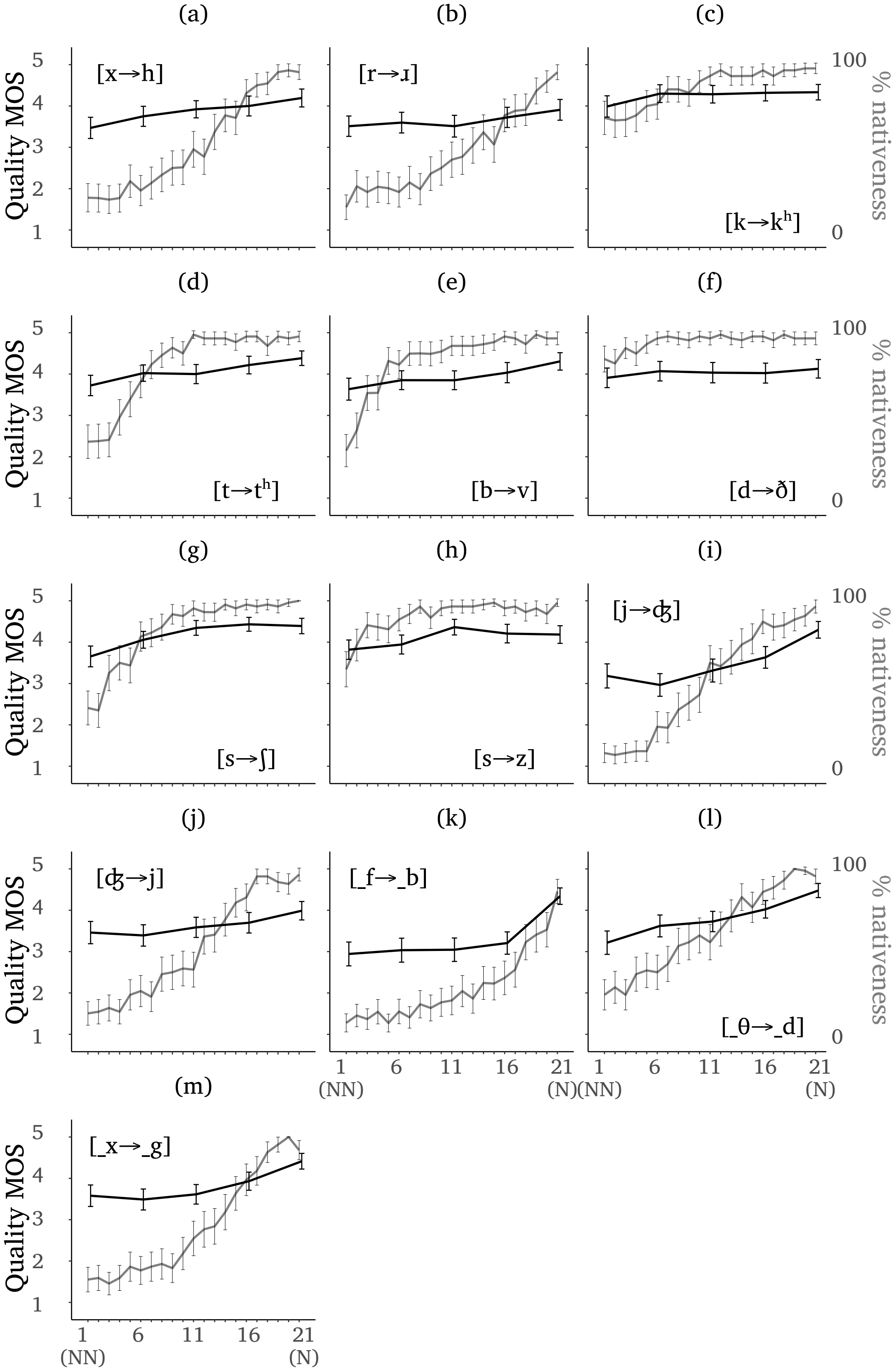


Figure 5.7: Quality (black) and nativeness categorisation (grey) responses for each consonant continuum.

As expected, by increasing the number of steps in the continuum the differences perceived from one step to the next were less noticeable. Even so, it is clear that some specific areas of each continuum conveyed a steeper rise towards nativeness than others, that is, a swift change in the category perceived. In order to illustrate the categorical nature of the continua and the area in which the change in category was perceived, an Akaike's Information Criterion was calculated for listeners' responses (table 5.4). Except for [d]→[ð], [r]→[ɹ] and [_θ]→[_d], all continua were perceived categorically. It is worth noting that the step at which the threshold between the *Native* and *Foreign* category was placed varied greatly across consonants.

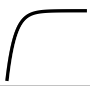
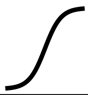
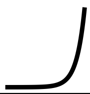
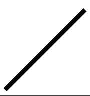
Perceptual pattern	Continua
Native biased 	[k→kʰ] [t→tʰ] [b→v] [s→ʃ] [s→z]
Balanced 	[x→h] [j→ʤ] [ʤ→j] [_x→_g]
Non-native biased 	[_f→_b]
Flat native	[d→ð]
Continuous 	[r→ɹ] [_θ→_d]

Table 5.4: Optimal fitting according to the AIC test for the categorisation results of the consonant continua.

A Pearson's correlation test was performed over the MOS task results and the equivalent five steps of the nativeness categorisation task. A significant correlation was found for [x]→[h], [r]→[ɹ], [ʤ]→[j], [_θ]→[_d], [_x]→[_g] [$p < .05$], [d]→[ð], [s]→[ʃ] and [_f]→[_b] [$p < .01$]. For the remaining five continua there were no significant correlations between the two dimensions. It is worth noting that no correlation was found mostly in continua that were categorised as native biased (see table 5.4), that is, in those cases in which the native segment required more acoustic deviation in order to be perceived as foreign. This implies that the more salient the foreign accented segment is (i.e. the more distant the native and the non-native segments are), the worse the perceived quality of the non-native steps of the continuum (e.g. continuum [_f]→[_b]).

Finally, we analysed the nativeness categorisation results of the 21st step and the original native recording (figure 5.8). A GLMM with *step* (2 levels) and *continuum* (13 levels) as the fixed factors and the individual results as random factor was generated. A significant effect was found for the *sound* factor [$p < .05$], but not for *step* nor the interaction between them. A post-hoc pairwise FLSD test was performed and no significant differences between the 21st step and the original recording were found for any continuum.

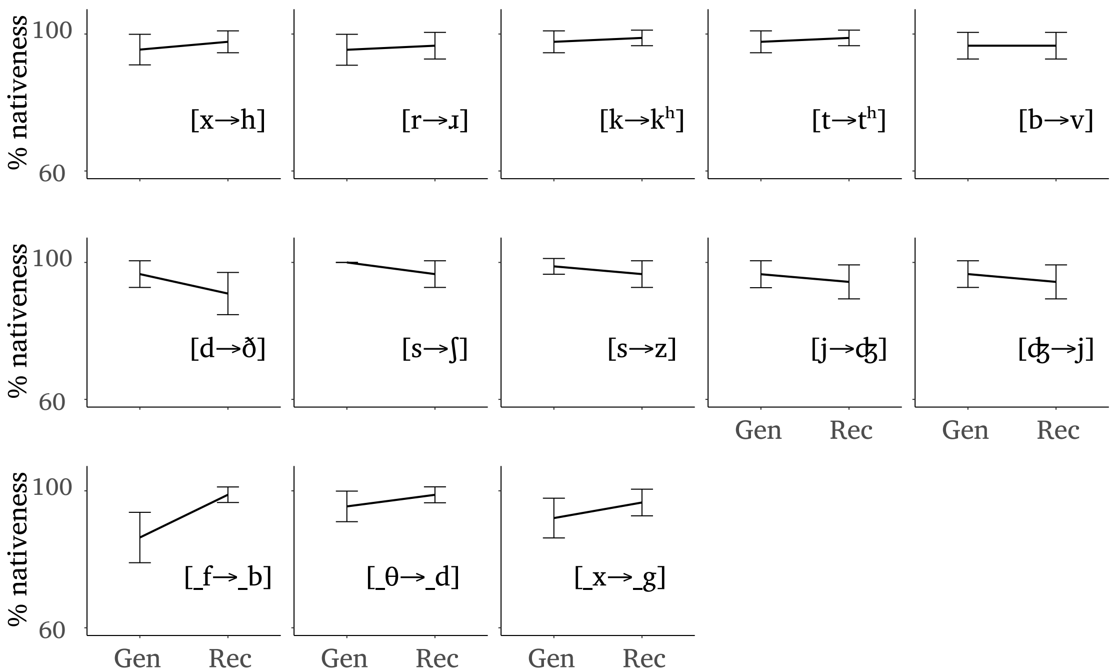


Figure 5.8: Quality (black) and nativeness categorisation (grey) tasks for the final step of the continuum (Gen) and the the original native stimuli (Rec) for each consonant. To improve visualization, the y-axis (nativeness) has been scaled to the range 60-100%.

Vowels

Overall responses to the vowel continua (figure 5.9) reveal that the quality of the generated tokens was GOOD, with a MOS of around 4 for all steps. The correlation between perceived quality and nativeness that we detected in the consonant continua was also present here [$r = 0.97$, $p < .01$]. It is noteworthy that nativeness did not reach 100% even at the native end of the continuum.

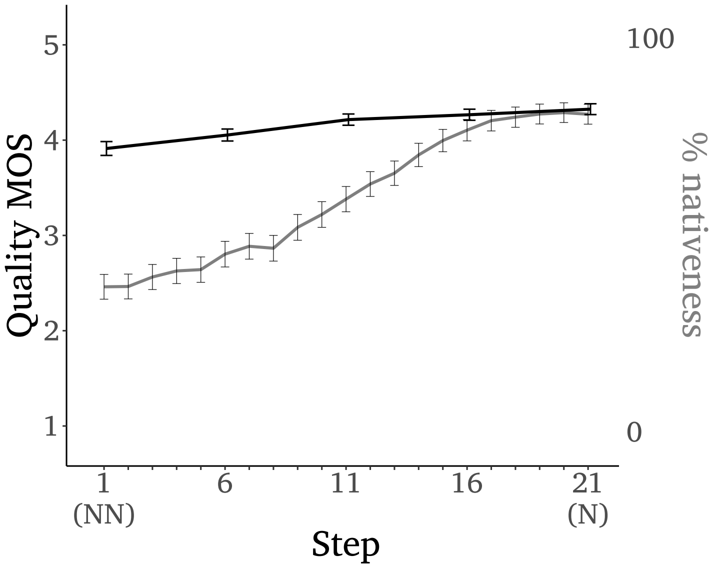


Figure 5.9: Quality (black lines) and nativeness categorisation (grey lines) results for vowels.

Individual vowel continua (figure 5.10) show that the quality of almost every token included in the vowel continua was around MOS = 4 (GOOD quality). Only in the [o]→[ɔː] continuum steps 1 and 6 were rated slightly lower, but never below 3 (FAIR quality). No consecutive steps were rated as significantly different, but for [e]→[ɜː] and [o]→[ɔː], the quality of step 21 was significantly higher than steps 1 [$p < .01$ and $p < .001$ respectively] and 6 [$p < .001$ in both continua].

As for the nativeness categorisation task, two of the vowel continua did not convey a foreign accent even at their non-native end, namely [a]→[æ] and [u]→[ʊ]. Similarly, the results of [a]→[ʌ] and [o]→[ɒ] suggest that participants did not perceive the non-native ends as particularly foreign-like, as they were judged with a nativeness score slightly below the upper tercile. For the remaining continua, listeners perceived different categories at both ends, that is, nativeness in the lower tercile at the non-native end and in the upper tercile at the native end.

According to AIC fitting (table 5.5), five of the vowel continua were perceived as non-categorical, either because no FA was detected or because the perceptual distance between acoustic steps was the same along the whole continuum.

There was a significant correlation between perceived signal quality and nativeness in five vowel continua: [a]→[ɑː], [i]→[ɪ], [o]→[ɔː] ($p < .05$), [e]→[ɜː] and [u]→[uː] ($p < .01$). It is

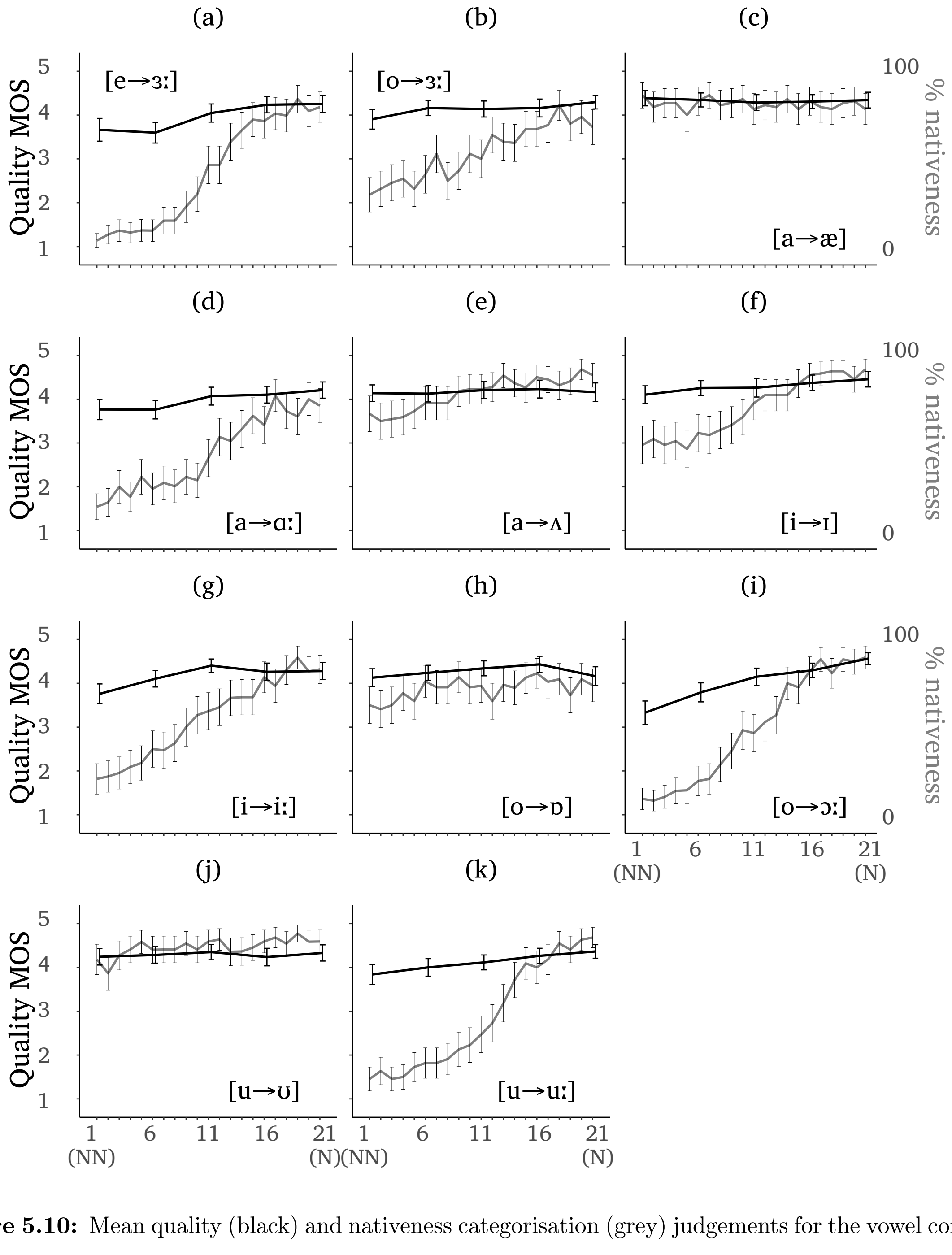


Figure 5.10: Mean quality (black) and nativeness categorisation (grey) judgements for the vowel continua.

worth noting that, out of these five continua, four of them involve durational change as an acoustic cue to convey a foreign accent.

Finally, the nativeness categorisation results of the final step of each continuum and the original native recording were compared (figure 5.11). A GLMM with *step* (2 levels) and *segment* (11 levels) as the fixed factors and the individual results as random factor was generated. A significant effect was found for the *segment* factor [$p < .001$], the *step*

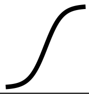
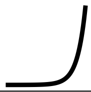
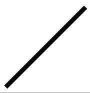
Perceptual pattern	Continua
Balanced 	[e→ɜː] [i→ɪ] [o→ɔː] [u→uː]
Non-native biased 	[a→ɑː]
Flat native	[a→æ] [u→ʊ]
Continuous 	[o→ɜː] [a→ʌ] [i→iː] [o→ɒ]

Table 5.5: Optimal fitting according to the AIC test for the categorisation results of the consonant continua.

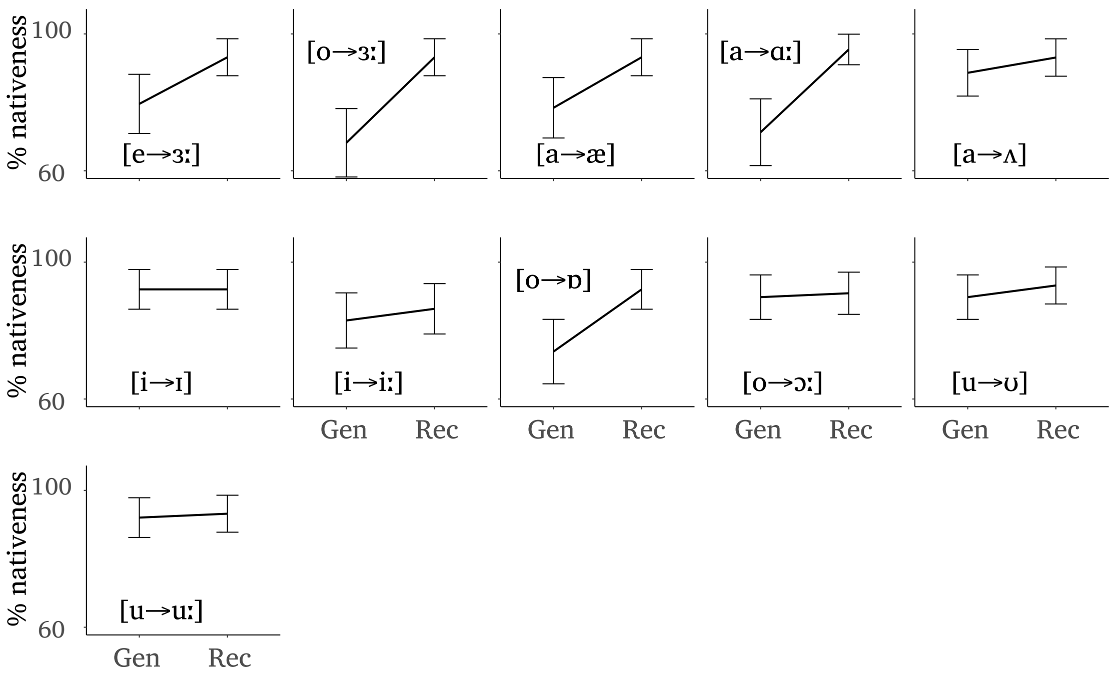


Figure 5.11: Quality (black) and nativeness categorisation (grey) tasks for the final step of the continuum and the the orignal native stimuli for each vowel. To improve visualization, the y-axis (nativeness) has been scaled to the range 60-100%.

factor [$p < .001$] and their interaction [$p < .01$]. A post-hoc pairwise FLSD test revealed a significant difference between these two tokens in the [o]→[ɜː] and [a]→[ɑː] continua (figures 5.11(b) and 5.11(d) respectively). As it has already been mentioned in the post-processing of the data, in these two continua some words were not perceived equally in the last steps (namely w[o]d→w[ɜː]d, w[o]rse→w[ɜː]rse and st[a]k→st[ɑː]k]). It is possible that in the case of these particular words, the gradation technique failed to extract proper formant values from the native recording. The implications of this will be discussed in the next section.

5.5 Discussion

The first research question asked whether listeners perceived acoustic deviations from the native segment as good tokens in terms of signal quality, measured using the MOS judgement scale. This question was motivated by a desire to determine whether perceptual results based on synthetically-graded segmental foreign accent are reliable measures of FA and not adversely affected by acoustic manipulations that lead to a reduction in quality. Overall, the generated continua were judged as GOOD in the MOS task. The four continua in which step 21 was considered higher in quality than the preceding ones (namely [j]→[ʤ], [_f]→[_b], [e]→[ɜː] and [o]→[ɔː]) were also rated to be close to 0% nativeness at their non-native end, and presented a balanced categorical perceptual pattern (except in the case of [_f]→[_b], which was non-native biased), which means that small to medium acoustic deviations from the native prototype were enough for them to be perceived as foreign. The fact that so little deviation was enough to change the perceived category highlights the saliency of foreign accent for these segments. Given this, and the fact that the quality perceived for these four continua was correlated with their nativeness, it cannot be ruled out that the extreme difference in perceived nativeness between both ends of these particular continua made listeners more prone to judge the non-native end as slightly worse in terms of signal quality. We also found that the generated native token was elicited the same degree of nativeness than the original recording. This finding also suggests that signal quality of the generated tokens is good.

The second research question aimed to extend the findings acquired in chapter 4 regarding nativeness perceived by native listeners in [non-native]→[native] continua. With this aim, three aspects were measured: first, the amount of nativeness perceived at the non-native end of the continua; second, whether the continua were perceived categorically or not; and finally, in those cases in which perception was categorical, the step of the continuum in which the change in perceived category took place.

One finding was that the foreign accented counterparts of English segments did not always convey a complete lack of nativeness. On the contrary, several fully non-native tokens (i.e. the first step of each continuum) scored above 25% on perceived nativeness, while some (viz. [d]→[ð], [a]→[æ] and [u]→[ʊ]) were not even perceived as foreign at all. The fact that no foreign accent was perceived even at the non-native end of the [d]→[ð] continuum suggests that the Spanish [d] segment is considered as an acceptable exemplar of the English phoneme /ð/, but the explanation for this behaviour might be at a higher level, namely the lexical frame. The words employed for this continuum (*than*, *that*, *this* and *thus*) were function words, a lexical category that might trigger different mechanisms in the perception process of listeners, since function words are a very small set with few competitors. Additionally, [d] is a regional realisation of /ð/. Even *than*, which in its foreign accented version is the name *Dan*, was not perceived as especially accented at the non-native end of the continuum. This finding led us to decide that the [d]→[ð] continuum should be excluded from further analysis in the present investigation. Two vowel continua also presented a flat native pattern (i.e. no foreign accent was perceived at any step of the continuum), namely [a]→[æ] and [u]→[ʊ]. In the case of [æ], it is known that in Northern English it tends to be pronounced with a higher F1 than in Southern British English (Evans & Iverson, 2004), and therefore similar to [a], its Spanish accented counterpart. Additionally, vowels [u] and [ʊ] may be neutralised in Scottish English and their lexical distribution differs in some Midlands varieties (Ferragne & Pellegrino, 2010; Stuart-Smith, 2008). It may be the case that listeners identified the Spanish accented versions of [æ] and [ʊ] with these regional variants.

A second finding is that not every continuum is perceived categorically. Repp (1984) points out that categorical perception may be specific to consonants and not to isolated vowels. Even though that work was not focused on FA and most of the experiments were performed using isolated segments rather than acoustically manipulated words, our findings support this affirmation. Besides the above-mentioned case of [d]→[ð], only two consonant continua were judged as non-categorical, namely [r]→[ɹ] and [_θ]→[_d]. On the other hand, six out of eleven vowel continua were perceived as non-categorical.

Looking into those continua which were perceived categorically, another finding is notable: the cut point between categories (i.e. the step of the continuum in which the perceived category shifted from foreign to native) was not always the same. In some continua, such as [t]→[t^h] or [b]→[v], the change in category took place during the initial steps, while others conveyed this change later, in the central area of the continuum (e.g.

[j]→[ʤ]) or in the last steps (e.g. [_x]→[_g]). This finding has important implications for non-native learners, since it helps in the identification of segments which are more prone to misperceptions on behalf of the listener. More importantly, the endurance of a native segment to be perceived as foreign through acoustic modifications seems to imply that some non-native pronunciation exemplars are perceptually more salient than others with respect to their native counterparts, e.g. the sound [b] is closer to [v] than the sound [x] is to [h] because it requires more acoustic deviation in order to convey a change in the nativeness category perceived by native listeners.

The outcome of this experiment demonstrates that the gradation technique does not in general lead to distortions or non-speech related artefacts that will be misperceived as non-native speech features. It can be assumed, therefore, that the results of the categorisation task are based only in the relationship between production and perception of foreign accent. Furthermore, we have demonstrated that each segment contributes to different extent to the perception of foreign accent, which encourages the segmental approach of this thesis.

Chapter 6

Conversion of acoustic continua of segmental foreign accent into perceptual continua

6.1 Introduction

Previous chapters have demonstrated that a range of English segments with a Spanish foreign accent can be perceived as foreign-accented, depending on the degree of acoustic deviation applied to them. While some segments are more robust and require a greater amount of acoustic deviation in order to be judged as non-native by native listeners, other segments appear to be more sensitive, and very small acoustic deviations are enough to convey a strong foreign accent. In chapter 4, [non-native]→[native] acoustic continua were classified into two major types: non-categorical continua, in which accentedness changes slowly from one step to the next, or categorical, where the perceived accentedness shifts rapidly in the course of small number of acoustic steps. We have seen that the acoustic domain does not always correlate with perception, but thanks to the gradation technique, the amount of acoustic deviation required to convey a certain degree of accentedness can be calculated. As the target language of this thesis is English, the goal of this chapter is to normalise the acoustic continua using native English listeners' perceptual judgements in order to remove cross-segmental variation in the generation of FA continua and facilitate inter-cohort comparison. Consequently, in this chapter we demonstrate how to convert a continuum in which steps are equally-sized in terms of acoustic deviation (or iso-acoustic continuum) into a continuum formed of iso-*perceptual* steps.

To support and evaluate the transformation methodology, a new set of acoustic continua will be generated from the recordings of four bilingual speakers. One of the main advantages of the gradation technique is the ability to convey different degrees of foreign accent with the voice of a single speaker. In this way, we can know that the differences perceived along an accentedness continuum are not motivated by inter-speaker variation. Here, we explore new datasets formed from several repetitions by different talkers of the 88 target words defined in chapter 5 with the exception of the [d]→[ð] and [a]→[ʌ] continua (see tables 5.2 and 5.3 for the complete list of words). Additionally, a group of native English listeners will evaluate the continua generated from the recordings of the four bilingual speakers in a 2-alternative forced choice (native/non-native) task in order to ascertain which talker is most clearly perceived as bilingual.

6.2 Iso-acoustic to iso-perceptual transformation

Continua generated for the experiments conducted in earlier chapters were formed using iso-acoustic steps i.e. using the same amount of acoustic deviation from one step to the next. In chapter 4, it was concluded that when the perception of a [non-native]→[native] continuum was non-categorical, the acoustic modifications were directly correlated with a change in perception, and every step involved the same amount of perceptual change as the others. When that change in perception was close to zero, the result was a flat pattern. On the other hand, when perception turned out to be categorical (either native biased, non-native biased or balanced), the same amount of acoustic manipulation had different perceptual outcomes dependent on the step chosen in the continuum.

In the light of the results of chapters 4 and 5, in which several continua presented categorical perceptual patterns, it can be argued that acoustic deviations do not necessarily lead to a change in perception. In this section the gradation technique will be used in order to transform the iso-acoustic continua into iso-perceptual continua, i.e. we will normalise the continua so the difference in perceived nativeness from one step to the next is nearly constant along the whole continuum. By applying this transformation, the cross-segment variability is expected to be reduced (i.e. all continua will be perceived non-categorically), and so the comparison between cohorts with different L1s that is described in the following chapter will remove one source of inter-segmental variation.

As an illustration, we examine three continua from chapter 5, namely [b]→[v], [_f]→[_b] and [i]→[iː], whose nativeness categorisation results (figure 6.1) showed different

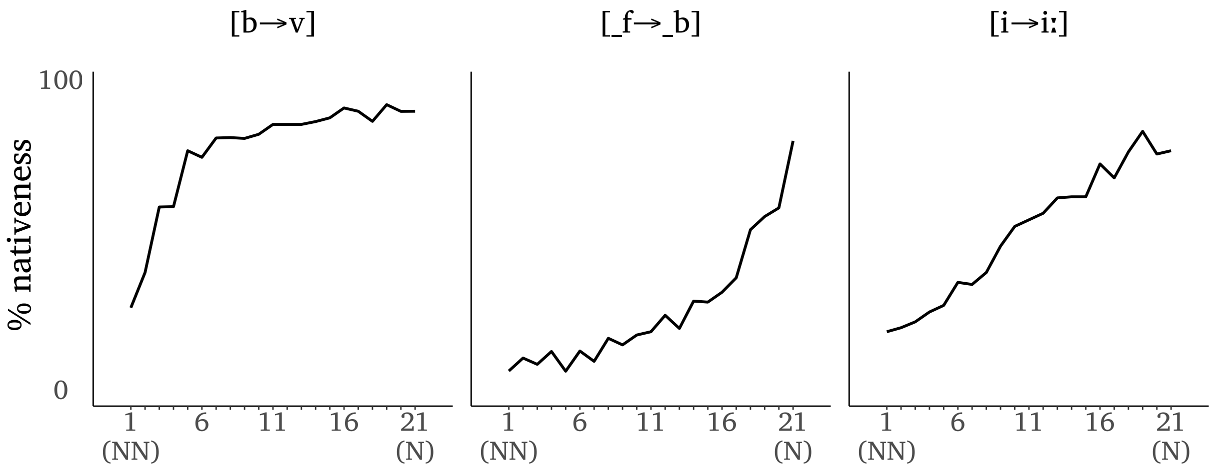


Figure 6.1: Categorisation results for the 21-step [b]→[v], [_f]→[_b] and [i]→[iː] continua of chapter 5.

patterns. According to AIC fitting, these continua can be considered as categorical native biased, categorical non-native biased and non-categorical continuous respectively. The acoustic deviation from one step to the next in these three continua is the same, but the perceptual distance between steps is not.

The different perceptual behaviour in the face of the three continua can be modelled by a four-parameter logistic (or sigmoid) function, equation 6.1. This function can be fit to the results of the nativeness categorisation task (figure 6.2) to highlight the underlying perceptual pattern. By fitting a sigmoid function to the nativeness categorisation results of a given continuum, the nativeness perceived at any step of a hypothetically infinite [non-native]→[native] continuum can be predicted.

$$f(x) = d + \frac{a - d}{1 + \left(\frac{x}{c}\right)^{b}} \tag{6.1}$$

It has already been demonstrated that certain foreign accented segments do not convey a complete lack of nativeness, and they might be perceived by native listeners as a close realisation of the native segment (see figures 5.7(h) and 5.10(b) for an example of a foreign accented segment not perceived as completely foreign). As a result, the perceptual range of a [non-native]→[native] continuum is taken to run from the lower limit of *perceived* nativeness (i.e. 0% of the possible perceived nativeness for that particular [native] target segment) to the higher limit of *perceived* natives (i.e. 100% of the possible perceived

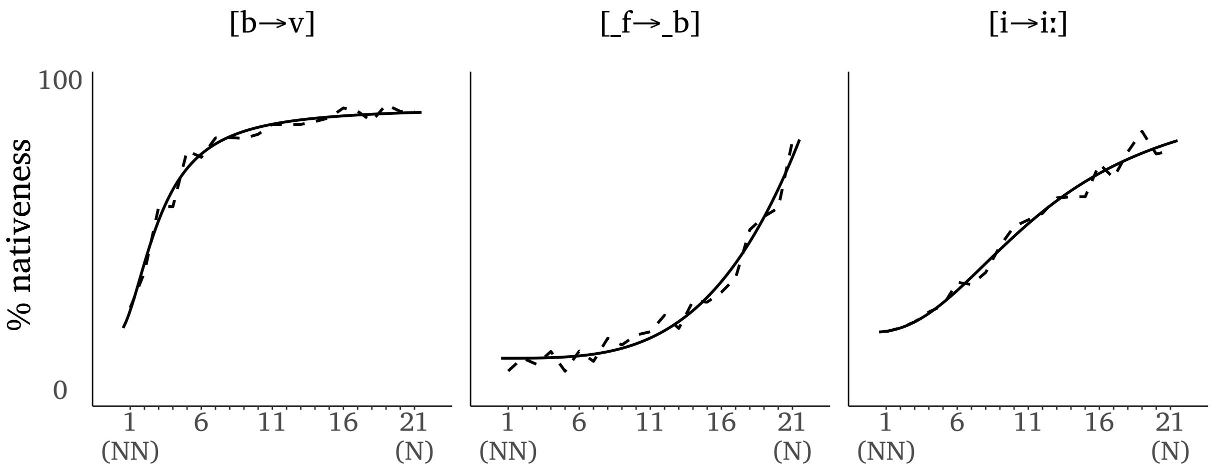


Figure 6.2: Four parameter logistic regression fits to the categorisation results for the [b]→[v], [_f]→[_b] and [i]→[iː] results from chapter 5. The dotted lines are the raw results, and the solid line is the resulting fitted sigmoid.

nativeness). Considering this, we now want to transform an iso-acoustic continuum of n_a steps into an iso-perceptual continuum of n_p steps such that its steps are equidistant in terms of perceived nativeness. First, the range of perceived nativeness is divided in n_p equidistant steps (figure 6.3).

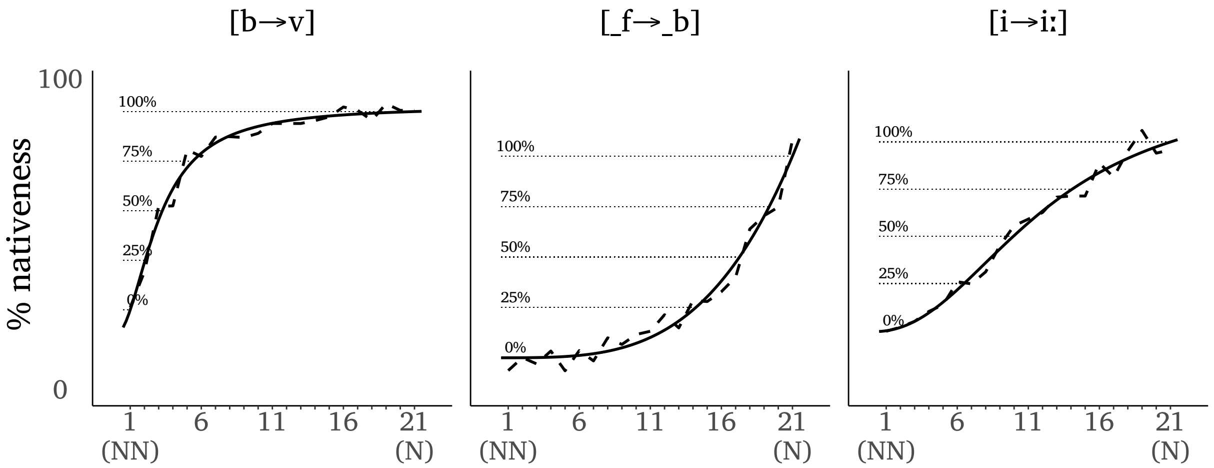


Figure 6.3: Division of the perceptual range of the [b]→[v], [_f]→[_b] and [i]→[iː] continua in 5 equidistant steps. Horizontal lines represent *perceived* nativeness in each of those steps.

Given the subdivision of the perceptual range (y axis), the corresponding acoustic values (x axis) can be extracted using the inverse function of the 4-parameter logistic function (equation 6.2) for each of the calculated y values. As a result, n_p values along the x axis are obtained, each of them representing an amount of acoustic deviation from the native prototype such as the difference in nativeness perceived between consecutive steps is equalised (figure 6.4).

$$f(y) = c\left(\frac{a-d}{y-d} - 1\right)^{\frac{1}{b}} \tag{6.2}$$

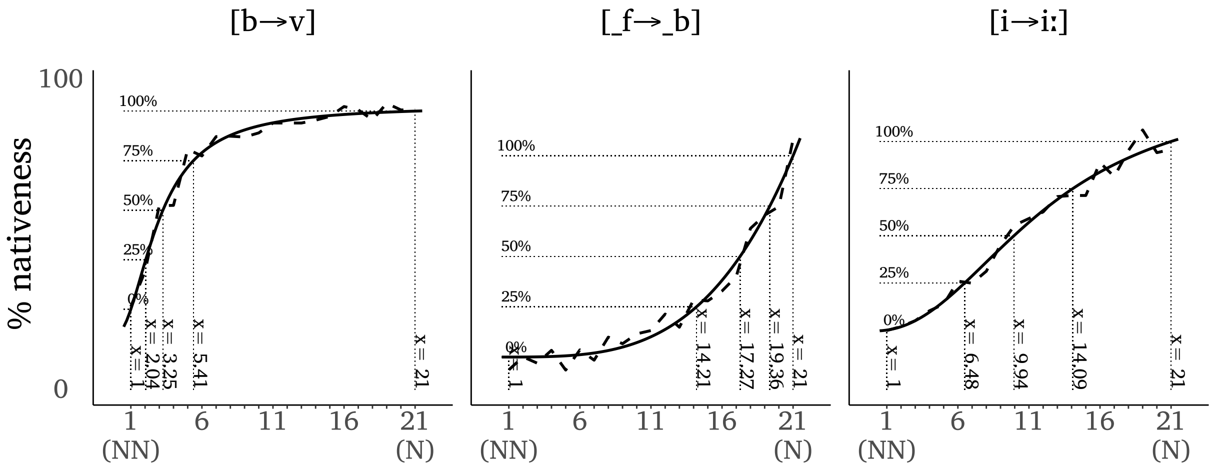


Figure 6.4: Generation of iso-perceptual continua for the [b]→[v], [_f]→[_b] and [i]→[i:] continua. Vertical lines are the steps of the continuum and horizontal lines represent perceived nativeness. The intersection of both values represents the acoustic deviation needed in order to convey a given nativeness in a hypothetical infinite [non-native]→[native] continuum.

The final process before generating the perceptual steps is to normalise the calculated values of the x axis to the range $0-1$ in order to obtain the amount of acoustic deviation that will be applied to the gradation technique. The value between 0 and 1 is what was termed λ in section 4.3. Note that the first and last steps are 0 and 1 respective in this normalised scale, and are the same in an iso-acoustic continuum and its corresponding iso-perceptual continuum.

6.3 Speaker selection

The iso-acoustic to iso-perceptual transformation requires a particular bilingual speaker: one who is able to produce both Spanish and English utterances without apparent foreign accent traits in either language. The speaker used in previous experiments was no longer available, so a group of potentially advanced bilingual speakers was recruited to provide speech material from which a dataset of 16-step continua was generated. The nativeness of these continua was then rated by a cohort of native listeners in a 2AFC categorisation task.

6.3.1 Materials

Five bilingual speakers of Spanish and English were recruited, one male (M1) and four females (F1, F2, F3 and F4), all between 18 and 25 years old. The speakers recorded two lists of words: the list of English words tested in chapter 5 (see tables 5.2 and 5.3) and a list of Spanish non-words. The latter list was designed to fit the specific contextual requirements of the target segments, e.g., for the English word "*spoon*", the Spanish non-word "*punu*" was recorded, so the native and the non-native target vowels ("*u*") were both in a similar p**X**n context. The two lists were recorded three times by each of the speakers.

The five bilingual speakers were recorded in a sound-attenuated booth at the University of the Basque Country. Other than the lists of words, the speakers were also asked to read two newspaper articles, one in Spanish and the other in English. These two recordings were later evaluated by the author and co-researchers to check for foreign accent traits in continuous speech. Following this analysis, a decision was made to discard the tokens of speaker F4, as her reading in English presented some non-native features.

Out of the three repetitions of each recorded word and non-word, one was chosen for manipulation, taking into account factors such as absence of background noise (e.g. chair movements, paper rustling) and voice clarity (e.g. no coughs, clear pitch tracks). For this experiment the continua generated consisted of 16 acoustic steps, including the native end and the foreign accented end. As mentioned earlier, these continua were the same as in chapter 5, with the exception of [d]→[ð] and [a]→[ʌ], which were removed from the final set of materials. In the case of [d]→[ð], the results of the categorisation task of chapter 5 gave rise to questions about the interference of the lexical frame (see section 5.4). The decision to discard the [a]→[ʌ] continuum was taken because [ʌ] is reported in Tollfree (2014) to be realised more in the region of [a] or [ɐ] in what the author calls 'South East London Regional Standard'. Consequently, the nativeness categorisation results might be compromised by inclusion of this continuum.

6.3.2 Participants

A group of 17 native English participants, none of whom was part of the previous experiments of this thesis, were recruited at the Anglia Ruskin University in Cambridge. The requirements for the recruitment were the same as in chapter 5: all were native speakers of English and none had knowledge of Spanish beyond the A1 (beginner) level.

They were not bilingual in any other language and did not report having lived for long periods of time in areas with a different language or strong regional accent. As in the previous experiment, participants underwent pure-tone audiometric screening; no hearing problems were detected.

6.3.3 Listening task

The categorisation task was a two alternative-forced choice task in which participants rated each of the presented tokens as either foreign accented or native. Stimuli were presented in a quasi-random fashion, with no consecutive same-continuum steps. The continua generated for this task were made up of 16 steps, resulting in a total of 5632 tokens (4 speakers × 22 continua × 4 words × 16 steps). Due to its length, the categorisation task was performed in two one-and-a-half-hour sessions on different days. On average, the entire experiment required 127 minutes to complete. Tokens were blocked by target segment (consonant continua/vowel continua) and speaker (four different types). Participants had a one-minute break after the completion of each speaker's stimuli (i.e. three breaks per session). The experiment was generated using the experimental interface of Praat (Boersma & Weenink, 2018).

Postprocessing

As in previous chapters, responses whose reaction time was below 300 ms or above 5000 ms were removed from the analysis ($<$1%). The differences between voices were calculated by generating a generalised linear mixed-effects model with *speaker* and *step* as the fixed factors and the individual results as random factors. All analyses were performed using R (R Core Team, 2017).

6.3.4 Results

Figure 6.5 plots nativeness categorisation judgements. It is clear that the tokens elicited by speaker F2 convey a much more marked degree of accentedness than the other three voices. For the consonant continua F2 was judged significantly more accented than M1 [step 1 = $p < .01$; remaining steps = $p < .001$], F1 [step 2 = $p < .05$; step 7 = $p < .01$; remaining steps = $p < .001$] and F3 [all steps = $p < .001$]. Similarly, speaker F2 was perceived as significantly more accented than the other three speakers in all steps of the vowel continua [$p < .001$].

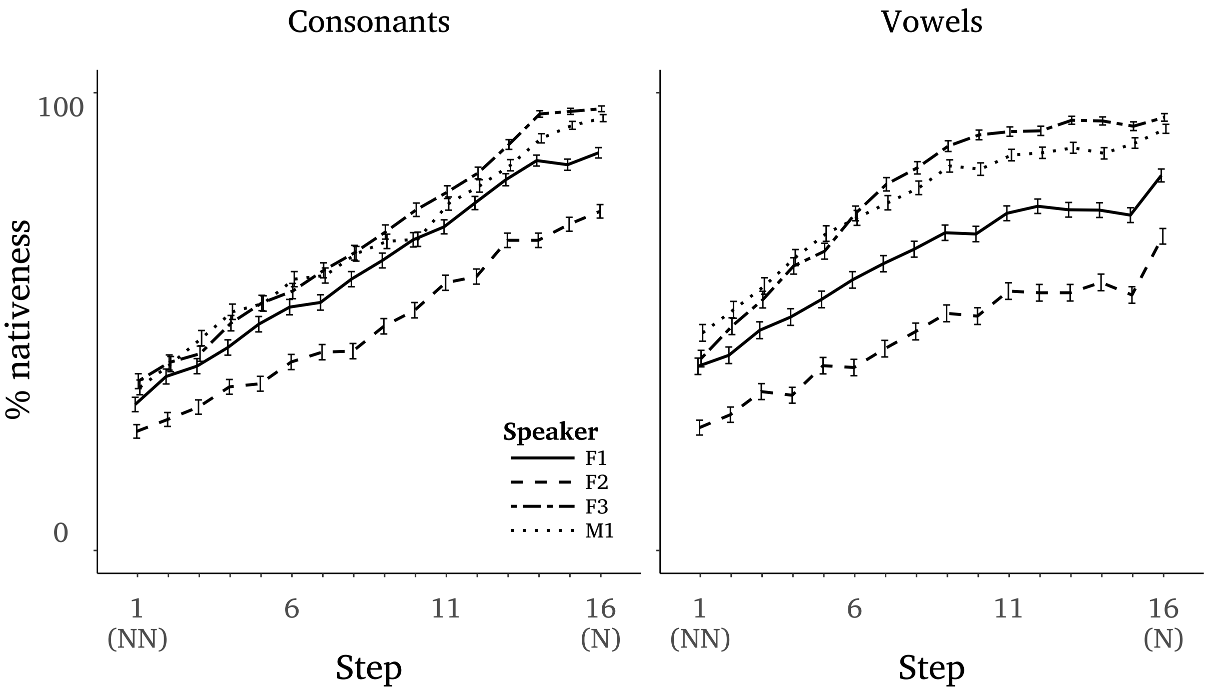


Figure 6.5: Results of the categorisation task for the four speakers across all consonants (left) and vowels (right). Vertical bars represent ±1 standard error. Results have been slightly jittered to prevent overplotting.

It is also noticeable that the degree of nativeness elicited by speaker F1 did not reach the expected values at the end of the continua, where listeners perceived the nativeness of speaker F1 as significantly lower than M1 [steps 15-16: $p < .001$] and F3 [steps 14-16: $p < .001$]. This behaviour is noticeably more pronounced in the vowel continua, where speaker F1 was perceived as significantly more accented than M1 [steps 4-16: $p < .001$] F3 [step 4: $p < .01$; step 5: $p < .05$; steps 6-16: $p < .001$].

Given these results, tokens generated from speakers F1 and F2 were considered not suitable for foreign accent evaluation. The differences detected between speakers M1 and F3 were marginal in both consonant [step 14 = $p < .05$] and vowel [steps 10, 14 = $p < .05$] continua, and both speakers reached nearly 100% of perceived nativeness at the native end of the continua. For consistency with the previous experiments in this thesis, in which tokens were generated from the voice of a female speaker, the decision was made to generate iso-perceptual continua from the results of speaker F3. Following the methodology described earlier (section 6.2), we calculated, for each of the 22 consonant and vowel continua, the set of 5 steps whose perceived nativeness was equidistant, this is, the 5 values of λ required in order for the continuum to be perceived as non-categorical by native listeners. As has already been mentioned, the first and last steps of an iso-perceptual continuum match the first and last steps of its corresponding iso-acoustic continuum, so steps 1 and 5 of the iso-perceptual continua have the same λ value as steps 1 and 16 of the iso-acoustic continua, i.e. 0 and 1 respectively. Corresponding λ values extracted from speaker F3 are presented in tables 6.1 and 6.2 for consonant and vowel continua respectively.

Continuum	Step 1	Step 2	Step 3	Step 4	Step 5
[x→h]	0	0.53	0.64	0.75	1
[r→ɹ]	0	0.42	0.59	0.77	1
[k→kʰ]	0	0.22	0.32	0.45	1
[t→tʰ]	0	0.17	0.23	0.32	1
[b→v]	0	0.16	0.25	0.38	1
[s→ʃ]	0	0.23	0.32	0.45	1
[s→z]	0	0.17	0.32	0.56	1
[j→ʤ]	0	0.33	0.47	0.64	1
[ʤ→j]	0	0.28	0.50	0.74	1
[_f→_b]	0	0.66	0.74	0.83	1
[_θ→_d]	0	0.53	0.65	0.78	1
[_x→_g]	0	0.65	0.76	0.86	1

Table 6.1: λ values required to generate iso-perceptual continua from the voice of speaker F3 in the consonant continua. Step 5 of the iso-perceptual continua is the same as step 16 of the iso-acoustic continua generated for this experiment.

Continuum	Step 1	Step 2	Step 3	Step 4	Step 5
[e→ɜː]	0	0.30	0.37	0.47	1
[o→ɜː]	0	0.22	0.34	0.49	1
[a→æ]	0	0.07	0.09	0.13	1
[a→ɑː]	0	0.32	0.47	0.65	1
[i→ɪ]	0	0.23	0.35	0.51	1
[i→iː]	0	0.21	0.31	0.45	1
[o→ɒ]	0	0.11	0.16	0.24	1
[o→ɔː]	0	0.23	0.31	0.42	1
[u→ʊ]	0	0.14	0.24	0.40	1
[u→uː]	0	0.26	0.35	0.47	1

Table 6.2: As table 6.1 for vowel continua.

Conclusion

In this chapter we have outlined a methodology that, by using the gradation technique, allowed us to transform iso-acoustic continua into iso-perceptual continua. The goal of the iso-perceptual transformation methodology was to gain the ability to elicit a specific degree of foreign accent among native listeners in a controlled way in order to remove cross-segment variability. The importance of generating the continua from the voice of a speaker whose pronunciation is perceived as foreign accent-free to native listeners has also been highlighted. In the next chapter the effect of iso-perceptual continua in the perception of listeners differing in their L1 will be studied.

Chapter 7

Assessment of perceptual degrees of foreign accent

7.1 Introduction

Summary of experiments in earlier chapters

This thesis has presented an analysis of the production of foreign accented segments seen through the responses of various cohorts of listeners differing in their L1 phonological system to a set of perceptual tasks. The present chapter explores how listeners of four cohorts differing in their L1 perceive segmental foreign accent in 22 iso-perceptual continua including consonants and vowels as target segments. To illustrate the progression towards the body of work presented in the current chapter, previous experiments presented in this thesis, along with the experiment of the current chapter, are summarised in table 7.1.

From the outset, this thesis has focused on how listeners with different L1s perceive foreign accented segments in terms of intelligibility and degree of foreign accent (chapter 3). Our findings indicate that each segment has a specific weight in the loss of intelligibility linked to the appearance of a foreign accent. This outcome led to a new question: how are different degrees of foreign accent perceived? In order to further investigate this factor, the gradation technique was developed as a way of generating acoustic continua spanning from a foreign accented token to a native one. We demonstrated that, even though previous holistic studies have observed that different degrees of foreign accent are perceived non-categorically, such an outcome is likely to be due to the accumulation of

Chapter	Goal	Cohorts	Tasks	Stimuli	Target segment
Ch. 3	Selection of technique for segmental foreign accent generation	English Spanish Czech	Intelligibility DFA	Words	9 consonants
Ch. 4	Categorical perception of accentedness	English Spanish HP Spanish LP	Nativeness Discrimination	Words (9-step continua)	6 consonants
Ch. 5	Nativeness in vowels and signal quality-nativeness correlation	English	Nativeness MOS	Words (21-step continua)	13 consonants & 11 vowels
Ch. 6	Speaker selection and iso-perceptual transformation	English	Nativeness	Words (16-step continua)	12 consonants & 10 vowels
Ch. 7	Segmental FA assessment by native and non-native listeners in iso-perceptual continua	English Spanish Czech Japanese	Intelligibility Discrimination Identification Nativeness	Words* (5-step continua iso-per)	12 consonants & 10 vowels

Table 7.1: Summary of goals, listener cohorts, tasks, stimuli and target segments of each experiment presented in this thesis, including the current chapter. *In the identification task syllables were presented instead of words.

individual traits pulling out in different directions (chapter 4) – a manifestation of the central limit theorem.

Chapter 5 validated the signal quality of the tokens generated via the gradation technique. Native English listeners evaluated several consonant and vowel continua in a mean opinion score task and, for the most part, signal quality was judged as good. Finally, chapter 6 introduced a methodology to transform the generated continua, in which steps were equidistant in acoustic deviation from the native prototype (iso-acoustic), into continua formed from steps equally-spaced in perceived nativeness by English listeners (iso-perceptual).

The current chapter brings all these threads together.

Motivation

This chapter introduces a set of perceptual tasks (7.1) designed to answer the following three research questions.

The first research question asks: What is the effect of a listener's L1 segmental system in the perception of segmental foreign accent? It has already been demonstrated that listeners differing in their L1 perceive Spanish segmental foreign accent in English words in dissimilar ways. So far, the only two cohorts to assess different degrees of foreign accent have been English and Spanish, that is, listeners whose L1 was either the target language or the source of FA. In this chapter two additional cohorts are included for comparison: (i) Czech listeners, whose results in chapter 3 revealed significant differences in the perception of some foreign accented segments regarding English and Spanish listeners; and (ii) Japanese listeners, whose phonological inventory is generally closer to the Spanish one, especially with regard to the vowels (Morrison, 2002). By comparing these four cohorts we expect to gain insight into the role of the phonological system in the perception of foreign accented segments.

In every experiment of this thesis, segmental foreign accent has been evaluated in carrier words. It still remains uncertain how listeners perceive non-native segments in a purely acoustic frame, that is, in a context totally devoid of lexical meaning. The second research question concerns this issue: does the lexical context affect the perception of foreign accented segments? The lexical context can be expected to help in the disambiguation of unclear accented segments, but it is possible that access to such lexical information imposes differential cognitive load across cohorts. In order to determine the importance of the lexical context, in this chapter listeners are asked to identify foreign accented segments in two different tasks: a word intelligibility task and an identification task in which the target segment is in an isolated syllable.

It has also been observed in chapter 5 that there are some differences in the perception of nativeness in vowels and consonants. The third research question inquires further into this difference with particular focus on the behaviour of the four experimental cohorts: do foreign accented vowels and consonants have a different effect on the perception of nativeness across listeners differing in their L1?

7.2 Methods

7.2.1 Participants

Four cohorts of listeners differing in their L1 were recruited to take part in this experiment: English (n=22, mean age=23.5), Spanish (n=24, mean age=19.4), Czech (n=14, mean age=21.8) and Japanese (n=25, mean age=21.4). Cohorts were recruited at the Anglia Ruskin University in Cambridge, United Kingdom (English listeners), the University of the Basque Country in Vitoria, Spain (Spanish listeners), Charles University in Prague, Czech Republic (Czech listeners) and Waseda University in Tokyo, Japan (Japanese listeners). None of the participants had taken part in any of the previous experiments presented in this thesis.

Native English participants were monolingual British English speakers from the south-east of England, with no knowledge of Spanish and no regular contact with Spanish speakers either in Spanish or English. All participants in the non-native cohorts had a medium-high proficiency level in English (B2-C1 according to the Common European Framework of Reference for Languages, Council of Europe, 2001). The requirements for the recruitment of non-native participants were the same as in previous chapters. Part of the recruitment process included a pure-tone audiometry test to ensure there were no hearing problems among the participants. The audiometry consisted of a set of pure tones at different frequencies (viz. 125, 250, 500, 1000, 2000, 4000 and 8000 Hz) for each ear. The minimum hearing level (dB HL) for each of these values was collected. All participants passed the audiometric screening.

7.2.2 Materials

The set of continua generated for this experiment was formed from the same target segments and carrier words as used in the previous chapter (i.e. 12 consonants, 10 vowels and 4 words for each; see section 6.3.1 and tables 5.2 and 5.3 for the complete list of words). For each word a continuum of 5 iso-perceptual steps was generated (see section 6.2). Stimuli were generated using speech material from speaker F3, the collection of which was described of chapter 6. The final dataset consists of 88 continua (12 consonants × 4 words + 10 vowels × 4 words), each made up of 5 iso-perceptual steps, for a total of 440 stimuli (22 continua × 4 words × 5 steps).

7.2.3 Tasks

The experiment consisted of four perceptual tasks. Once these tasks were finished, members of the three non-native cohorts were asked to carry out two control tasks to assess their English language knowledge. Due to the number of tasks and tokens, the experiment was divided into two one-hour sessions. Participants were paid after completion of the tasks. The first session consisted of a word intelligibility task and an AX discrimination task. During the second session, participants underwent a segment identification task, a nativeness categorisation task and two control tasks measuring perceptual phonemic competence in English and familiarity with English words. All tasks are described in more detail below.

The experiment was carried out in four different locations: the Anglia Ruskin University in Cambridge for the British cohort, the Basque Country University in Vitoria for the Spanish cohort, the Charles University in Prague for the Czech cohort and the Waseda University in Tokyo for the Japanese cohort. The conditions of the testing room (a noise-attenuated room) and the equipment employed were similar in each location. A sample of the experimental interfaces used are provided in Appendix C.

Intelligibility task

Participants had to listen to each of the 440 words in the corpus and write down the word they heard. The instructions, as shown on-screen, were: “You will hear single words. Your task is to type the word you have heard.” The researcher verbally asked participants to always write a real word, and informed participants that the stimuli would not be repeated. Once the response was given, participants were instructed to press *return* so the next token could automatically start; in this way, the experiment was self-paced. For each token, the response and the reaction time were collected.

Stimuli were presented blocked by type of target segment (vowel or consonant) in pseudo-random order with no consecutive stimuli from the same continuum, as described in previous chapters. A 1-minute pause was introduced halfway through the task. A 5-stimulus practice session was included at the beginning of the task with tokens from different words of a [p]→[p^h] continuum. The completion time for this task was, on average, 19 minutes (English), 22 minutes (Spanish), 26 minutes (Czech) and 26 minutes (Japanese).

AX Discrimination task

For this task, the following instructions were presented on-screen at the beginning of the task: “You will hear two words. Your task is to say if they are pronounced exactly the same or if the pronunciation is different”. Participants were presented with pairs consisting of stimuli from two steps of the same continuum. For each continuum, 8 pairs of stimuli were presented: 4 pairs one step apart, 3 pairs two steps apart, and, for control reasons, one *Same* pair formed by the middle stimulus (i.e. step 3) presented twice. As the results of the discrimination task presented in section 4.4.3 revealed no major significant differences between upwards and downwards pairs, all the stimuli were presented in the upwards direction, as shown in table 7.2.

<table>
<tr><td>Same step</td><td colspan="12">3-3</td></tr>
<tr><td>One step apart</td><td colspan="3">1-2</td><td colspan="3">2-3</td><td colspan="3">3-4</td><td colspan="3">4-5</td></tr>
<tr><td>Two steps apart</td><td colspan="4">1-3</td><td colspan="4">2-4</td><td colspan="4">3-5</td></tr>
</table>

Table 7.2: Pairs presented for each continuum in the AX discrimination task. Step 1 refers to the non-native end of the continuum, while step 5 is the native end.

The total number of tokens in the discrimination task was 704 (22 segments × 4 words × 8 pairs). The task was presented in two blocks, one for consonants and another for vowels. The order of these two blocks was randomised across participants, and the tokens were presented in a pseudo-random order (no two pairs of the same continuum were presented consecutively). A 1-minute break was introduced halfway through each block and between blocks. The two stimuli of each pair were presented with a 0.5 seconds interval between them. The completion time for this task was, on average, 24 minutes (English), 22 minutes (Spanish), 26 minutes (Czech) and 22 minutes (Japanese).

Identification task

The identification task was the only one in this experiment in which stimuli consisted of syllables instead of words (CV for identification of consonants in word initial position and vowels and VC for identification of consonants in word final position). These syllables were generated from the 5-iso-perceptual-step continua tokens by extracting the required syllable from the words, e.g. the tokens of the *toad* continuum, [təʊd]→[^{h}əʊd], were truncated to extract only the first syllable, resulting in a [tə]→[t^{h}ə] continuum. To avoid an abrupt cut in the signal, which might be perceived as unnatural, syllables were smoothed with a half-Hamming window at the trimmed end. No interference from the lexical level was expected in this task, so only two words per target segment were included.

This task was divided into three blocks: one for the onset consonant continua, one for the vowel continua and one for the final consonant continua. For the consonant continua, a total of 90 CV syllables were presented to the participants (9 sounds × 2 words × 5 steps). Participants were then asked to choose, from the whole inventory of English consonants, which one corresponded to the consonant heard in the CV syllable. Similarly, for the vowel continua, 100 CV syllables were presented to the participants (10 sounds × 2 words × 5 steps). Participants had to choose, from the vowel repertory of English, the vowel pronounced in the CV syllable. For the final consonant continua, 30 VC syllables were presented to the participants (3 sounds × 2 words × 5 steps). Participants were also asked to choose the consonant heard. The specific instructions as written on-screen were: "You will hear a sequence of a consonant and a vowel. Choose the word which contains the consonant you heard from all the possibilities" (similar instructions were provided for the three types of task). These three blocks were presented in a random fashion across participants, and the stimuli within each block were pseudo-randomised so that no two contiguous tokens were part of the same continuum.

The interface of this task presented a button for each of the possible answers of the participants. On each button, an orthographic representation of the sound and an example word were displayed (tables 7.3 and 7.4).

B Beat	CH CHip	D Deer	DH THat	J Jump	F Fox
G Girl	H Home	K Key	L Lift	M Man	N Nose
P Pull	R Rat	S Sun	SH SHip	T Tea	TH Thief
V Very	W Wall	Y Yes	Z Zip	ZH meaSure	

Table 7.3: Orthographic representations and word examples for the inventory of English consonants.

Listeners undertook a practice session consisting of tokens from a native English speaker pronouncing each of the target syllables. If the correct answer was not selected, the participant had the chance to repeat the sound as many times as desired. In this way intra-language variability could be avoided to some extent, since participants were aware

A rat	UH gun	AR park	
ER shirt	E pet	I sit	EE need
O lot	OR short	U push	OO goose

Table 7.4: Orthographic representations and word examples for the inventory of English vowels.

from the beginning of the task the nature of the linguistic variant that they were being exposed to.

Nativeness task

In the nativeness categorisation task the spelling of the word was presented on-screen, so it was important for it to be the final task in the experiment, avoiding as much as possible a learning effect on the participants. The categorisation task was a two-alternative forced choice task, in which participants were asked to decide whether each of the 440 tokens of the corpus was pronounced with a foreign accent or with a native accent.

Stimuli were presented in two blocks (one for consonants and one for vowels), and these two blocks were presented in a random order across participants. Inside each block, stimuli were presented in a pseudo-random order, with no consecutive same-continuum steps. Participants were aware that stimuli could not be repeated, and that once one of the two options was selected, the next stimulus would automatically be reproduced. Participants had a 1-minute break between blocks.

Postprocessing

As in previous chapters, those responses whose reaction time was below 300 ms or above 5000 ms were removed from the analysis (<1% for all tasks and cohorts). For the analysis of the intelligibility results, homophonic words were detected by comparing the phonetic transcription of the target word and the typed word. The transcriptions employed to this end were those included in the BEEP dictionary (Beep, 2017). As in chapter 3, if the target segment was correct, the answer was considered good (e.g. the word *tool* was considered a correct answer to the token *told*, as the target segment [t^h] was correctly perceived).

Statistical analyses reported in the results section of this chapter (section 7.3) were performed using R (R Core Team, 2017). Since responses in the four tasks are binary (correct/incorrect identified segment, same/different stimuli or foreign/native token), results were analysed with the general linear mixed model function *glmer* included in the *lme4* package (Elzhov et al., 2016). Post-hoc analysis was performed using FLSD pairwise comparisons via the same software.

Two control tasks were included in the experiment: a perceptual phonemic competence task and a familiarity task. Only non-native cohorts (i.e. Spanish, Czech and Japanese) were asked to fulfil the control tasks, as the goal of these tasks was mainly to ascertain that the participants had a comparable level of competence in the target language (i.e. English).

In the phonemic competence task, listeners heard unmanipulated CVC and VCV clusters pronounced by a native English speaker. Participants were asked to identify the central segment (the vowel in the CVC token and the consonant in the VCV token) by choosing the corresponding sound from the English phoneme inventory. The interface and procedure for this task were similar to those used for the identification task (see section 7.2.3), with the exception that, in the case of the phonemic competence task, participants did not receive any feedback and there was no practice session.

Similar capabilities of segment identification were detected for the four cohorts. Some marginal differences were detected in non-experimental segments [segment [l], Japanese-Spanish = $p < .05$; segment [ʒ], Czech-Spanish and Japanese = $p < .001$]. The only experimental segment for which significant differences across cohorts were found was [iː] [Spanish-Czech and Japanese = $p < .01$].

The second control task was a lexical familiarity task. Participants were asked to rate, from 1 to 7, how familiar they were with the words included in the experiment, which were presented orthographically in the visual modality only (i.e. no sound was reproduced during this task). The scale presented to the participants was accompanied by the labels in table 7.5.

Results of the familiarity task were analysed separately for words carrying a consonant or a vowel as a target segment (table 7.6). No significant differences were found between cohorts either for the consonant words or the vowel words, but Czech and Japanese listeners found significantly more familiar vowel words than consonant words [$p < .05$ for both cohorts].

1	I have never seen this word before
2	I might have seen this word before
3	I am nearly sure I've seen this word before but I'm not certain
4	I recognise the word because I've seen it before, but I don't know its meaning
5	I have seen this word before but I only have a vague idea of its meaning
6	I think I know the meaning of the word but I'm not sure it's the correct one
7	I know the word and I'm sure I know its meaning

Table 7.5: Labelling of the 7-point scale in the familiarity task.

Type of continuum	Spanish	Czech	Japanese
Consonants	6.1 (1.77)	6.08 (1.8)	6 (1.95)
Vowels	6.29 (1.6)	6.39 (1.44)	6.23 (1.75)

Table 7.6: Mean familiarity reported by participants of the three cohorts. Values in parentheses are the standard deviations.

7.3 Results

Segment identification in isolated syllables, word intelligibility and nativeness categorisation are analysed for consonants and vowels separately. Discrimination results for the 22 continua are analysed for the 1 step apart and the 2 steps apart pairs separately.

7.3.1 Consonants

The overall results for the consonant continua (figure 7.1) show a monotonic increase in the score for segment identification, word intelligibility and nativeness categorisation tasks. This means that, as the stimuli become more native-like, listeners of the four cohorts identified the target segments more easily both in isolated syllables and in words, and perceived them as more native-like.

A GLMM with *task* (2 levels) and *cohort* (4 levels) as fixed factors and individual results as a random factor revealed that the occurrence of an accented segment in the context of a word as opposed to in an isolated syllable resulted in a significant detriment in the scores of the non-native cohorts [Spanish = $p < .01$; Czech = $p < .001$; Japanese

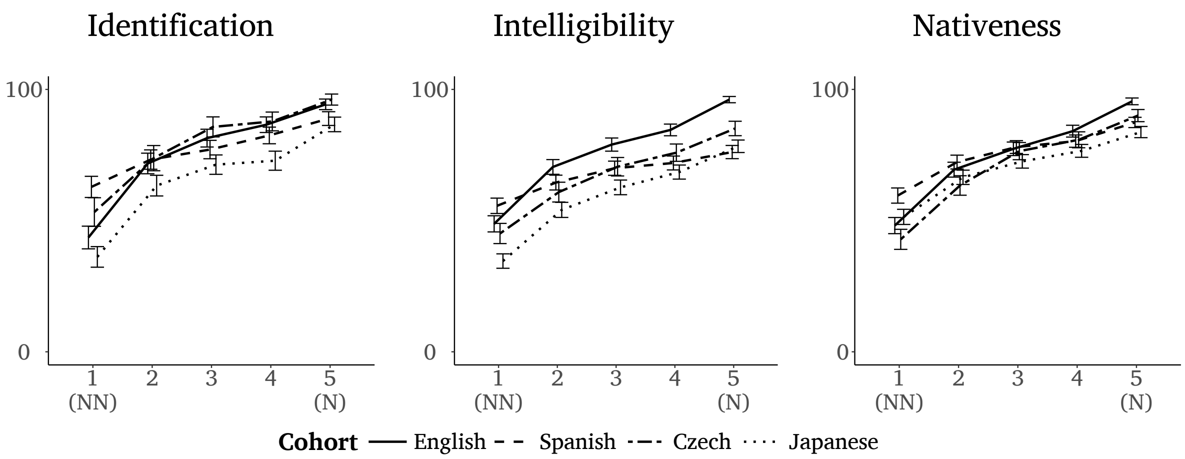


Figure 7.1: Overall results for consonant continua in segment identification on isolated syllables, word intelligibility and nativeness categorisation tasks. Here and elsewhere vertical bars represent ±1 standard error and results are jittered to prevent overplotting.

= $p < .05$], an effect that was not observed in native listeners. This finding suggests that the effort required to decode the signal at its lexical level penalised non-native listeners, but not native listeners, for whom lexical information is more easily accessible.

Regarding the results for individual segments (figures 7.2, 7.3 and 7.4), different patterns were detected in the segment identification responses of the cohorts. The four cohorts identified all steps of the [x]→[h] continuum as [h] regardless of the degree of accentedness. Out of the 12 continua, the four cohorts showed an increment from the non-native end to the native end in correct segment identification in 6 of them (see row 2 of table 7.7). No significant differences were found between cohorts in any step of these 6 continua. The three non-native cohorts perceived the [t]→[t^h] continuum as native in its entirety, while native listeners had trouble identifying the non-native end of the continuum as [t^h]. Two different trends were detected in [r]→[ɹ], [k]→[k^h] and [_x]→[_g]: English and Japanese listeners were unable to correctly identify the initial steps, while Spanish and Czechs identified the native target segment along the whole continuum, no matter the degree of foreign accent. Similarly, in [s]→[z], no significant increment in segment identification was detected for any cohort except for the Japanese, for whom the non-native end was rarely identified as [z]. These trends are summarised in table 7.7.

Segment identification scores in isolated syllables did not always coincide with word intelligibility results, even though only the target segment was accented. As already seen in the across-consonant results (figure 7.1), lexical factors appear to have influenced the correct identification of foreign-accented segments. By analysing the results of individual

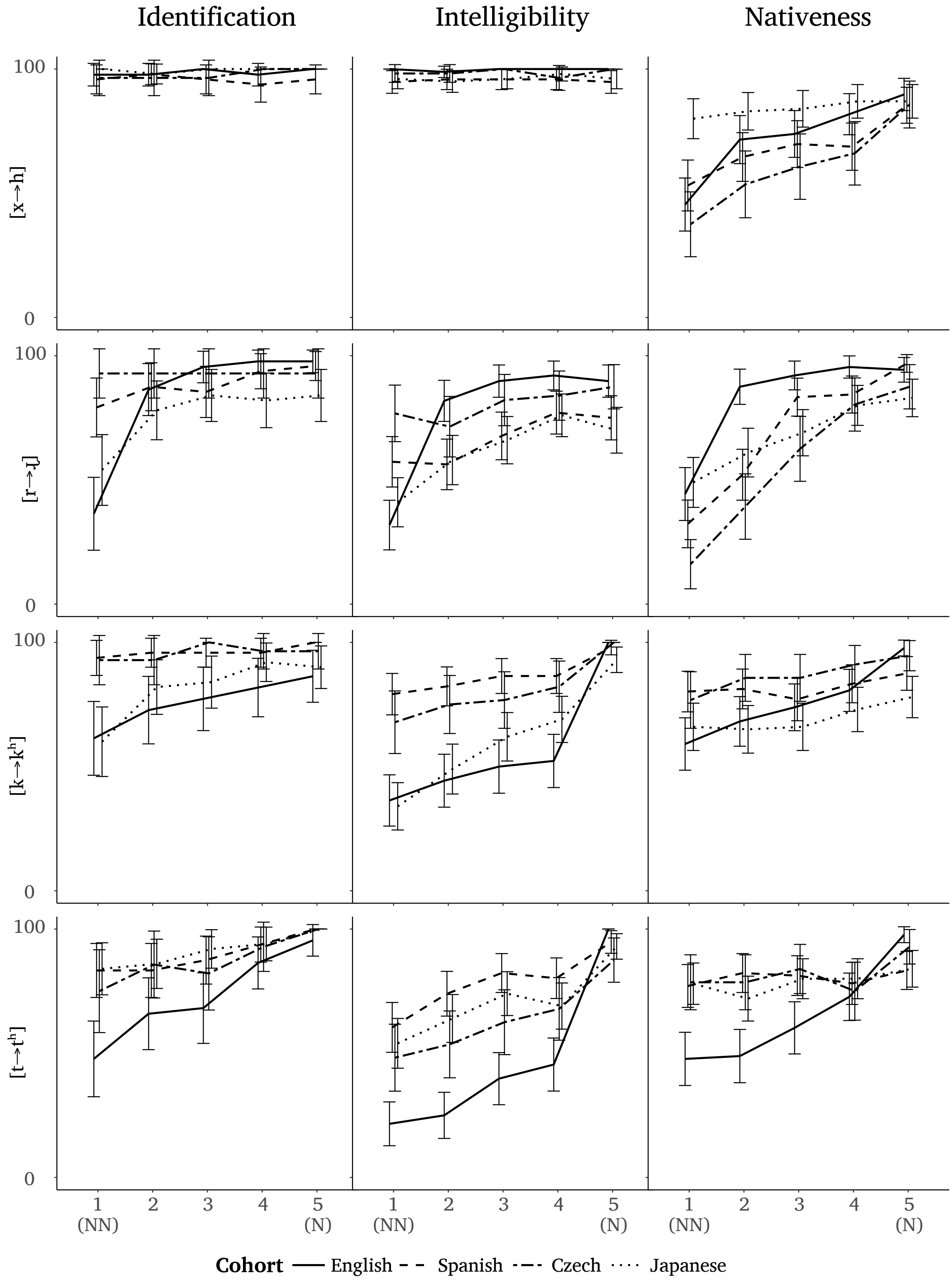
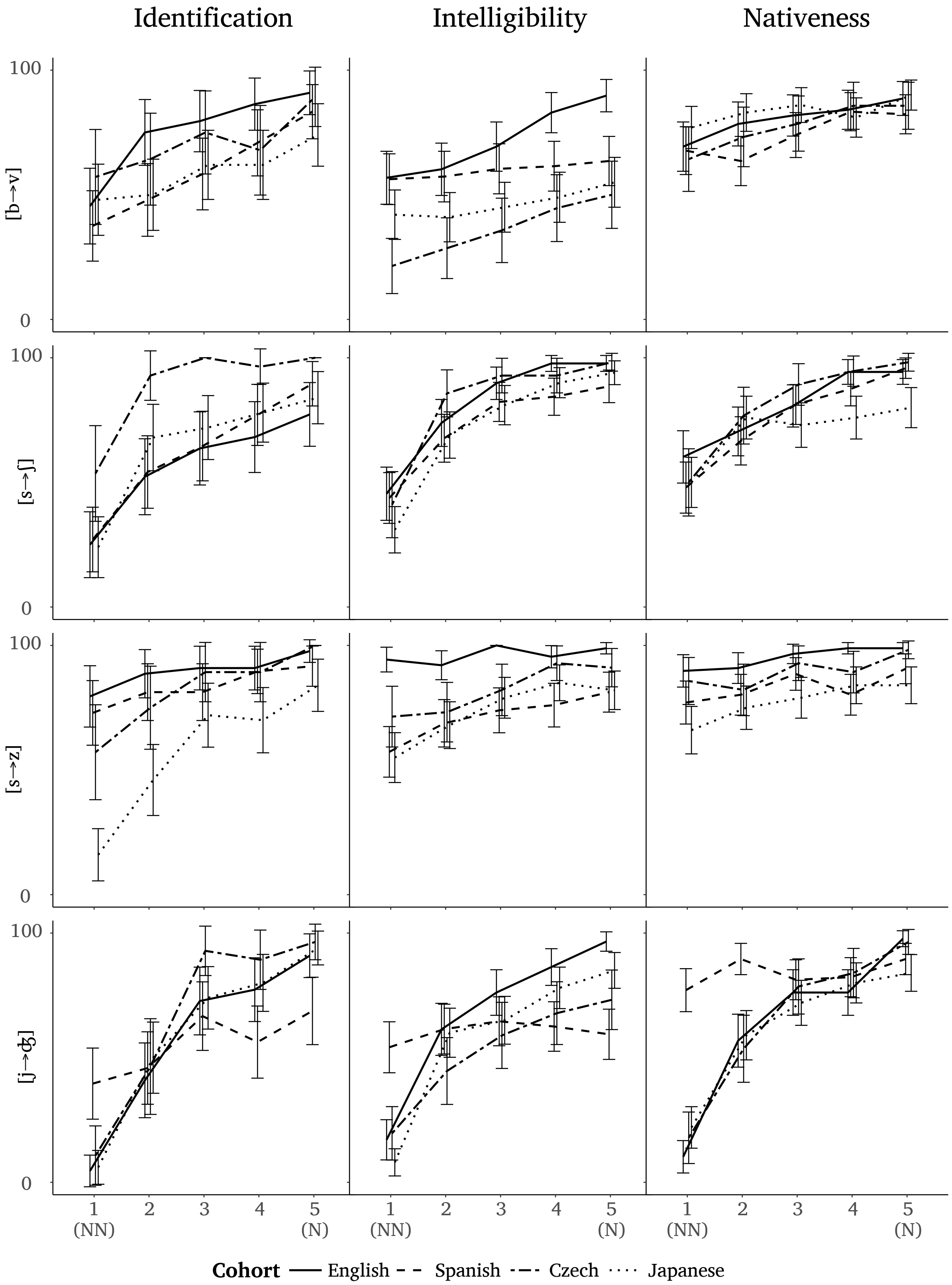


Figure 7.2: This figure and figures 7.3 and 7.4: Nativeness categorisation results for individual consonant continua for the four experimental cohorts.



Identification
Intelligibility
Nativeness
100
0
[b→v]
[s→ʃ]
[s→z]
[j→ʤ]
1 (NN)
2
3
4
5 (N)
Cohort English Spanish Czech Japanese

Figure 7.3

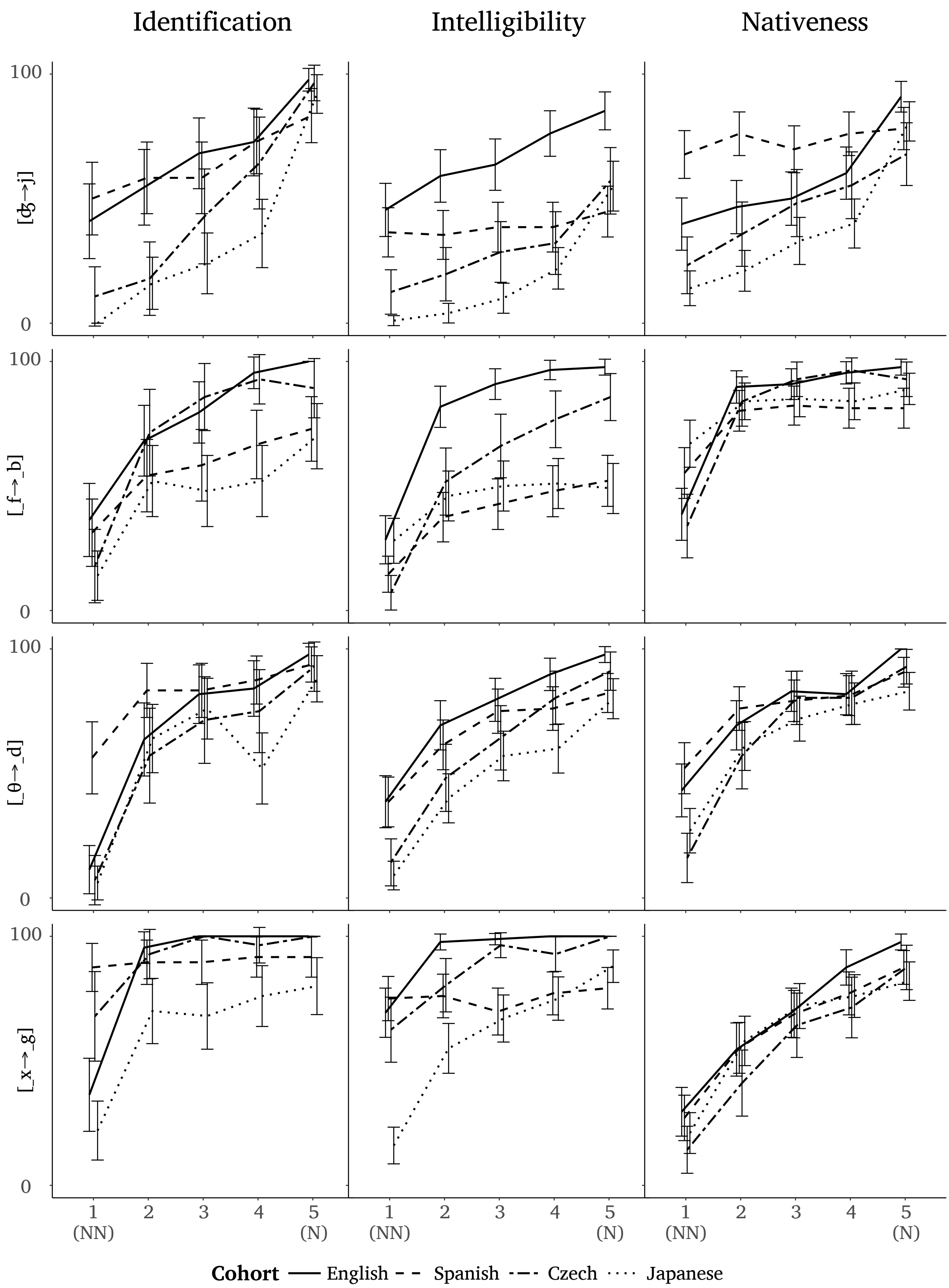

Identification
Intelligibility
Nativeness
100
0
[ʤ→j]
[_f→_b]
[_θ→_d]
[_x→_g]
1
(NN)
2
3
4
5
(N)
Cohort
English
Spanish
Czech
Japanese

Figure 7.4

Continuum	Natives	All non-natives
[x→h]	—	—
[b→v] [s→ʃ] [j→ʤ] [ʤ→j] [_f→_b] [_θ→_d]	/	/
[t→tʰ]	/	—
	English & Japanese	Spanish & Czech
[r→ɹ] [k→kʰ] [_x→_g]	/	—
	English, Spanish & Czech	Japanese
[s→z]	—	/

Table 7.7: Syllable-isolated segment identification trends for the four cohorts in consonant continua. The patterns represent those continua in which the target segment is correctly identified along the whole continuum (—) and those in which the non-native end does not reach the upper tercile (/).

continua, we found that this influence could be either positive (when the target segment is more often correctly identified in a word), or negative (when the appearance of the target segment in a word results in lower identification scores). For brevity in what follows, the term ‘lexical effect’ will be used to refer to the influence of the lexical context in the correct identification of a foreign-accented segment.

According to the comparison of the syllable-isolated segment identification task results and the word intelligibility task results, the English listener cohort benefitted most from the lexical context. A negative lexical effect was detected for English listeners only in the two continua involving plosives ([k]→[kʰ] and [t]→[tʰ]). This drop in correct segment identification was located in the first 4 steps of these continua, which means that native listeners found it especially problematic to identify the foreign accented counterpart of [tʰ] and [kʰ] when presented in words. A positive lexical effect was found in the first steps of [s]→[z], [_f]→[_b] and [_θ]→[_d] continua, and in [s]→[ʃ] a positive lexical effect was observed in the final steps of the continuum.

Spanish listeners suffered a significant drop in identification of the [ɹ] segment along the whole continuum when in lexical context. In the [ʤ]→[j] continuum, when the target segment was in an isolated syllable, Spanish listeners were able to correctly identify the native end more than the 80% of the time. Accuracy dropped to values below 50% when in lexical context, which implies a negative lexical effect for this contrast among Spanish listeners. Similarly, it is worth mentioning that the identification scores of the [b]→[v] continuum dropped significantly when in lexical context for Czech [$p < .001$] and Japanese [$p < .05$] cohorts, i.e. as we observed in the [ʤ]/[j] contrast for Spanish, the [b]/[v] contrast was also more salient when in isolated syllables for the Czech and Japanese groups.

Some of the effects detected among Spanish listeners were also observed in the Japanese cohort. The above-mentioned contrasts (namely [ʤ]/[j] and [b]/[v]) were also better perceived in isolated syllables. As in the English cohort, Japanese listeners also suffered a significant drop in segment identification in continua involving plosives (i.e. [k]→[kʰ] and [t]→[tʰ]). They were the group that showed most benefit when the non-native end of the [s]→[z] continuum appeared in a lexical context.

Any lexical effect – which was generally positive – typically occurred in the initial steps of the continua for all cohorts, i.e. the lexical context appears to have a more important role the more foreign-accented the segment is. Nevertheless, in some cases, the lexical effect was detected in the final steps of a continuum, e.g. for Spanish and Japanese listeners in the [b]→[v] and [ʤ]→[j] continua. In these particular cases, the word interfered with correct identification of the non-accented target segment.

The iso-acoustic to iso-perceptual transformation of continua worked as was expected in most of the cases, i.e. English listeners perceived the majority of the continua as formed from equally-sized steps in terms of nativeness. Significant differences between contiguous steps were only found between steps 1 and 2 of [r]→[ɹ] and [_f]→[_b] [$p < .001$] and, to a lesser extent, between steps 4 and 5 of [ʤ]→[j] [$p < .05$]. The inter-cohort analysis for the nativeness categorisation task revealed that Japanese listeners perceived the first step of the [x]→[h] continuum significantly [$p < .05$] more native than the other cohorts, which means that the [x]/[h] contrast was not as salient as it was for English, Czech and Spanish listeners. A significant difference was also found in the way English listeners categorised the initial steps of the [t]→[tʰ] continuum, identifying them as much less native than the other three cohorts. This effect was not significant in [k]→[kʰ], even though a similar trend can be observed. The final steps of [s]→[ʃ] and the initial steps of [s]→[z] were perceived as significantly [$p < .05$] less native by Japanese listeners. In [j]→[ʤ] and [ʤ]→[j] Spanish listeners perceived the first two steps as significantly [$p < .001$] more native than the other three cohorts, which means that, as observed in previous chapters, they did not consider any step of these continua to be foreign-accented. Finally, regarding consonants in final position, only a marginal difference between Spanish and Czech listeners was found in the first step of the [_θ]→[_d] continuum [$p < .05$].

7.3.2 Vowels

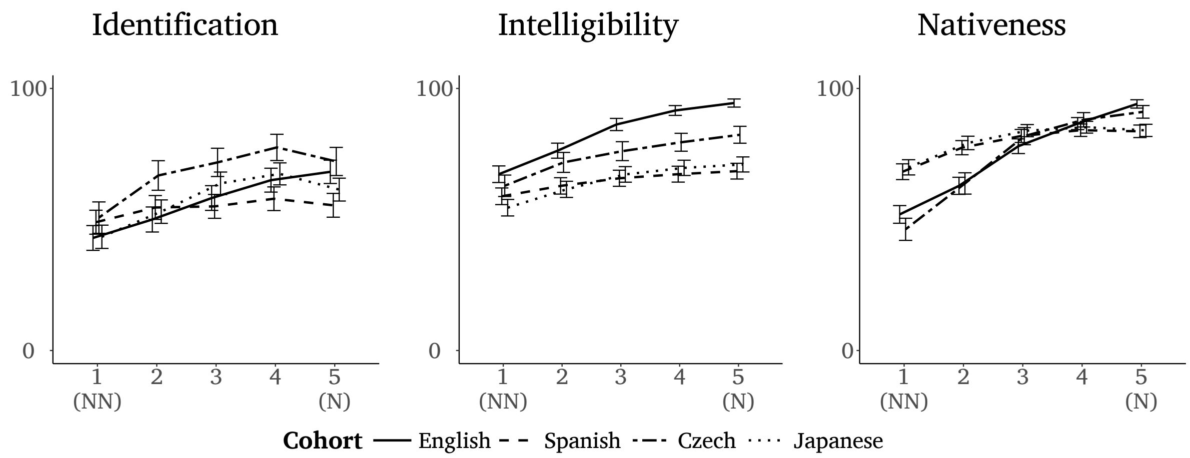


Figure 7.5: Overall results for vowel continua in segment identification in isolated syllables, word intelligibility and nativeness categorisation tasks.

The overall results for vowel continua (figure 7.5) show increases in identification scores as the tokens become more native-like both in isolated syllables and in words. This increment was more pronounced for English and Czech listeners in both tasks, i.e. the difference in the identification scores of the first and last steps of the continua is bigger for English and Czech listeners than for Spanish and Japanese. The latter pair of cohorts produced lower scores in the intelligibility task along the whole continuum of foreign accented vowels, particularly on steps 4 and 5, in which significant differences with the Czech [$p < .01$] and English [$p < .001$] cohorts were detected.

The four groups perceived an increment in nativeness, but Japanese and Spanish listeners perceived the first steps as significantly more native than English and Czech listeners [$p < .01$]. In the light of these results, we interpret that the contrast between native English vowels and their corresponding foreign accented counterpart was not as salient for Spanish and Japanese listeners as it was for English and Czech listeners.

A lexical effect can also be observed in the overall results for vowel continua. Unlike the effect seen for consonant continua, in the case of the vowels it was not that the non-native cohorts suffered a detriment in segment identification when the target segment was in a word, but rather we found that English listeners improved their scores significantly [$p < .01$] in the lexical context. No significant differences between tasks were found for any other cohort.

Continuum	Natives	Non-natives
[a→æ] [o→ɒ]	‾	‾
[o→ɜː] [o→ɔː] [u→ʊ]	_	_
[e→ɜː] [i→iː]	/	/
[u→uː]	‾	/

	English, Spanish & Czech	Japanese
[a→ɑː]	/	_

Table 7.8: Segment identification trends for the four cohorts in the vowel continua. The patterns represent those continua in which the target segment is correctly identified along the whole continuum (‾), those in which both ends of the continuum are correctly identified below the upper tercile (_) and those in which the non-native end is identified below the upper tercile (/).

Regarding the results for individual vowel continua (figures 7.6, 7.7 and 7.8), the trends of segment identification in isolated syllables were not constant across segments, but more agreement between cohorts was found than in consonants. Three vowel continua

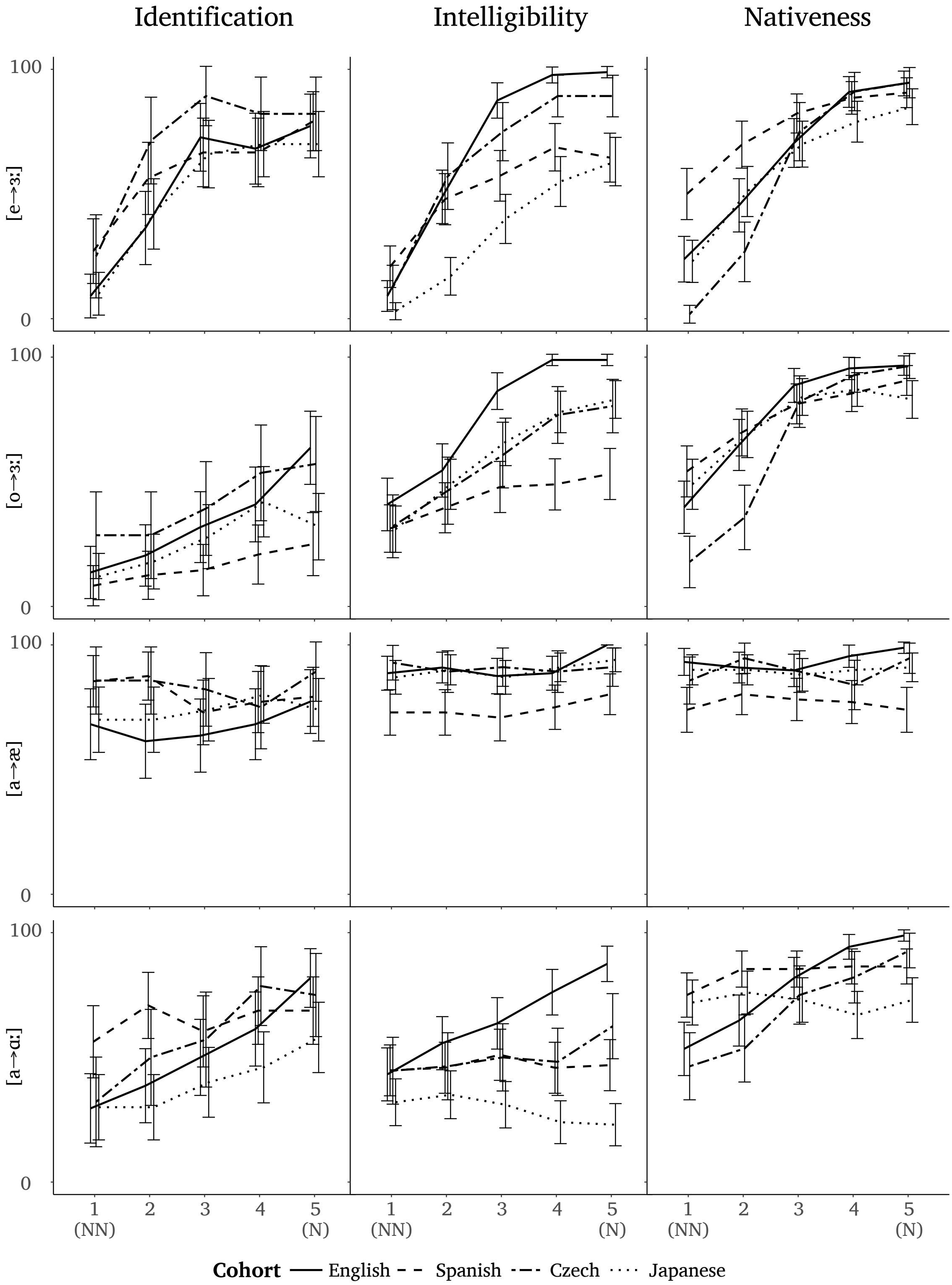
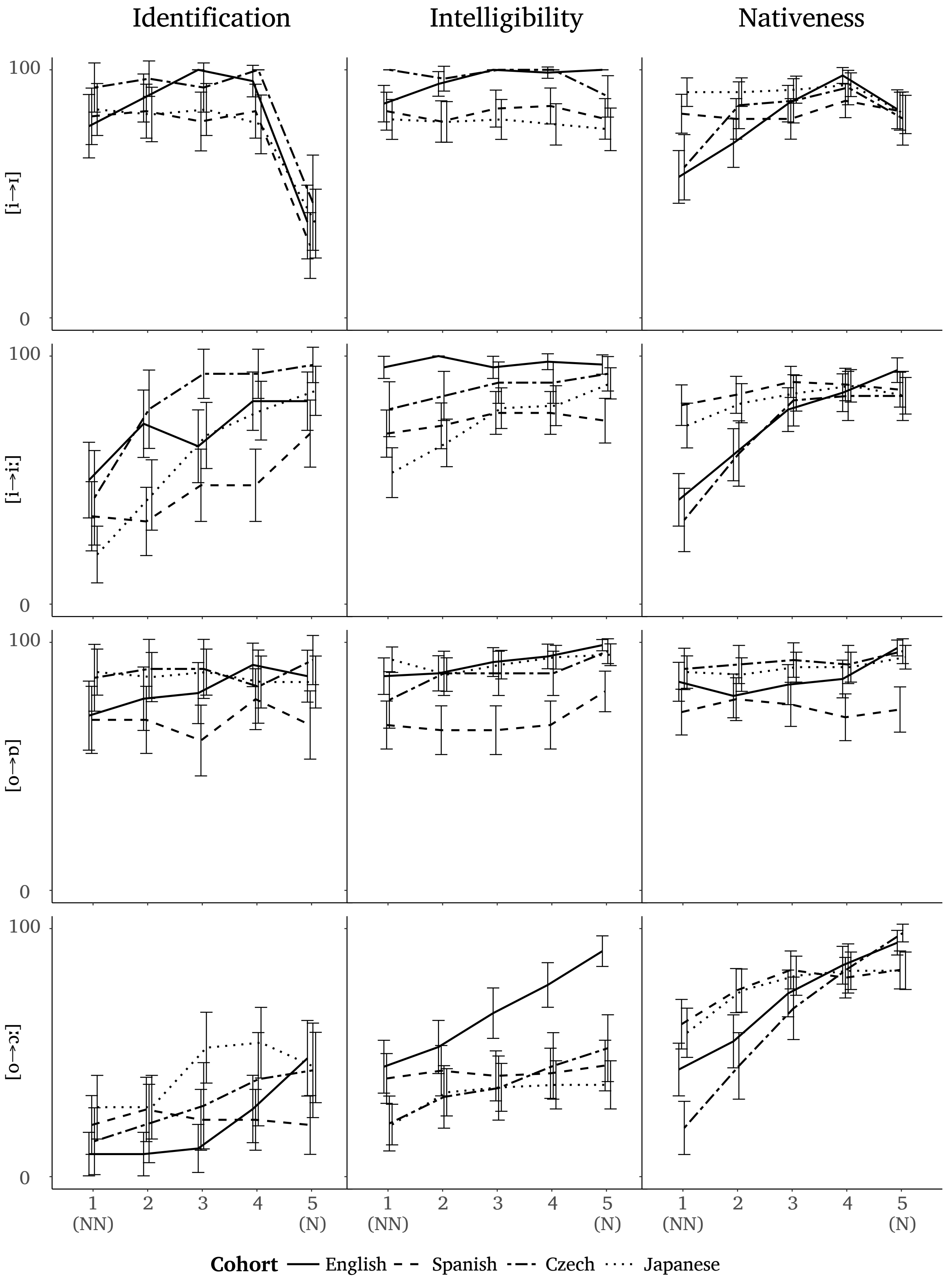


Figure 7.6: This figure and figures 7.7 and 7.8: Nativeness categorisation results for individual vowel continua for the four experimental cohorts.



Identification
Intelligibility
Nativeness
[i→ɪ]
[i→iː]
[o→ɒ]
[o→ɔː]
100
0
1 (NN)
2
3
4
5 (N)
Cohort — English – – Spanish –·– Czech ··· Japanese

Figure 7.7

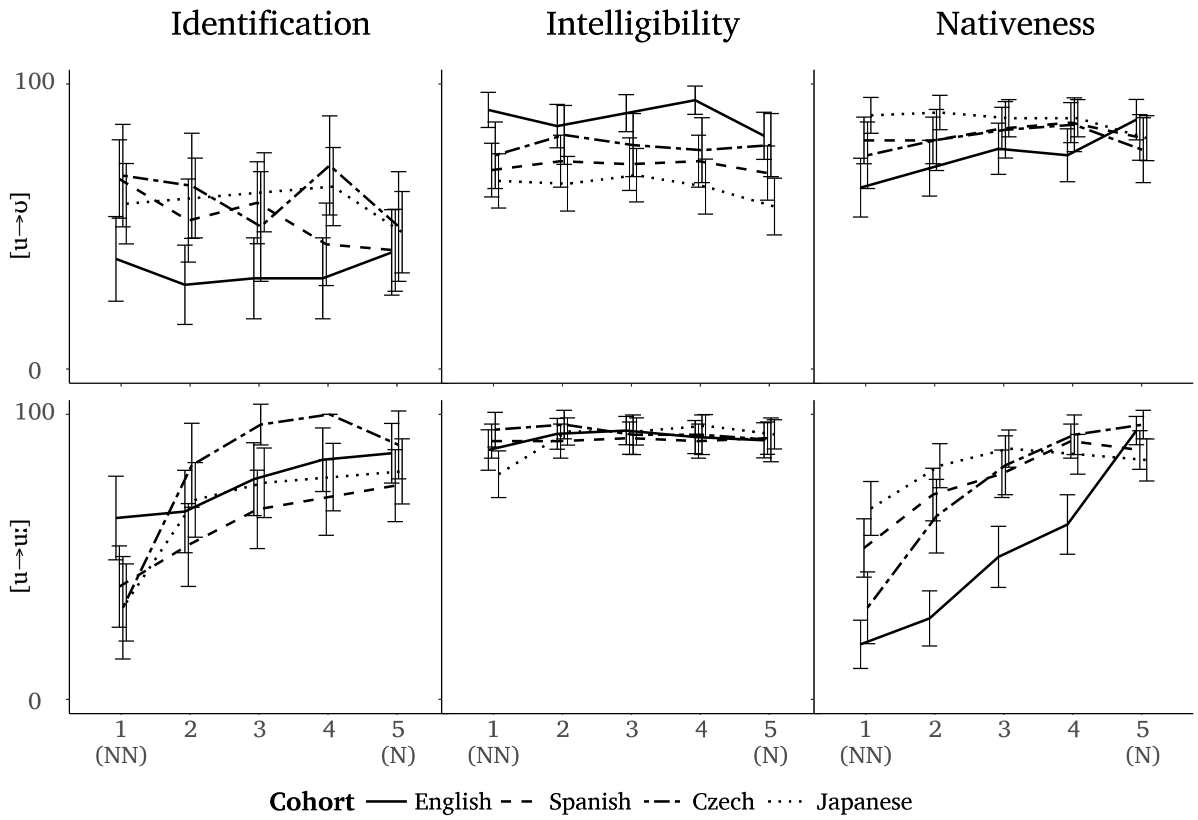


Figure 7.8

(namely [o]→[ɜː], [o]→[ɔː] and [u]→[ʊ]) did not reach 70% correct identification by any group at the native end of the continuum. Even native listeners failed to reach the upper tercile in identification, which may imply that the speaker's pronunciation was not accurate enough to be identified as the native segment in isolated syllables, but good enough to be identified as native in lexical context. Nevertheless, English listeners identified correctly the native end of [o]→[ɜː] significantly more than Spanish [$p < .01$] and Japanese [$p < .05$] listeners. A significant difference was also found in the third step of the [o]→[ɔː] continuum between Japanese and English listeners [$p < .001$]. This may imply that Japanese listeners have a wider [ɔː] category in their L2 phonological system, which makes them able to identify this segment with a higher degree of deviation from the native prototype than native listeners, who have a more restricted space for this category in their L1 phonological system.

Two continua, [a]→[æ] and [o]→[ɒ], were correctly identified at each step by all cohorts. As we observed in the results of nativeness categorisation earlier in this thesis, these two continua never conveyed foreign accent, even at their non-native ends.

Listeners from all cohorts reported a clear change in the segment identified at both ends of [e]→[ɜː] and [i]→[iː]. No significant differences were found between cohorts for the former, but in the central steps of the latter (i.e. steps 2, 3 and 4 of the [i]→[iː] continuum) a significant difference was found between Spanish and Czech listeners [$p < .01$]. This central area was identified as [iː] by Czech listeners more often, which reveals that long vowels are much more prominent for them than for Spanish listeners.

There was a difference in the segment identification pattern in isolated syllables by English listeners compared to non-native listeners in the [u]→[uː] continuum. Even though the four cohorts identified the native end correctly, English participants were the ones who correctly identified the non-native end more frequently. The scores of native listeners regarding this continuum contrast with their identification rate of [u]→[ʊ]. These two continua share the same non-native segment (namely [u]) which, in the light of these results, appear to be perceived in isolated syllables as [uː] rather than as [ʊ].

Japanese listeners were the only ones who did not reach the upper tercile in correct segment identification in the [a]→[ɑː] continuum, hence we classified their results as flat non-native. Even so, it is worth mentioning that no significant differences among cohorts were found in this continuum, and the highest rate of correct segment identification in the native end was 82% (English listeners), which makes us think that, every cohort found problems identifying the [ɑː] segment when in isolated syllables.

Finally, results for the [i]→[ɪ] continuum were not included in table 7.8 because the pattern of segment identification in isolated syllables was unexpected. The four groups identified steps 1-4 as [ɪ], but in step 5 the identification scores dropped below 50% for all cohorts. The possible reason for this will be discussed in section 7.4.

Comparing these results with the responses of the word intelligibility task, we detected, as in the consonants, a lexical effect in correct segment identification. As a general impression, the appearance of a segment in lexical context had a positive effect on its correct identification among all cohorts. As we observed in consonants, nevertheless, English listeners benefitted the most from the lexical context: a positive lexical effect was found for them in every continuum except in [a]→[ɑː] (in which no effect was found for English listeners, while the other cohorts suffered a negative lexical effect) and [o]→[ɒ].

English listeners produced a significant positive lexical effect in the three continua whose native end was not successfully identified in isolated syllables (namely [o]→[ɜː], [o]→[ɔː] and [u]→[ʊ]). As we observed in [s]→[ʃ], the lexical context helped native listeners identify a non-accented segment which, in isolated syllables, was not correctly identified. On the other hand, none of the non-native groups reported a lexical effect in [o]→[ɔː]; only for [o]→[ɜː] (Spanish and Japanese) and [u]→[ʊ] (Spanish) a positive lexical effect was detected. No lexical effect was detected for Czech listeners in any of these three continua.

All non-native cohorts suffered a significant detriment in correct identification of the native end of the [a]→[ɑː] continuum when placed in word context. It seems that, when devoid of lexical context, duration becomes an important cue for non-native listeners vowel identification, but the placement of the [ɑː] segment into a word makes them more hesitant and more prone to perceive it as [æ] or [ʌ]. This negative lexical effect was not observed among native listeners.

We observed a positive lexical effect among all cohorts in the first steps of [i]→[iː] and [u]→[uː]. In syllable identification, the most frequent confusions were the corresponding short vowels (namely [ɪ] and [ʊ]), but the lexical context provided enough information for each cohort to correctly identify the target segment.

Finally, a significant positive lexical effect was found for all cohorts in the final step of the [i]→[ɪ] continuum. The unexpected results of the syllable-isolated segment identification were not mimicked in lexical context, where all cohorts identified the target segment correctly at least 80% of the time for every step of the continuum. It is worth noting that, except for the two ends of the continuum and in spite of the high identification rate of all cohorts, English and Czech listeners scored significantly higher on intelligibility than Spanish and Japanese listeners [$p < .05$].

Regarding the nativeness categorisation results, all vowel continua were perceived as non-categorical by native listeners, as expected by the iso-perceptual transformation. Only marginal significance was found between steps 4 and 5 of the [u]→[uː] [$p < .05$], but no other pair of contiguous steps was perceived as significantly different by the English cohort in terms of nativeness. No differences were found in the perceived nativeness among the four cohorts in the [o]→[ɒ] and [u]→[ʊ] continua, and only a marginal difference between English and Spanish listeners [$p < .05$] was found in the last step of the [a]→[æ] continuum. Czech listeners perceived the first two steps of the [o]→[ɜː] continuum as significantly less native-like than Spanish and Japanese listeners [$p < .01$] and the first two steps of the [e]→[ɜː] continuum significantly less native than Spanish [$p < .01$]. In the remaining five continua ([a→ɑː, i→ɪ, i→iː, o→ɔː and u→uː]) the four listener cohorts were grouped into two clear blocks: on one hand, English and Czech listeners, for whom the perception of the continuum was incremental and both ends were perceived with significantly different nativeness (i.e. there was a change in category along the continuum). On the other hand, Japanese and Spanish listeners, whose nativeness ratings were usually more flat-native. In the case of the latter cohorts, only in two cases a significant difference in perceived nativeness was detected between both ends: Spanish listeners in [u]→[uː] [$p < .001$] and Japanese listeners in [o]→[ɔː] [$p < .01$].

Discrimination

Regarding the results of the discrimination task, no significant differences across cohorts were found for the *same step* trials in a 4-sample test for equality of proportions without continuity correction, and the success rate in this particular pair was above 94% for all cohorts. The results of the discrimination task for consonant and vowel continua (figures 7.9 and 7.10 respectively) revealed, as a broad observation, agreement between the four cohorts in the general shape of the discrimination curves and, particularly, in the discrimination peak of each continuum. A clear exception to this general rule is the behaviour of Japanese listeners in the [s]→[z] continuum: even though in the 1 step apart trials the four groups agreed with the discrimination pattern, in the 2 steps apart trials Japanese perceived the discrimination peak between steps 1 and 3, while the other three cohorts discriminated better steps 3 and 5.

A GLMM was generated with three fixed factors (*continuum* = 22 levels, *listener cohort* = 4 levels and *condition* = 2 levels —1 step apart and 2 steps apart pairs—) and individual responses as random factor. Regarding consonants, significant differences were found between the two experimental conditions (1 step apart and 2 steps apart) in the results of the [_x]→[_g] continuum for Spanish listeners [$p < .05$], Czech listeners [$p < .05$] and Japanese listeners [$p < .001$]. Significant differences were also found across conditions in [r]→[ɹ], [s]→[z] and [_θ]→[_d] [Japanese = $p < .05$] and [j]→[ʤ] [English = $p < .05$]. The inter-cohort analysis revealed significant differences in the 2 steps apart condition between Japanese and Czech listeners in [b]→[v] [$p < .01$], between Japanese and English listeners in [s]→[z] [$p < .01$] and between Japanese and Spanish listeners in [ʤ]→[j] [$p < .01$].

In vowel continua, on the other hand, all groups reported significant differences across experimental conditions in the discrimination task for the [e→ɜː, o→ɜː, i→iː, o→ɔː and u→uː] continua [$p < .001$]. Spanish listeners were the only ones who did not discriminate differently across conditions in the [a]→[ɑː] continuum [$p < .001$ for the other three cohorts],

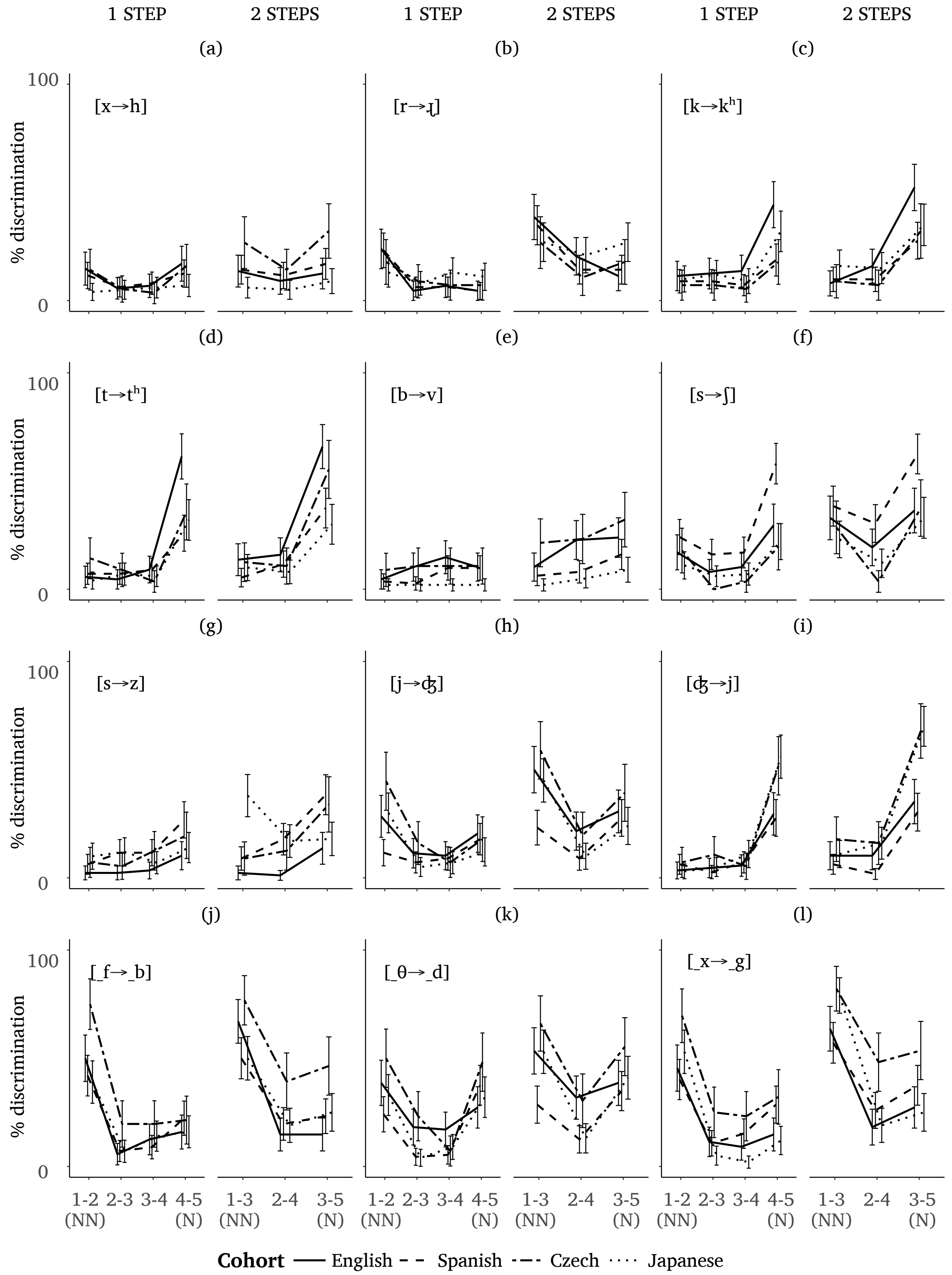


Figure 7.9: AX discrimination results for consonant continua.

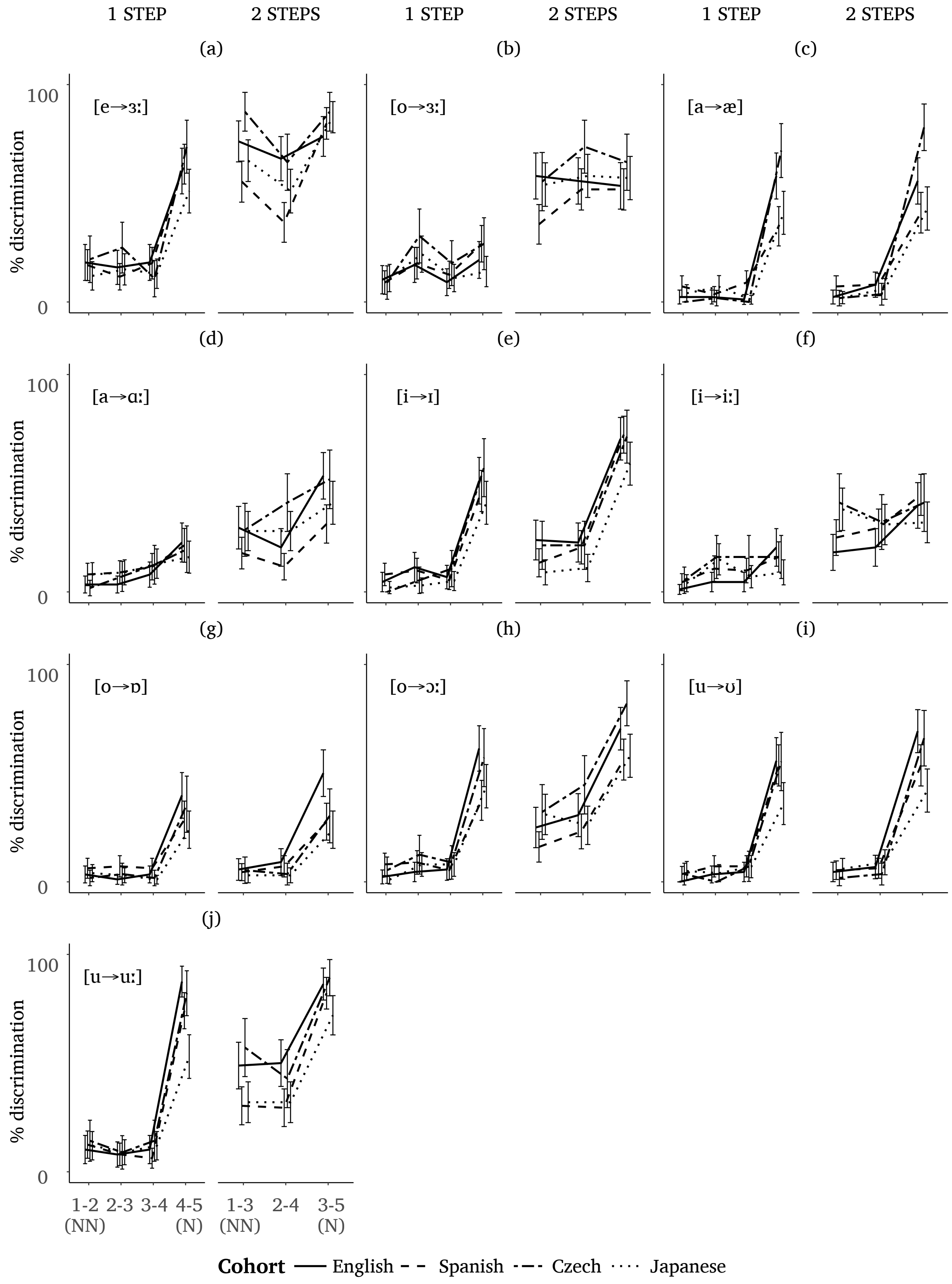


Figure 7.10: AX discrimination results for vowel continua.

and significant differences were found in [i]→[ɪ] [English = $p < .001$; Spanish = $p < .01$; Japanese = $p < .05$]. There were no significant differences across experimental conditions for [a]→[æ], [o]→[ɒ] and [u]→[ʊ]. No significant differences were found between cohorts in any vowel continuum nor experimental condition.

Generally, it can be said that cohorts behave similarly from one experimental condition to the other in the consonant continua, i.e. they exhibited similar discrimination capabilities in 1 step apart pairs than in 2 steps apart pairs. However, this is not the case for the vowel continua, in which the 2 steps apart condition elicited a higher degree of discrimination mostly in continua involving duration as a phonemic cue. Only Japanese listeners behave slightly different than other cohorts in a few consonant continua, but these differences are not sufficient to claim that the discrimination capabilities of certain listener cohorts are better than others.

7.4 Discussion

This chapter describes an experiment consisting of a series of perceptual tasks intended to evaluate the realisation of English segments varying in their degree of Spanish foreign accent. Listeners with four different L1s (English, Spanish, Czech or Japanese) took part in this experiment, assessing the stimuli for segment identification, word intelligibility, AX discrimination and nativeness.

As a general first finding, the results of the segmental approach suggest that the impact of foreign accent is highly dependent on the analysed segment and its representation in the phonological system of the listener. This complexity is hidden in holistic studies of accent and suggests that the overall outcomes seen in such studies do not represent patterns at the level of individual segments. When several FA traits are analysed in connected speech (words, sentences or other utterances), their effects may neutralise one another. This can result in linear FA effects which are generally correlated to the degree of foreign accent. In the light of our results, we argue that segments have specific weights on the perception of foreign accent, not necessarily triggering linear-like perception when analysed individually.

RQ1: What is the effect of the L1 segmental system of listeners in the perception of segmental foreign accent?

As the main focus of this thesis is to expand knowledge regarding segmental foreign accent, in the first research question we inquired into the perception of foreign accented segments by listeners with different phonological systems. To neutralise the effect of lexical knowledge on non-native listener populations we developed a syllable-isolated segment identification task, which allowed us to analyse the behaviour of listeners towards accented segments in a lexical-free context.

The results suggest that the perception of foreign accented segments is not always equal across cohorts. In this chapter it has been demonstrated that the consideration of a [non-native]/[native] contrast is highly dependent upon the phonological system of the listener, even in contexts that are free of lexical constraints. Overall, our first research question can be answered by saying that the production of foreign accented segments does not affect perception equally for listeners with different L1s, or, in other words, the ability of listeners to assign the correct native category to accented segments is highly dependent upon their first language.

One of the main findings related to this question is that, in some cases, listeners are able to identify the target segment even when it is highly accented. This is the case for [x]→[h]: even though listeners were able to identify this segment as [h], they reported a high degree of accentedness at the non-native end of the continuum. Previous holistic studies have demonstrated in sentences that a strong foreign accent does not always lead to a loss of intelligibility (Derwing & Munro, 1997; Munro & Derwing, 1995a; Rajadurai, 2007). Rather, the relationship between these two dimensions is more linked to the representation of the non-native segment in the phonological system of the listener. In terms of the perceptual magnet effect hypothesis (Kuhl, 1991), it can be said that [x] falls into the [h] category of English listeners, but it is a bad representative of this category.

The significant difference between English and Spanish listeners in the identification of plosives in isolated syllables can be explained by differences in the VOT of stops in these two phonological systems (see figure 5.5). Similarly, the absence of aspiration in the Czech phonological system (Doleželová, 2009) makes them unaware of the difference between both ends of the continuum, and no problems of identification were detected. Japanese results, on the other hand, presented a mixed picture: on the one hand, they had no problem identifying correctly every step of the [t]→[t^h] continuum (as we already mentioned for Spanish and Czech listeners), but found the first steps of the [k]→[kʰ] continuum hard to identify, as observed in English listeners. In Japanese, initial voiceless stops are slightly aspirated (Riney, Takagi, Ota, & Uchida, 2007), that is, the VOT is not as long as in native English realisations, but it is also not as short as among Spanish and Czech speakers. Specific values can be found in Kim (2011): VOT of /t/ has, on average, a duration of 107 ms among English speakers and 47 ms in Japanese exemplars; on the other hand, VOT of /k/ is 95 ms for English speakers and 69 ms for Japanese. We found a much bigger difference in the VOT of /t/ between cohorts than in /k/. This may be the reason why Japanese listeners perceive [t]→[tʰ] in the same way as Spanish and Czech listeners, while [k]→[kʰ] is perceived in a manner closer to English participants.

Japanese listeners also behaved like the English cohort for [r]→[ɹ] and [_x]→[_g]. Segments [x] and [r] are not part of the English nor the Japanese phonological systems, even though they may appear in regional variants (e.g. Scottish English, Watt, Llamas, & Johnson, 2014) or sociolects (e.g. the speech of the Yakuza, Streetharan, 2004). On the contrary, both segments can be found in the Spanish and Czech phonological systems, where [ɹ] has no phonemic value and, even though [g] is a possible segment both in Spanish and Czech, its appearance in word final position is not productive. This outcome seems to imply that non-native listeners with a phonological category associated with the foreign accented segment but not to the native category (Czech and Spanish listeners, in this case) are less sensitive to the contrast than native listeners or non-native listeners without a category for the accented segment (in our experiment, Japanese listeners). In other words, in a [non-native]/[native] contrast such that the [non-native] segment is not part of the target language phonological system, listeners whose L1 includes the [non-native] segment as a phonological category but not the [native] segment (Spanish and Czech) will identify the [non-native] segment as equivalent to the [native] one more easily than non-native listeners without any of the segments in their phonological system (Japanese) or native listeners (English). This finding is related to the interlanguage speech intelligibility benefit (Bent & Bradlow, 2003) in the sense that non-native listeners whose phonological systems exhibit a similar relationship between the accented and native segments are less affected by foreign accent than native listeners or other non-native listeners.

Japanese listeners were the only ones who perceived a total lack of nativeness in the non-native end of the [s]→[z] continuum. This pair is not contrastive in Spanish (Whitley, 2002), and there is a strong voice assimilation process in Czech (Skarnitzl & Šturm, 2014) which may interfere in the perception of this contrast. On the other hand, in Japanese and English, /s/ and /z/ are phonemic categories, but there are acoustic differences in the realisation of [s] by the two cohorts (Li, Edwards, & Beckman, 2009). One explanation for this effect could be that the Spanish realisation of [s] was closer to that of the Japanese listeners, resulting in a clearer distinction between both segments and, subsequently, a detriment in correct segment identification in the non-native end of the continuum, but further research into this topic is necessary to confirm this hypothesis.

Despite these differences across cohorts in segment identification, it is worth noting another finding already mentioned in this research: even though differences were found among cohorts in their segment identification skills and in the nativeness categorisation task, no significant differences were found in the results of the AX discrimination task. As stated in chapter 4, discrimination tasks trigger in listeners what has been called acoustic or auditory stage of perception, as opposed to the phonetic stage of perception. The difference between these two stages is defined in Pisoni (1973): "the auditory stage refers to the analysis of the acoustic wave form into a set of time-varying psychological dimensions [...] whereas the phonetic stage refers to the transformation of auditory dimensions into abstract phonetic features". Pisoni refers here to what is known as bottom-up and top-down processing (auditory and phonetic modes respectively, Field, 2004). While in phonetic mode, listeners retrieve information from large entities such as the lexicon (e.g. "*this word exists, this word does not*") or their phonological system (e.g. in Spanish, the /j/ phoneme is more often realised as [ʤ] than as [j] in word initial position). This knowledge varies greatly from one group of listeners to another, even among same-language speakers, and so it has great influence in how listeners perceive a certain utterance. On the other hand, bottom-up processing implies the use of smaller units (such as individual segments, in our case) or sensory information in order to reshape bigger structures. Our results show how bottom-up processing of the [non-native segment]/[native segment] contrasts are similar across cohorts, no matter their L1. In other words, listeners are equally capable of discriminating pairs of segments regardless of the representation of those segments in the phonological system of their L1.

RQ2: Do lexical factors affect the perception of foreign accented segments?

We wanted to know if and how higher order factors interfere in the perception of foreign accented segments. To check the effect of such factors, we compared the identification of segments in isolated syllables with the intelligibility of words in which a single segment was accented. It is important to remember that, in order for the comparison to be as accurate as possible, in the lexical level only the target segment was accented, and the rest of the word and other prosodic features remained unaltered.

One of the main findings in this regard is the fact that overall the lexical level does affect the perception of foreign accented segments. By comparing the results of the syllable-isolated segment identification task and the intelligibility of words in which only one segment was accented, we saw that the appearance of a segment in a word as opposed to in a non-lexical context affected the correct identification of the target accented segment, i.e., the answer to RQ2 is that indeed there is a lexical effect in the identification of foreign accented segments.

Listeners can disambiguate obscure target segments by their appearance in a word, especially if that word does not generate a minimal pair with the foreign accented segment (top-down processing). It is important to remember here that listeners of all cohorts reported a high level of familiarity with the experimental words, which means that this positive effect should apply to all of them equally. Nevertheless, we found that this effect could go in two directions: either positive, meaning that the appearance of an accented segment in a word had a beneficial influence on its correct identification, or negative, when the lexical context was detrimental for its correct identification.

Generally speaking, English listeners experienced a positive lexical effect. This could be because, even though all cohorts reported a high degree of familiarity with the words, native listeners have more access to the complementary information provided by the lexical level. Such an effect was not found among non-native listeners, who even experienced a negative lexical effect in specific cases. This behaviour seems to imply that rapid availability of lexical information (and, presumably, other higher order effects) has an impact on the perception of foreign accent, and that intelligibility of non-native words by native listeners is facilitated by this additional higher order context (Engen & Peelle, 2014), despite similar word familiarity.

The importance of the richness of lexical information for segment identification is further evinced by the fact that some non-phonemic contrasts suffered a negative lexical effect, like [ʤ]/[j] among Spanish listeners. Additionally, the discrimination capabilities of Spanish listeners were similar to those of native listeners. This means that native-like bottom-up processing capabilities are blocked when the non-native listener needs to access lexical information. Native listeners are generally benefited from this lexical context. On the other hand, non-native listeners do not have this lexical information as readily available as natives, so identification of a foreign accented segment in a word becomes harder than in a non-lexical context.

RQ3: Do the production of foreign accented vowels and consonants affect listener's perception in a similar manner?

We inquired into any contrasts between accented consonants and vowels in terms of their effects on listeners, and found that the two type of segments elicit different perceptual behaviour in listeners. Overall results for consonant continua (figure 7.1) and vowel continua (figure 7.5) revealed a subtle but critical difference: in consonant continua, non-native listeners suffered a negative lexical effect, while, in vowel continua, native listeners experienced a positive lexical effect. This means that non-native listeners were generally better at identifying accented consonants in non-lexical contexts than in word contexts, that is, [non-native]/[native] consonant contrasts were better identified at the acoustic level by non-native listeners. On the other hand, native listeners benefitted more from lexical information, especially in vowel continua. As we have already explained, this kind of behaviour is most likely due to the richness of the lexical information and the cognitive load associated with deciphering foreign accented messages, but the fact that it is more prominent in vowels than in consonants for native listeners implies that native listeners use vowels as a more salient cue for word identification than consonants. This effect has also been observed in higher order features. For instance, (Kewley-Port, Burkle, & Lee, 2007) demonstrated that vowel-only sentences are significantly more intelligible than consonant-only sentences for both young normal-hearing and elderly hearing-impaired listeners. Nevertheless, it is worth noting that this effect may not be universal. Poltrock and Nazzi (2015) claimed that 11-month-old French-learning infants tolerate misproductions in vowels better than in consonants, i.e., they rely more on the pronunciation of consonants in order to decode the message. It remains unclear whether French listeners would behave similarly when confronted with foreign accented speech instead of same-language realisations.

The saliency of vowels in the English phonological system is also apparent in the differences between cohorts detected in the nativeness categorisation task. Two groups were sharply distinguished: on the one hand, English and Czech listeners, for whom native accent was lower in the non-native end than in the native end of the continua; on the other, Spanish and Japanese listeners, who perceived almost no differences from one end to the other. Volín and Skarnitzl (2010) claimed that Czech listeners are able to use vowel duration information to perceive nativeness in English. This means that both English and Czech listeners have two cues at their disposal to identify vowels: vowel quality and vowel quantity. On the contrary, Spanish and Japanese share a similar vowel system in terms of quality, i.e. 5-vowel system, with the main difference that Japanese, but not Spanish, does have vowel length contrasts (Morrison, 2002). Nevertheless, it is known that Japanese listeners disambiguate [short]/[long] English pairs of vowels using primarily spectral information (Nishi & Kewley-Port, 2007), displaying a behaviour closer to Spanish listeners than to Czech listeners. This kind of two-way disambiguation of cohorts was not observed in consonant continua (see figure 7.1).

Further evidence that vocalic differences are more salient in English than consonant differences can be found in the results of the discrimination task for vowel continua (figure 7.10). As opposed to the results of the same task for consonant continua (figure 7.9), in some vowel continua an increment in the discrimination rate was found when the distance between steps was increased. This finding implies that the bigger the deviation in the pronunciation of vowels, but not consonants, the greater the perceived difference between them.

Conclusion

Three main findings can be extracted from the study reported in this chapter. First, the perception of foreign accented segments is directly linked to the phonological system of the listener and the representation of such segments in it. Second, native listeners benefit more from a lexical context; as the frame of analysis became smaller, the capabilities of the four analysed cohorts became more similar, to the point that no differences between them were appreciated in bottom-up processing. And third, foreign-accented vowels are more likely to trigger intelligibility impairments than foreign-accented consonants.

In the final chapter, these findings are summarised and contextualised within the framework of this thesis in order to provide insights about possible future lines of work and practical applications.

Chapter 8

General conclusions

This closing chapter highlights the major findings and contributions of the thesis. Some limitations of the work are also discussed, and applications and impact of the findings are considered. Finally, potential future lines of research arising from the investigation are described.

Novel contributions

This thesis set out to present an understanding of the segmental basis of foreign accent. During the course of the work, in addition to experimental findings, new speech materials, methods and algorithms have been developed. The principal findings as well as theoretical and methodological contributions are summarised below.

Findings

- Perception of segmental foreign accent is highly dependent on the phonological system of the listener [Chapter 7]
- There is great cross-segmental variability regarding the amount of acoustic deviation from the native prototype required to convey the same amount of foreign accent [Chapter 5]
- Intelligibility loss due to non-native speech is less when the listener shares the same L1 as the speaker and greater when the listener is non-native with a different L1 than the speaker [Chapter 3]

- Native listeners benefit more from the lexical context than non-native listeners, that is, they can use lexical information more effectively in situations where segmental foreign accent appears [Chapter 7]
- Perceived nativeness in continua spanning from a native segment and its foreign accented counterpart differs across native and non-native listeners [Chapter 5]
- Listeners with high proficiency in a second language exhibit more native-like capabilities regarding perception of segmental foreign accent than listeners with low proficiency [Chapter 4]
- Some segments are perceived monotonically in a [non-native]→[native] continuum (i.e. non-categorical perception), while others present an abrupt change in perceived category at some point along the [non-native]→[native] continuum (categorical perception) [Chapter 5]. In particular, categorical perception is generally seen more for consonants than for vowels [Chapter 7]
- If the accented counterpart of a native segment (but not the native realisation) is part of the phonological system of the non-native listener, the non-native realisation is more readily assigned to the category of the native segment than otherwise [Chapter 7]
- The most possible accented realisation of a native segment does not always convey a complete lack of nativeness: the more salient the foreign accented segment is, the less perceived nativeness [Chapter 5]

Theoretical contributions

- A strong FA does not always imply loss in intelligibility at the segmental level
- Results of segmental analyses are not entirely correlated with the results of holistic analysis in the field of FA
- The role and background of the listener is crucial in segmental foreign accent perception
- New phonological boundaries can arise in the L2 system of non-native speakers through training in the target language
- Listeners with different linguistic background exhibit similar bottom-up capabilities

Methodological contributions

- Generation of segmental foreign accent: evaluation of three techniques, including a novel bilingual synthesis approach
- Gradation of segmental foreign accent: development of semi-automated algorithmic procedures to produce segments with controlled degrees of foreign accent
- Iso-perceptual continua of segmental foreign accent: a new approach to normalise differences across segment continua to ensure equal perceptual rather than acoustic steps
- Schematic patterns of accent judgements: a set of schematic templates for patterns of accent judgements, alongside a flexible parametric fitting procedure and decision criterion

Limitations

In this thesis, the focus has been placed on English words in which one segment was pronounced with a Spanish foreign accent. Even though some of the results may be generalisable to other foreign accent traits, our findings are population-oriented, i.e. they can only be applied to the languages within the scope of the experimental languages (English and Spanish). Although a large number of extensive perceptual experiments have been carried out in 4 countries in the course of investigating the topic of the thesis, it is always possible to do more, and it is recognised that time constraints (particularly in terms of experimental participant time and the scope of visits to laboratories) have limited the work to certain segments, positions and target languages, as well as the number of lexical and non-lexical tokens employed.

Applications

Pedagogical applications

The most direct application of the knowledge acquired in this thesis is in the pedagogical domain. Depending on the motivations of the student, English as foreign language (EFL) teachers will be able to adapt their pedagogical approach based on outcomes for the segment inventory explored here. For example, a Spanish student whose aim is to be understood in English ought to focus training episodes on those segments that are responsible for the greatest loss in intelligibility among native listeners (e.g. [ʤ], [ɜː]). On the contrary, students whose goal is to be perceived as native, will prefer to put additional emphasis in segments that are perceived as strongly accented by native listeners (e.g. [h], [uː]), even though they may not cause a reduction in intelligibility.

Automatic speech recognition

Speech recognition systems can be a source of frustration among speakers (Assefi, Liu, Wittie, & Izurieta, 2015; Taylor, Dey, Siewiorek, & Smailagic, 2015), and that frustration can increase if the accent of the speaker does not match the expectations of the machine. Several studies have focused on automatic recognition of foreign accent (Behravan, Hautamaki, Siniscalchi, Kinnunen, & Lee, 2016; Hansen & Arslan, 1995; Kat & Fung, 1999) in order to improve user experience when dealing with this kind of systems. It can be expected that, by incorporating segmental information specific to a given target accent and predicted possible input accents, speech recognition systems can improve their performance.

Computer-based language learning

Speech recognition is also widely employed in mobile applications and computer-based platforms for English learning (*Duolingo*, 2015; *Elsa Speak*, 2018; Martínez-Paricio, Koreman, Husby, Eggesbø Abrahamsen, & Bech, 2014, among others). Generally speaking, these tools ask the user to record a set of words or sentences in order to provide feedback, advice, or even generate entire lessons based on the most problematic pronunciation issues of the user. Some tools base their assessment on statistical models, that is, they compare the pronunciation of the user with a model generated from a large corpus (Ryu, Hong, Kim, & Chung, 2016). In such a directed context, specific segmental knowledge combined with statistical models could greatly help enhance computational assessment of foreign accent. Particularly, the tool could boost automatic processes such as detection of improvements on the pronunciation of the user or better detection of false positives or negatives, improving positive feedback and reducing frustration.

Further studies

Our focus has been on communication problems related to intelligibility and degree of foreign accent, but results regarding segmental foreign accent such as the ones presented on this thesis can have applications in other areas. Fields such as sociolinguistics and psychology could benefit from the techniques and methods used in this thesis in order to detect the roots of phenomena such as linguistic profiling (see chapter 2).

The set of segments evaluated in this thesis is not comprehensive. The decision was made to focus on consonants in initial position, vowels in medial position and a small set of highly-confusable consonants in final position. It remains unknown if and how the pronunciation of these segments in other word positions could affect the perception of foreign accent. This thesis opens the door to check the effect of accented clusters, diphthongs or other structures that would require segmental manipulation (e.g. one syllable completely accented in a multi-syllabic word or native sentences in which only one word is pronounced with a foreign accent).

In this thesis the pronunciation of English by Spanish speakers has been mapped, but the methodology employed to perform the assessment is not language-specific. The gradation technique, the iso-perceptual method and, in general, the methodology for evaluation of segmental foreign accent proposed here can be extended to other L1/L2 combinations.

It is expected that some of the proposed paths for further studies require modifications in the gradation technique, e.g. to incorporate the possibility of modifications in other word-positions or to simultaneously modify several segments or all the occurrences of a segment in a sentence. As the scientific knowledge advances, synthetic speech systems are expected to generate more human-like sounds, and the possibility of a fully computer-based segmental foreign accent generator should not be ruled out. Other than the possible pitfall of being unnatural sounding, computational systems are highly flexible in performing acoustic manipulations, and may enable the scientific community to reach new levels of experimentation and more advanced applications to further explore segmental foreign accent.

Conclusion

Segmental foreign accent has an impact on speech perception in ways that have not been revealed by holistic studies to date. The methodology proposed in this thesis opens the door to further research in a field that can be pivotal for people living among speakers with other accents.

Appendices

Appendix A

Linguistic questionnaire

QUESTIONNAIRE FOR LISTENERS IN PERCEPTION TASKS

All the information provided in the following questionnaire will be treated confidentially, anonymously and will be used for the sole purposes of this experiment

If there are any questions you are uncertain about, please ask the experimenter for clarification

BIOGRAPHICAL INFORMATION

1- Participant ID (from experimenter):

2- Age:

3- Place of birth:

4- City/town of current residence and since when:

5- Please state any locations where you have lived for at least one year, indicating how long and when:

-
-
-

6- Highest education level achieved or currently enrolled in. If university degree please specify which one:

7- Current occupation:

8- How would you (or others) describe your English accent?

Languages Information

1. Can you speak any other language apart from English?

2. If you answered yes to the previous question, please specify below for each language your level of competence in each of the skills mentioned (1 means beginner and 5 means native):

	Speaking					Listening					Reading					Writing				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

3. How often do you interact with **foreign** speakers of English **in English**? (circle one option)

1 never	2 hardly ever	3 from time to time	4 quite often	5 constantly

4. How often do you interact with **Spanish** speakers of English in **English**? (circle one option)

1 never	2 hardly ever	3 from time to time	4 quite often	5 constantly

5. Degree of **oral** contact with other **languages** (circle one option):

Spanish

1 never	2 hardly ever	3 from time to time	4 quite often	5 constantly

Other (specify)____________

1 never	2 hardly ever	3 from time to time	4 quite often	5 constantly

Other (specify)____________

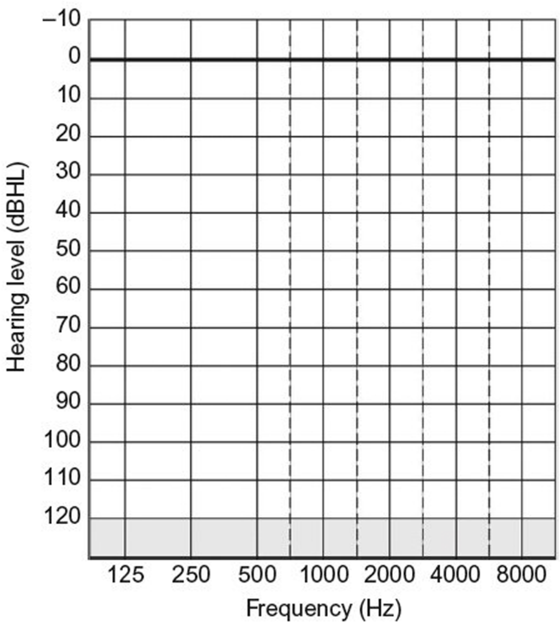
1 never	2 hardly ever	3 from time to time	4 quite often	5 constantly

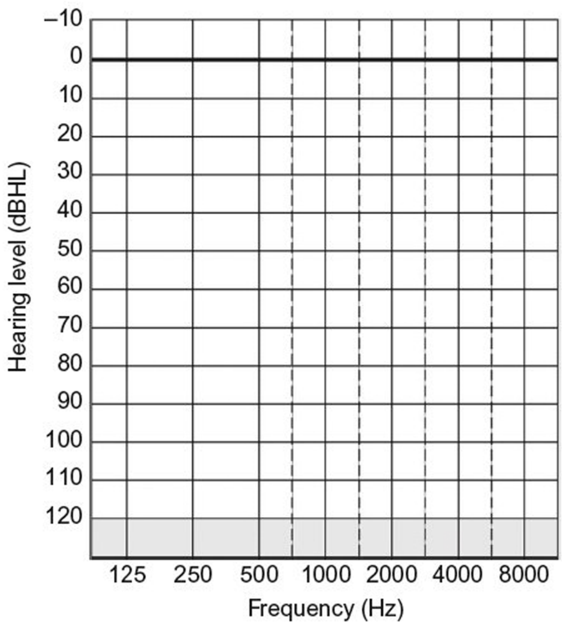
Other (specify)____________

1 never	2 hardly ever	3 from time to time	4 quite often	5 constantly

6. Have you ever lived in a foreign country for any length of time? If yes, state where, when and for how long.

7. Do you have any musical knowledge? If yes, please specify.





Appendix B

Consent form

CONSENT AND INFORMATION FORM

Research Project: Dissecting Foreign Accent.

This project involves the collaboration of CSTR at the University of Edinburgh, the University of the Basque Country and Charles University Prague. With the collaboration of Bettina Beinhoff, from Anglia Ruskin University.

Researchers: Rubén Pérez, M. Luisa García Lecumberri, Martin Cooke, Simon King.

Description: In this project we are trying to understand foreign accents in depth. For that we need estimations from native English speakers about different accents in English, both native and non- native. Some of the stimuli we present have been manipulated.

Procedure: There are four different tasks:

- In the first task we present one word at a time via headphones and your task is to type the word yo heard.

- In the second task we present two words at a time via headphones and your task is to say if the pronunciation of both words is the same or different

- In the third task we present CVC or VCV structures and you have to tell us which sound category it fits in.

- In the fourth task we present one word at a time via headphones and we also present it in spelling on the computer screen. Your task is to judge whether that word has been spoken with a native or a foreign accent.

At the beginning of each task, there will be a short practice session at the end of which you can call the researcher if you have any questions.

I , ..., declare that:

- I have read and understood the information sheet given to me by the researcher on (date) ___________.

- I had the chance to ask questions and get additional information.

- I understand that the aims of the project are research-based and ultimately the improvement of language teaching methods.

- I have been told how the confidentiality of the answers will be preserved.

- I have been assured that I am free to withdraw from the experiment at any point without having to explain why.

Date	Participant's signature

Researcher's declaration: I have explained the project aims and testing procedures in detail and have answered all questions conscientiously.

Date	Researcher's signature

Appendix C

Experimental interfaces

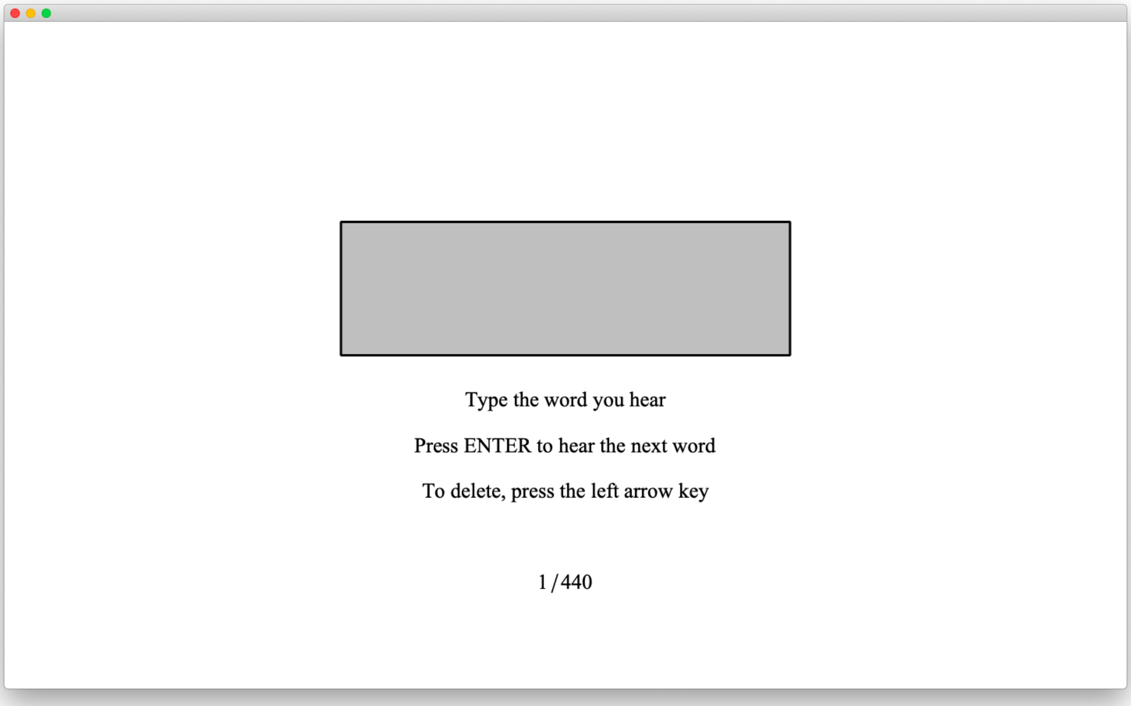


Figure C.1: Intelligibility task interface

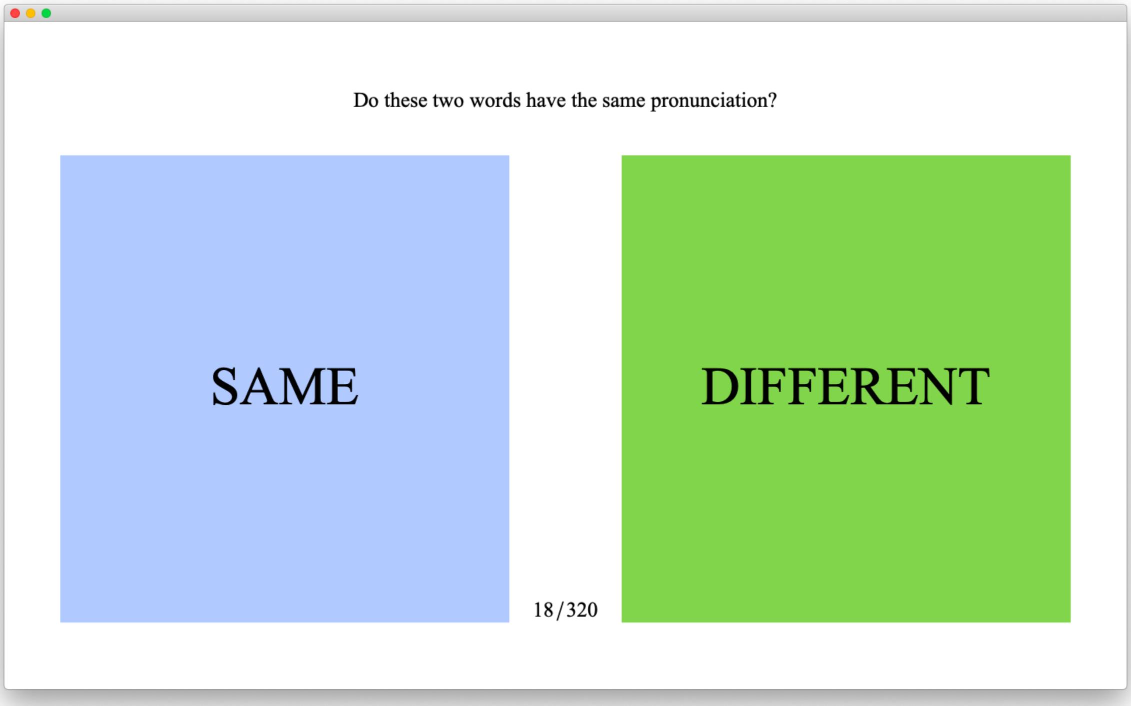


Figure C.2: Discrimination task interface

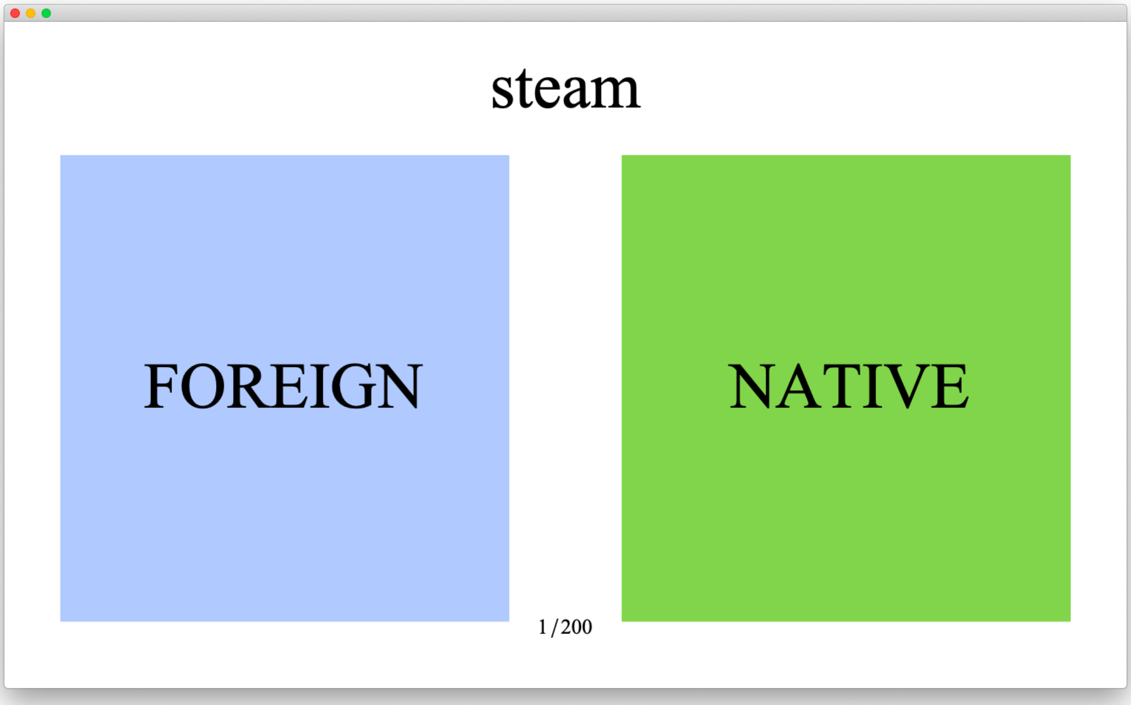


Figure C.3: Categorisation task interface

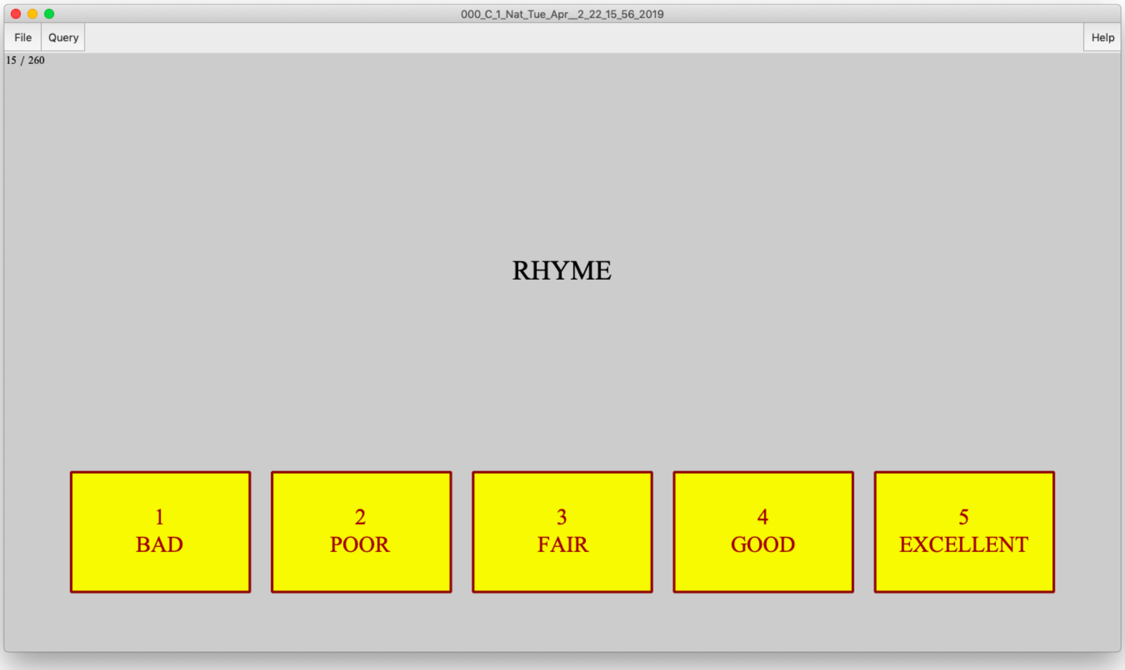


Figure C.4: MOS signal quality task interface

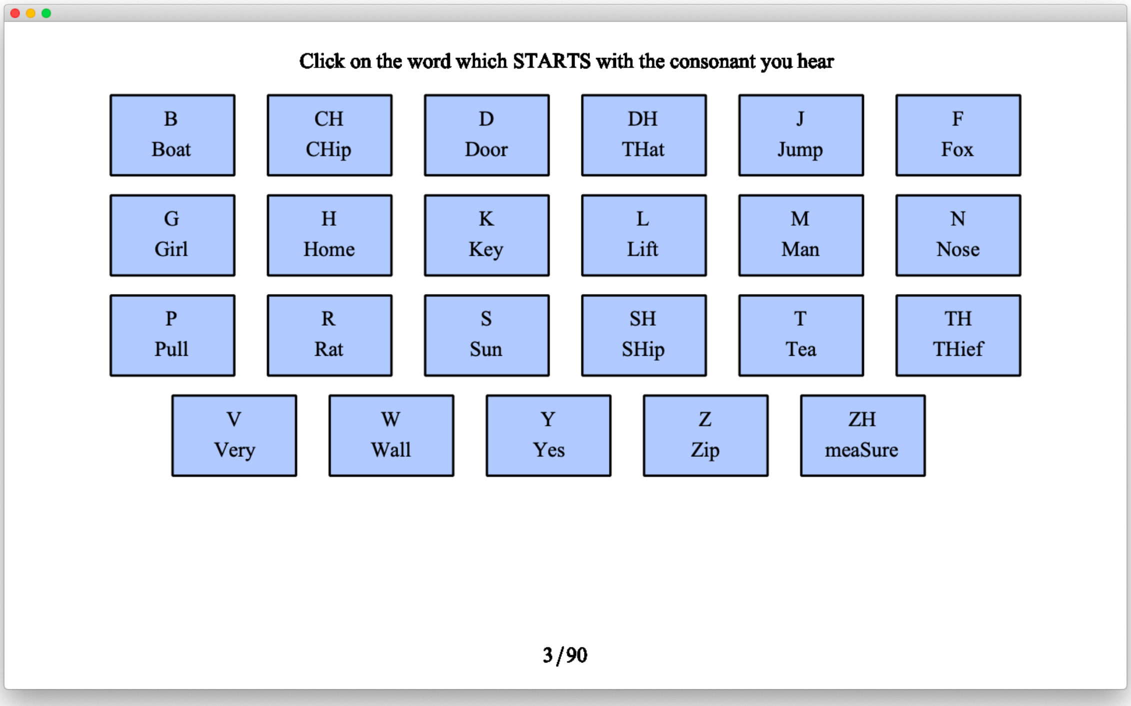


Figure C.5: Identification/Competence task interface

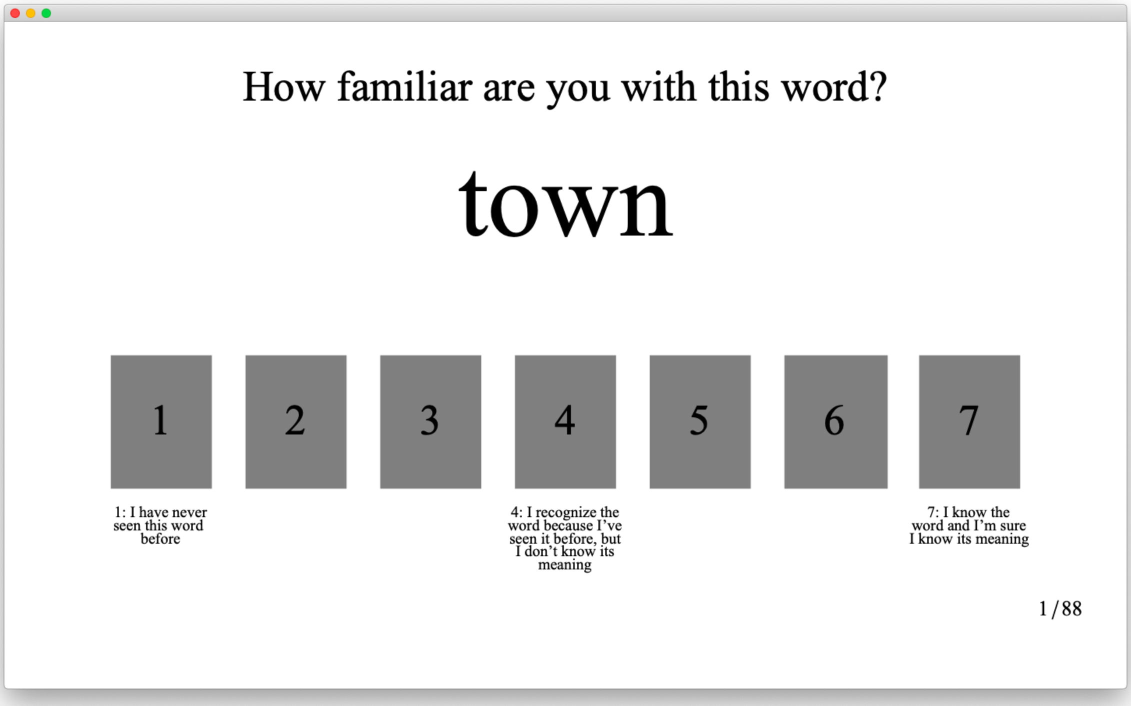


Figure C.6: Familiarity task interface

Appendix D

Hinton diagrams for word identification task (Chapter 3)

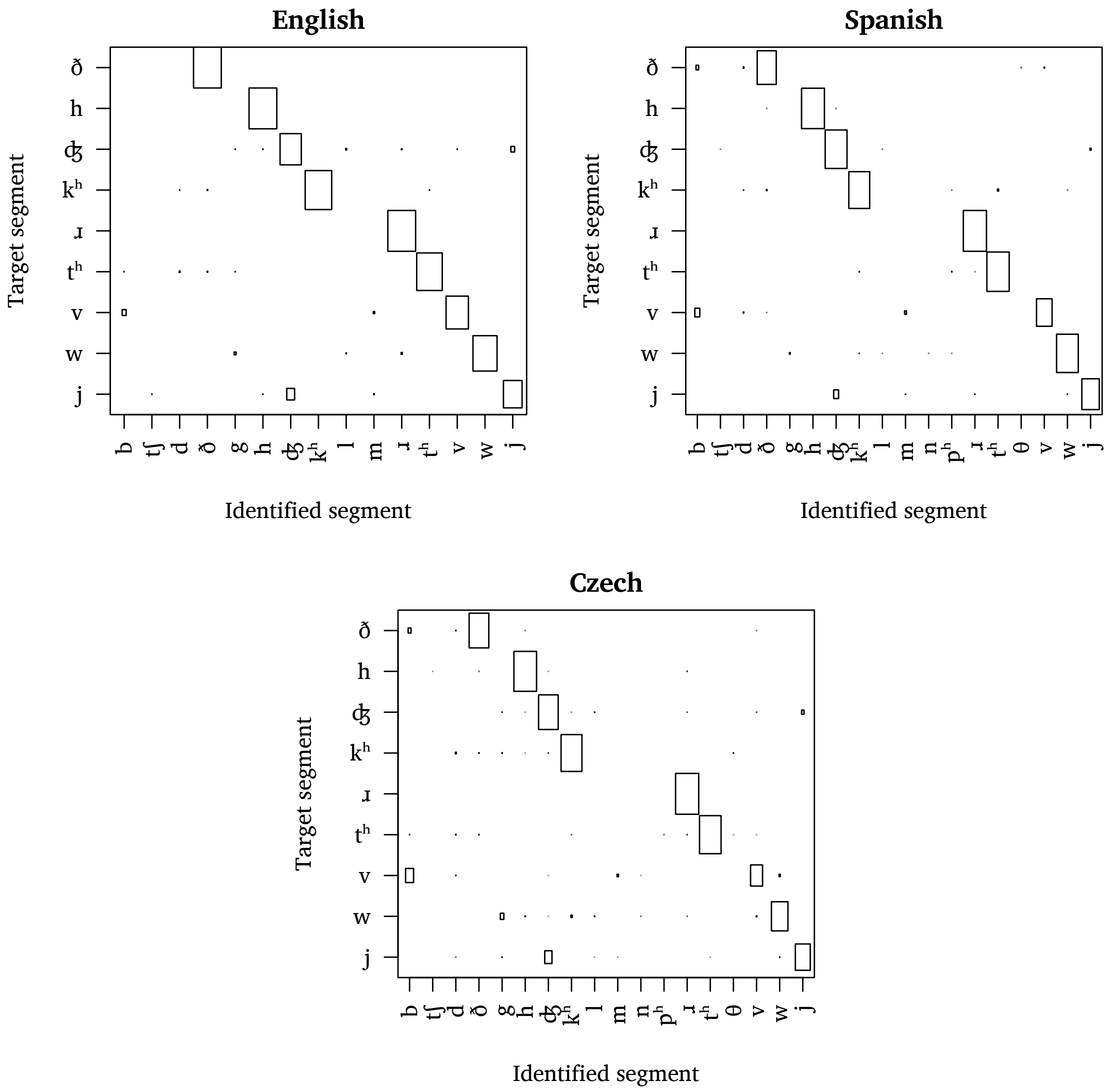


Figure D.1: Word intelligibility confusions of the English, Spanish and Czech cohorts for the intelligibility task of chapter 3. The area of the squares represents the proportion of responses for each category. The ⊘ category represents the elision of the target segment.

Appendix E

Hinton diagrams for segment identification task (Chapter 7)

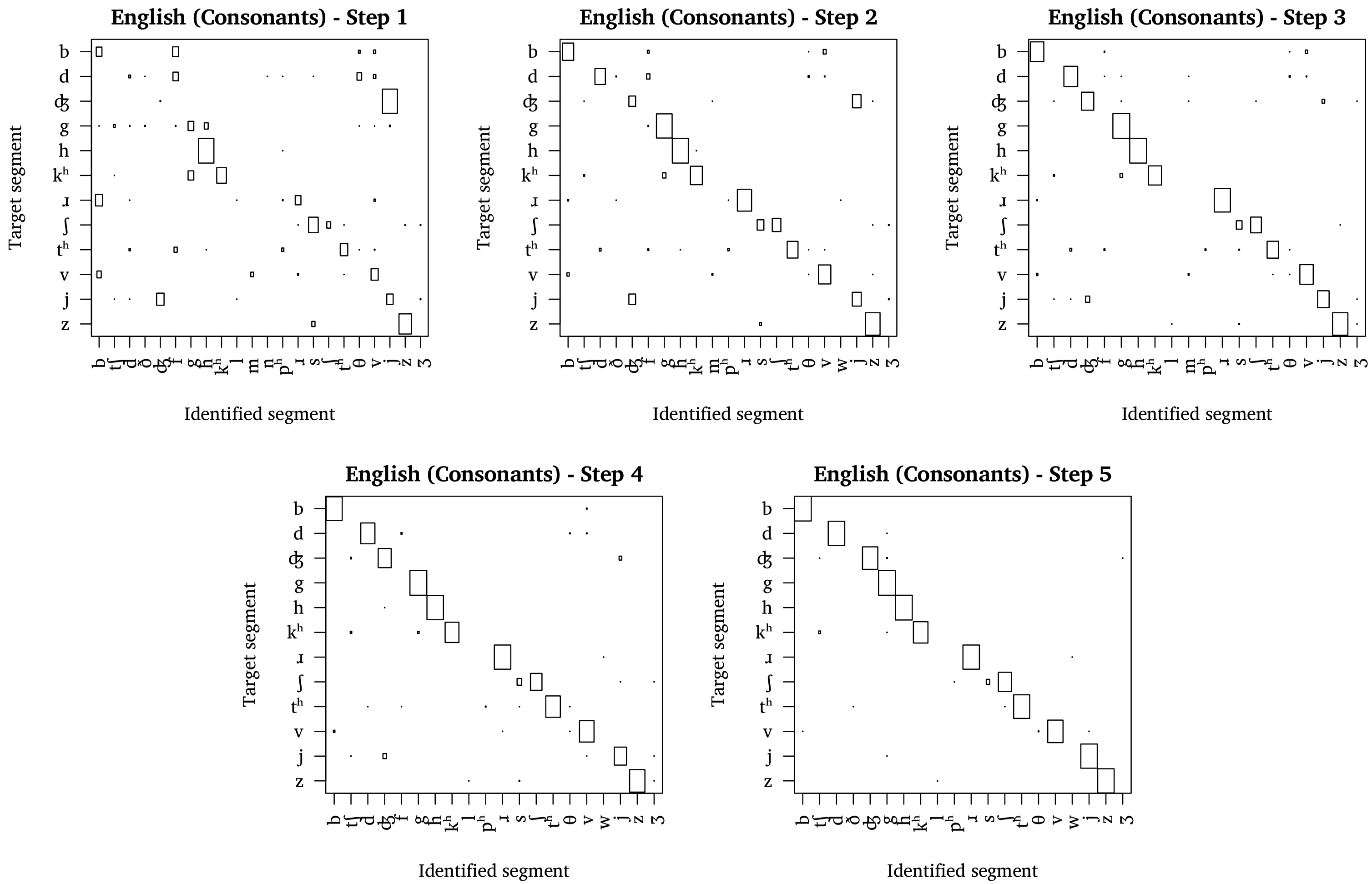


Figure E.1: Segment identification confusions of the English cohort for consonants for the segment identification task of chapter 7. The area of the squares represents the proportion of responses for each category. The ⊘ category represents the elision of the target segment.

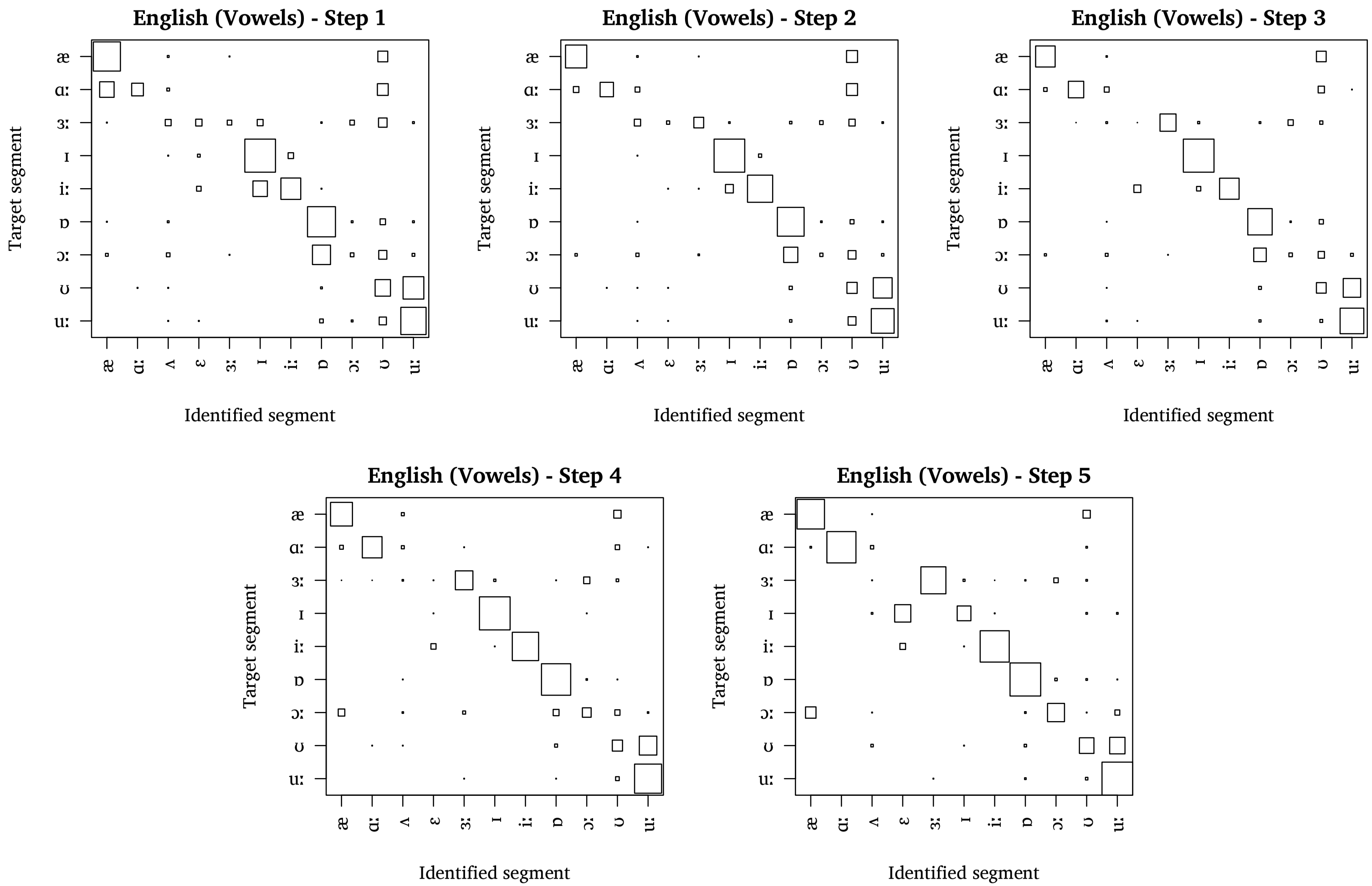


Figure E.2: Segment identification confusions of the English cohort for vowels for the segment identification task of chapter 7. The area of the squares represents the proportion of responses for each category.

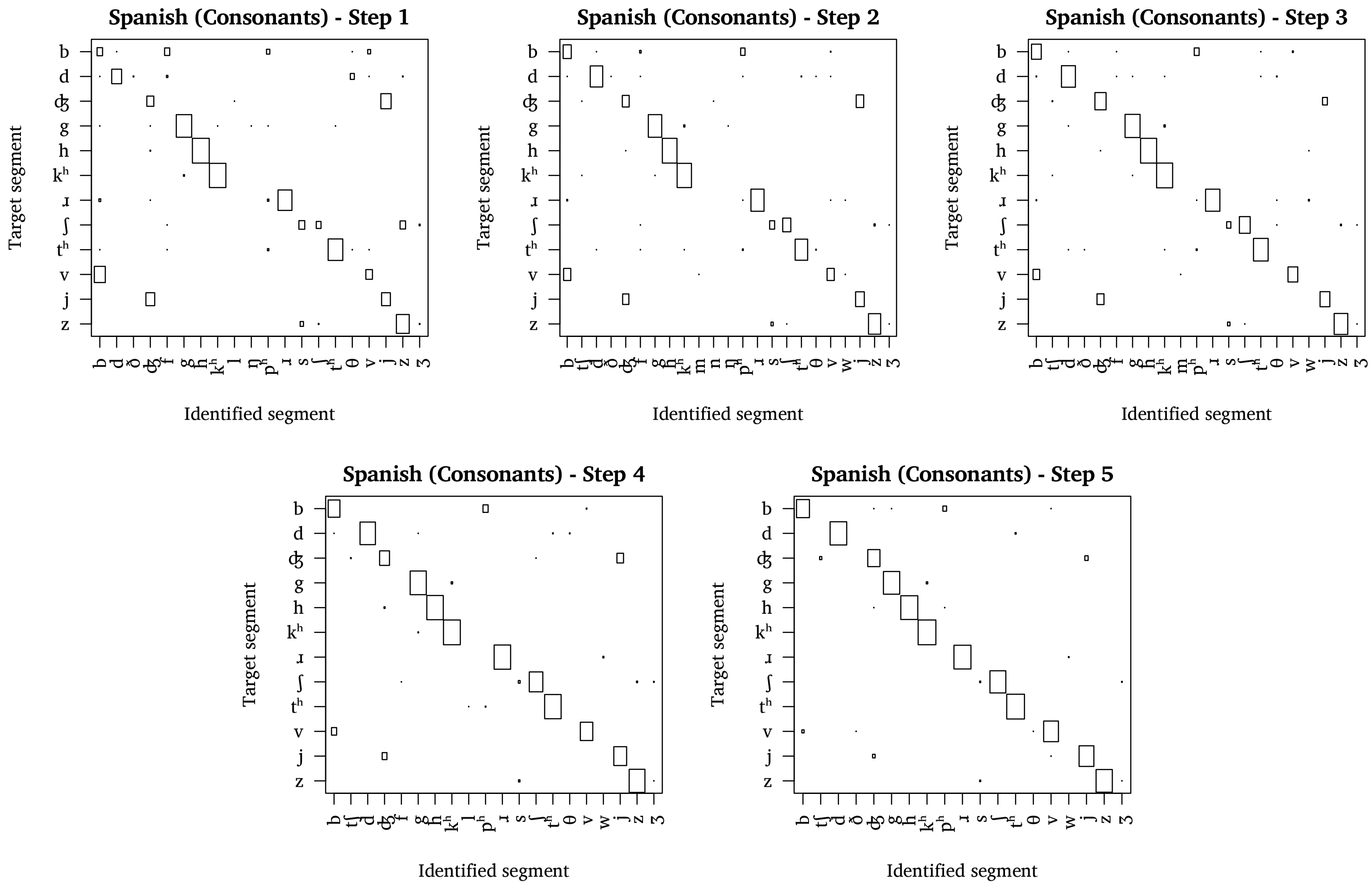


Figure E.3: Segment identification confusions of the Spanish cohort for consonants for the segment identification task of chapter 7. The area of the squares represents the proportion of responses for each category. The ⊘ category represents the elision of the target segment.

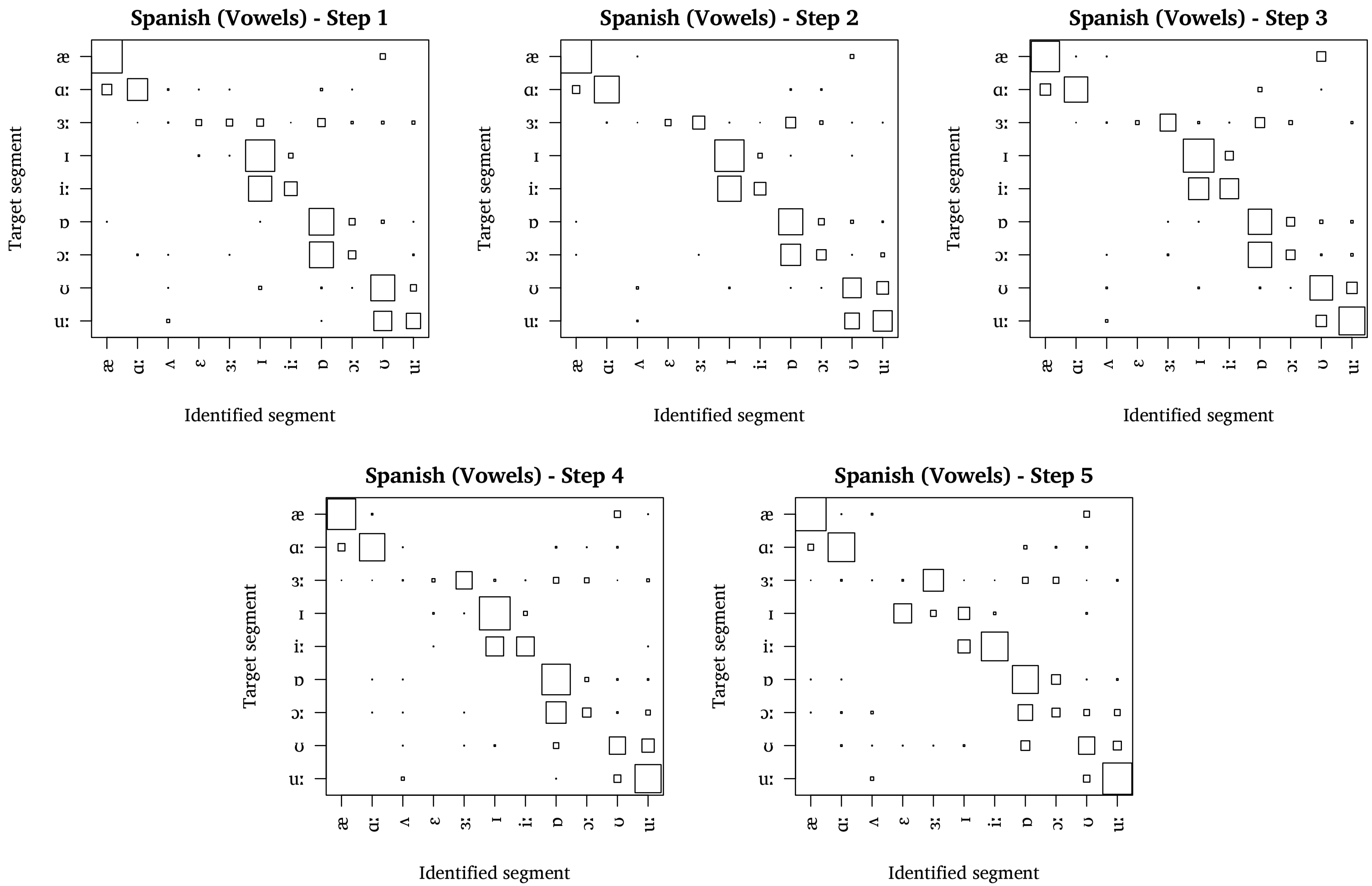


Figure E.4: Segment identification confusions of the Spanish cohort for vowels for the segment identification task of chapter 7. The area of the squares represents the proportion of responses for each category.

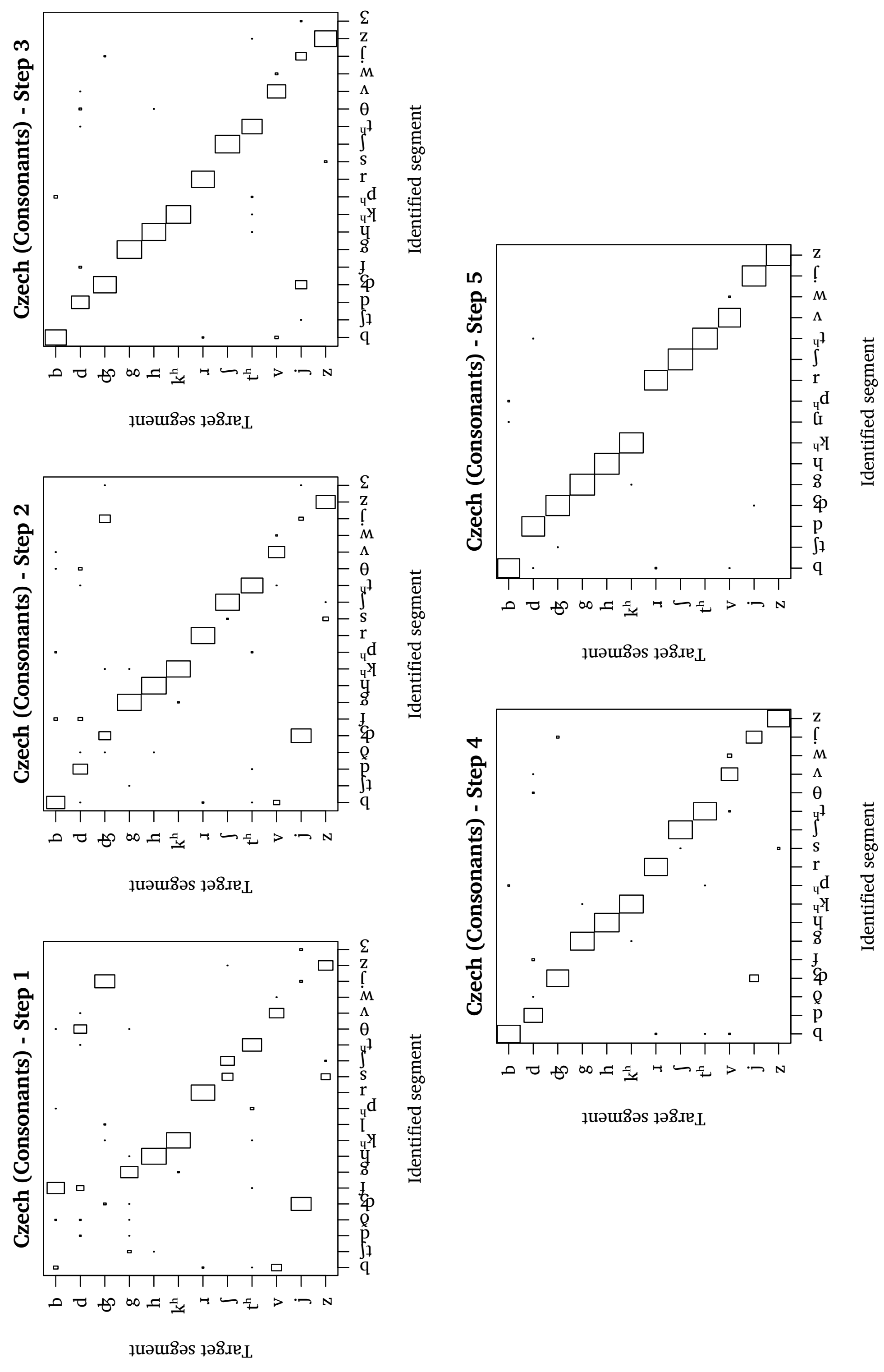


Figure E.5: Segment identification confusions of the Czech cohort for consonants for the segment identification task of chapter 7. The area of the squares represents the proportion of responses for each category. The ⊘ category represents the elision of the target segment.

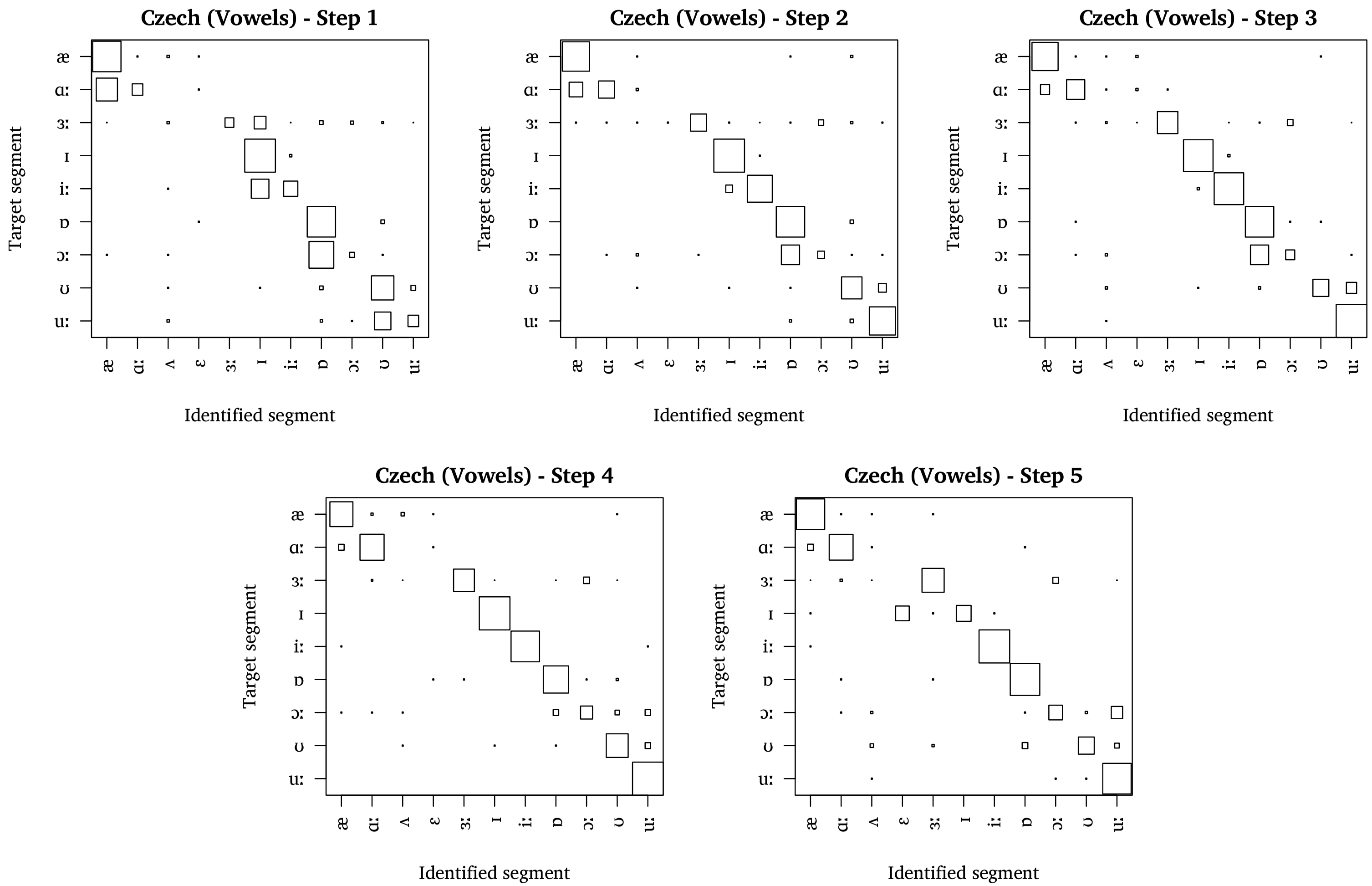


Figure E.6: Segment identification confusions of the Czech cohort for vowels for the segment identification task of chapter 7. The area of the squares represents the proportion of responses for each category.

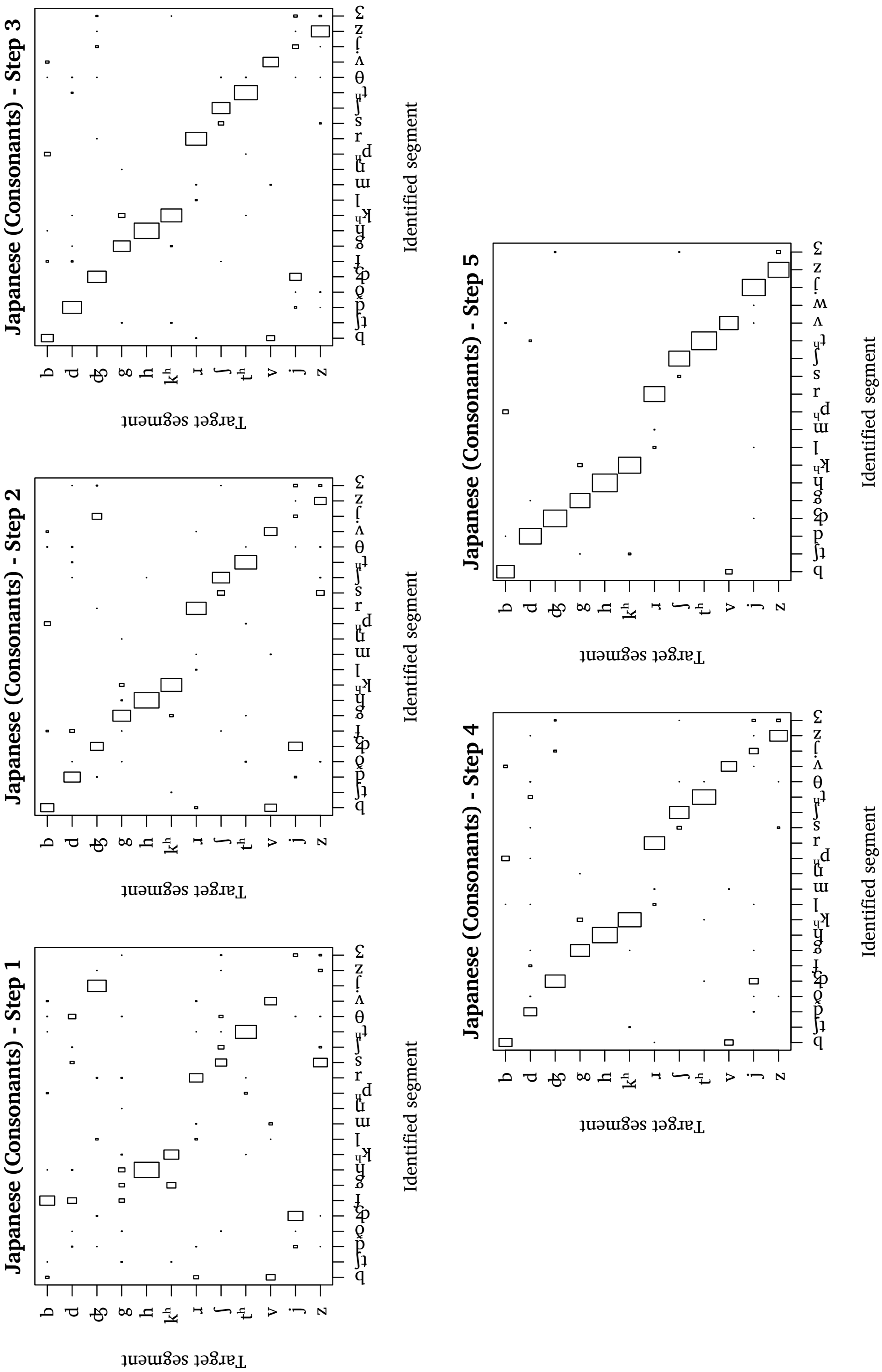


Figure E.7: Segment identification confusions of the Japanese cohort for consonants for the segment identification task of chapter 7. The area of the squares represents the proportion of responses for each category. The ⊘ category represents the elision of the target segment.

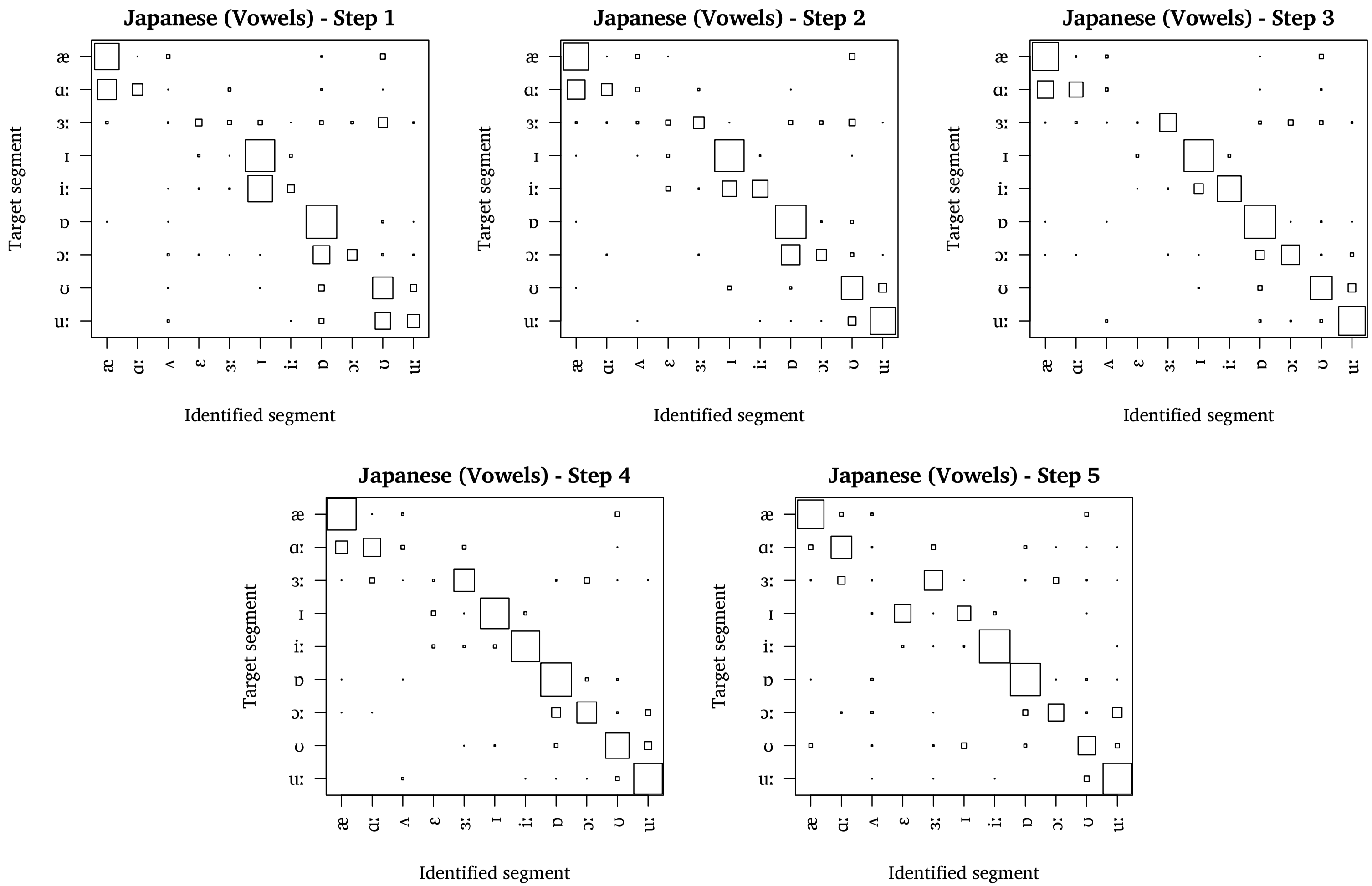


Figure E.8: Segment identification confusions of the Japanese cohort for vowels for the segment identification task of chapter 7. The area of the squares represents the proportion of responses for each category.

Appendix F

Hinton diagrams for word intelligibility task (Chapter 7)

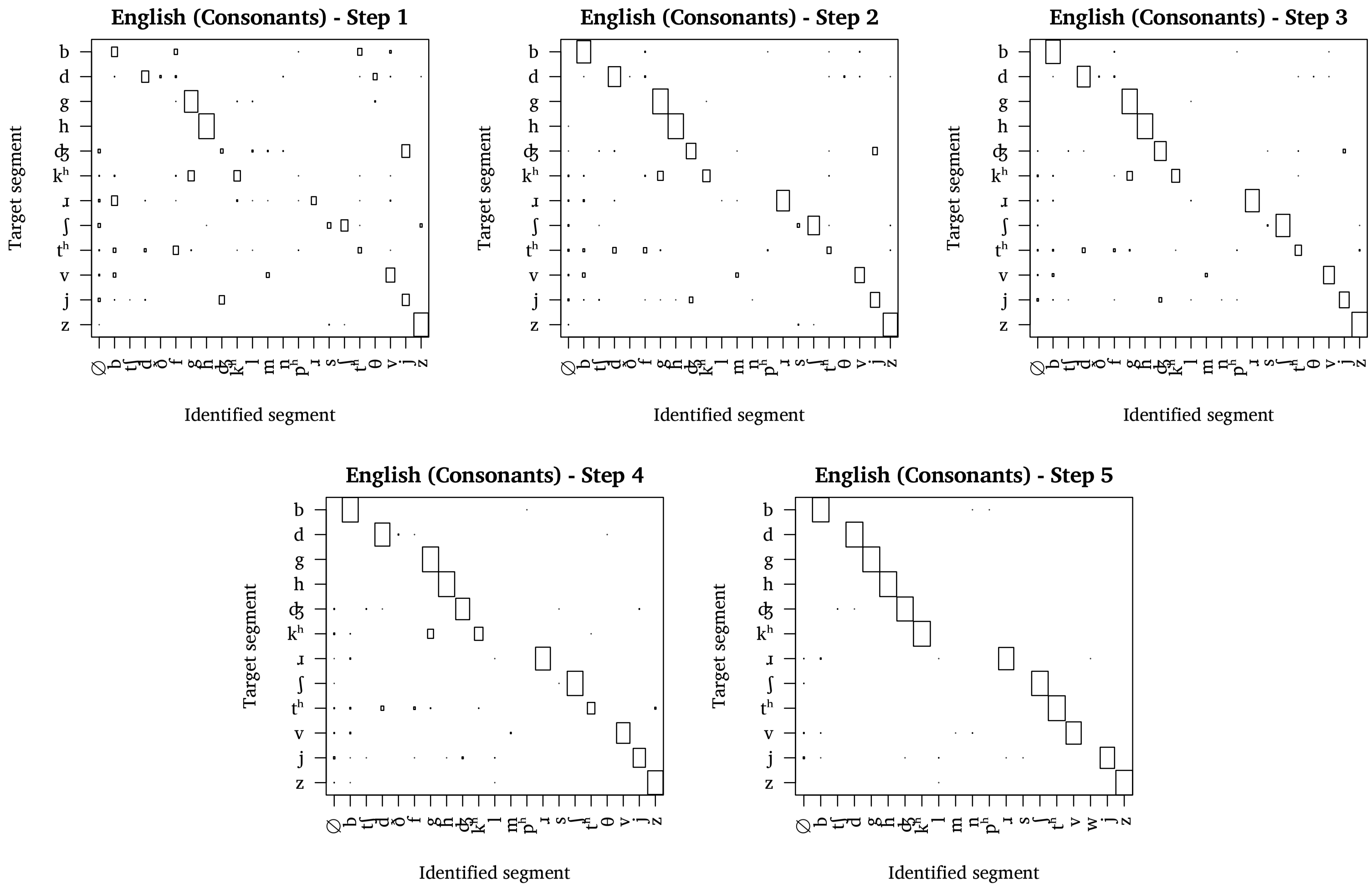


Figure F.1: Word intelligibility confusions of the English cohort for consonants for the word intelligibility task of chapter 7. The area of the squares represents the proportion of responses for each category. The ⊘ category represents the elision of the target segment.

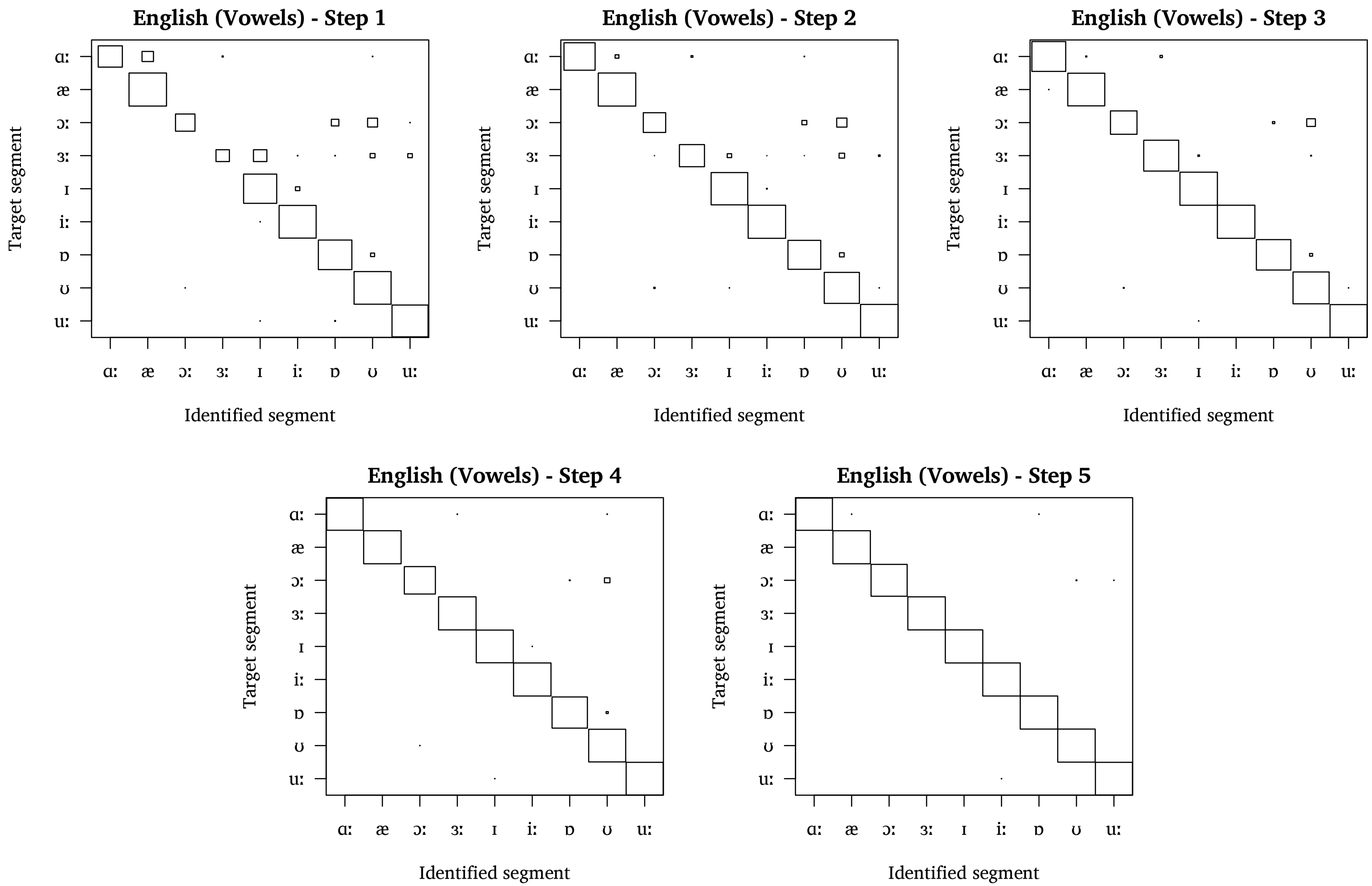


Figure F.2: Word intelligibility confusions of the English cohort for vowels for the word intelligibility task of chapter 7. The area of the squares represents the proportion of responses for each category.

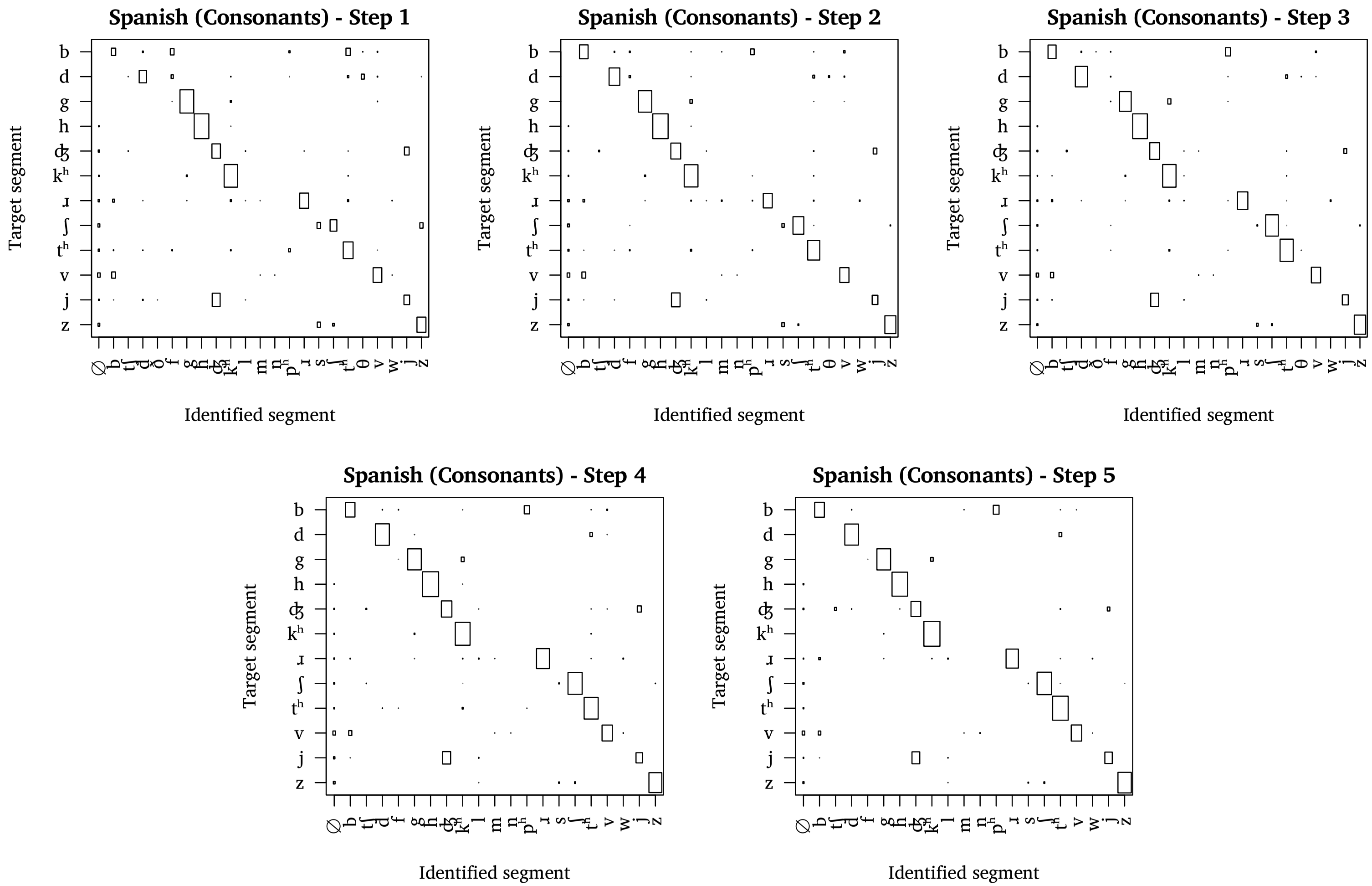


Figure F.3: Word intelligibility confusions of the Spanish cohort for consonants for the word intelligibility task of chapter 7. The area of the squares represents the proportion of responses for each category. The ⊘ category represents the elision of the target segment.

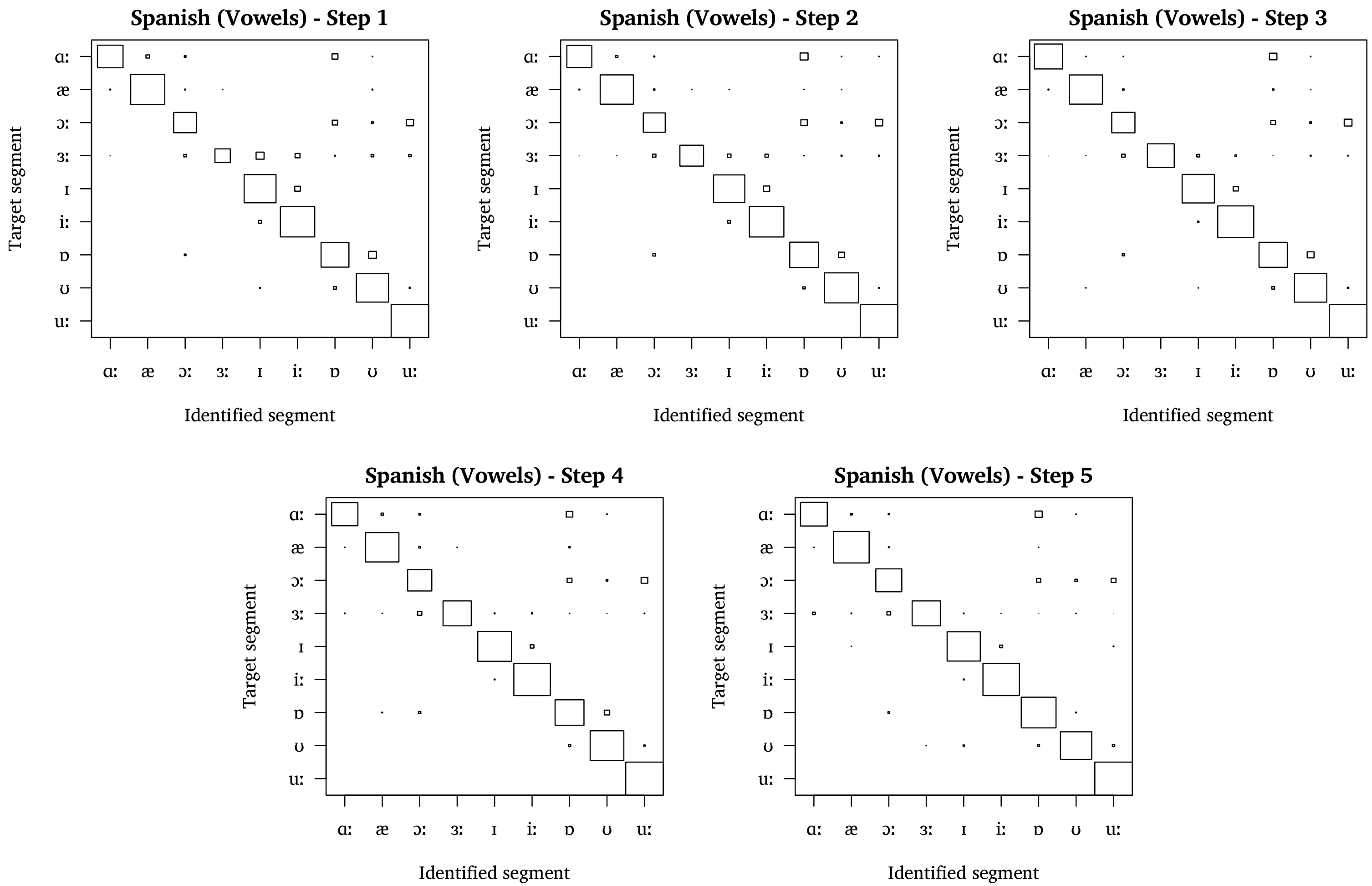


Figure F.4: Word intelligibility confusions of the Spanish cohort for vowels for the word intelligibility task of chapter 7. The area of the squares represents the proportion of responses for each category.

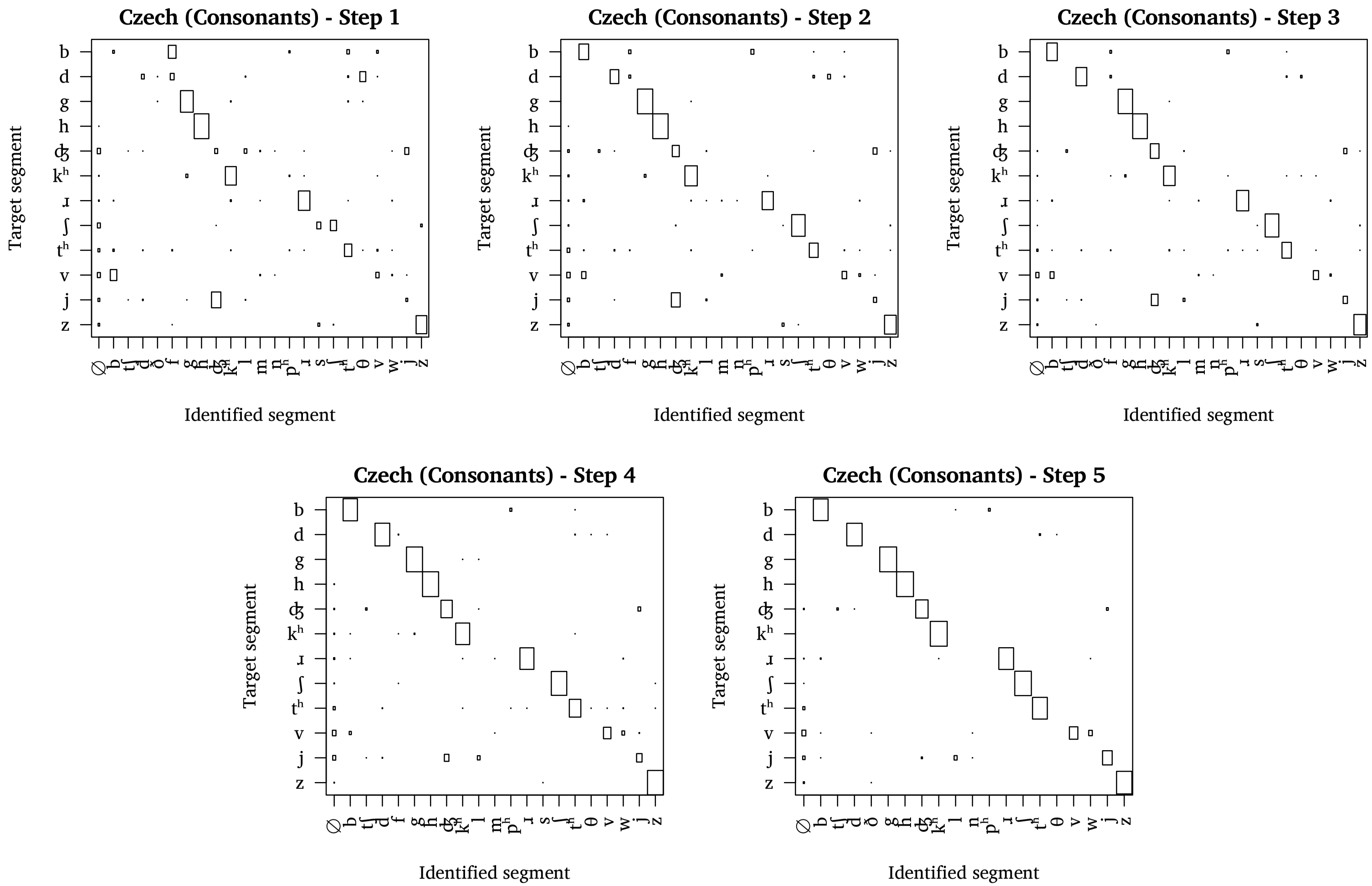


Figure F.5: Word intelligibility confusions of the Czech cohort for consonants for the word intelligibility task of chapter 7. The area of the squares represents the proportion of responses for each category. The ⊘ category represents the elision of the target segment.

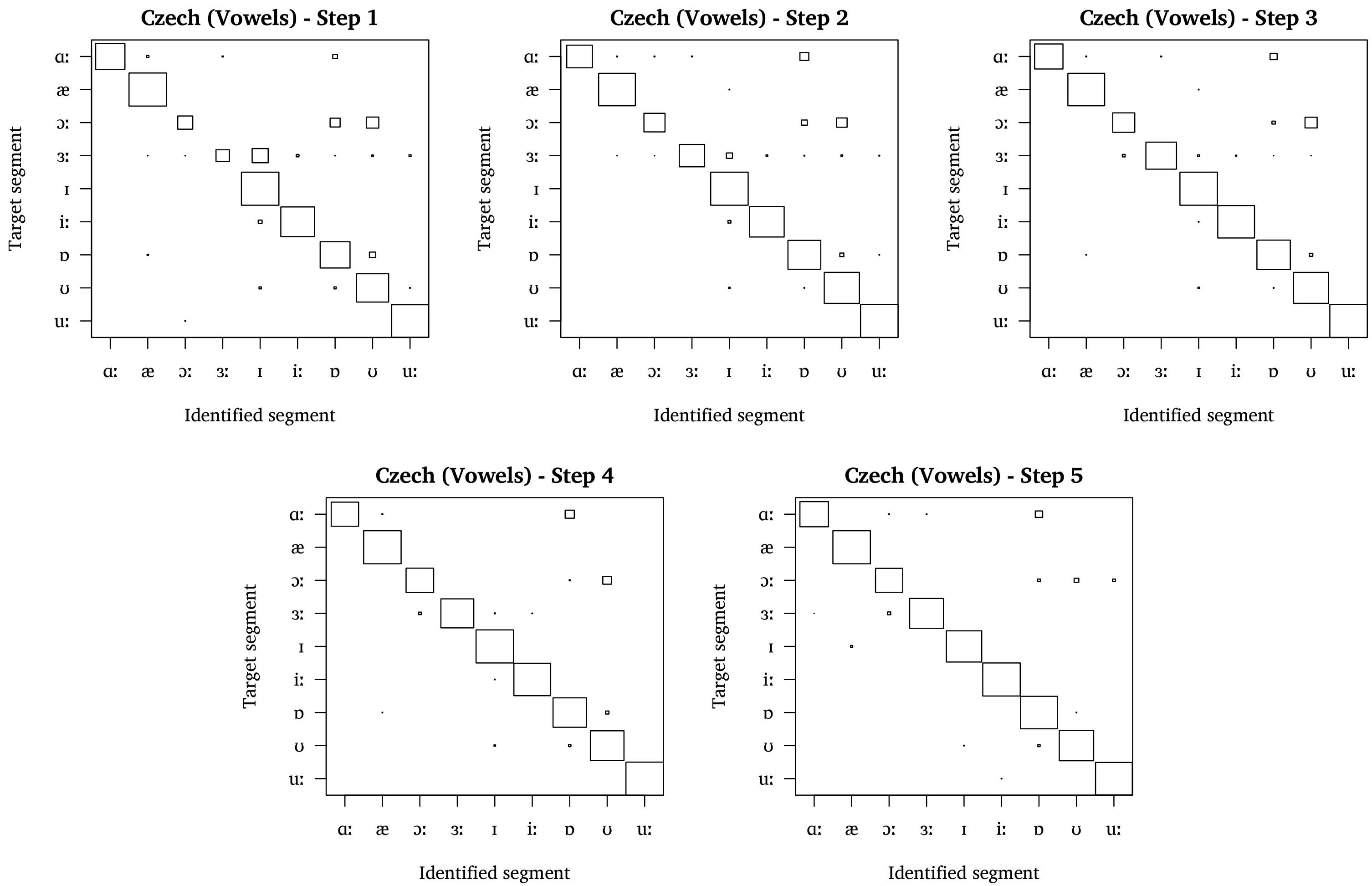


Figure F.6: Word intelligibility confusions of the Czech cohort for vowels for the word intelligibility task of chapter 7. The area of the squares represents the proportion of responses for each category.

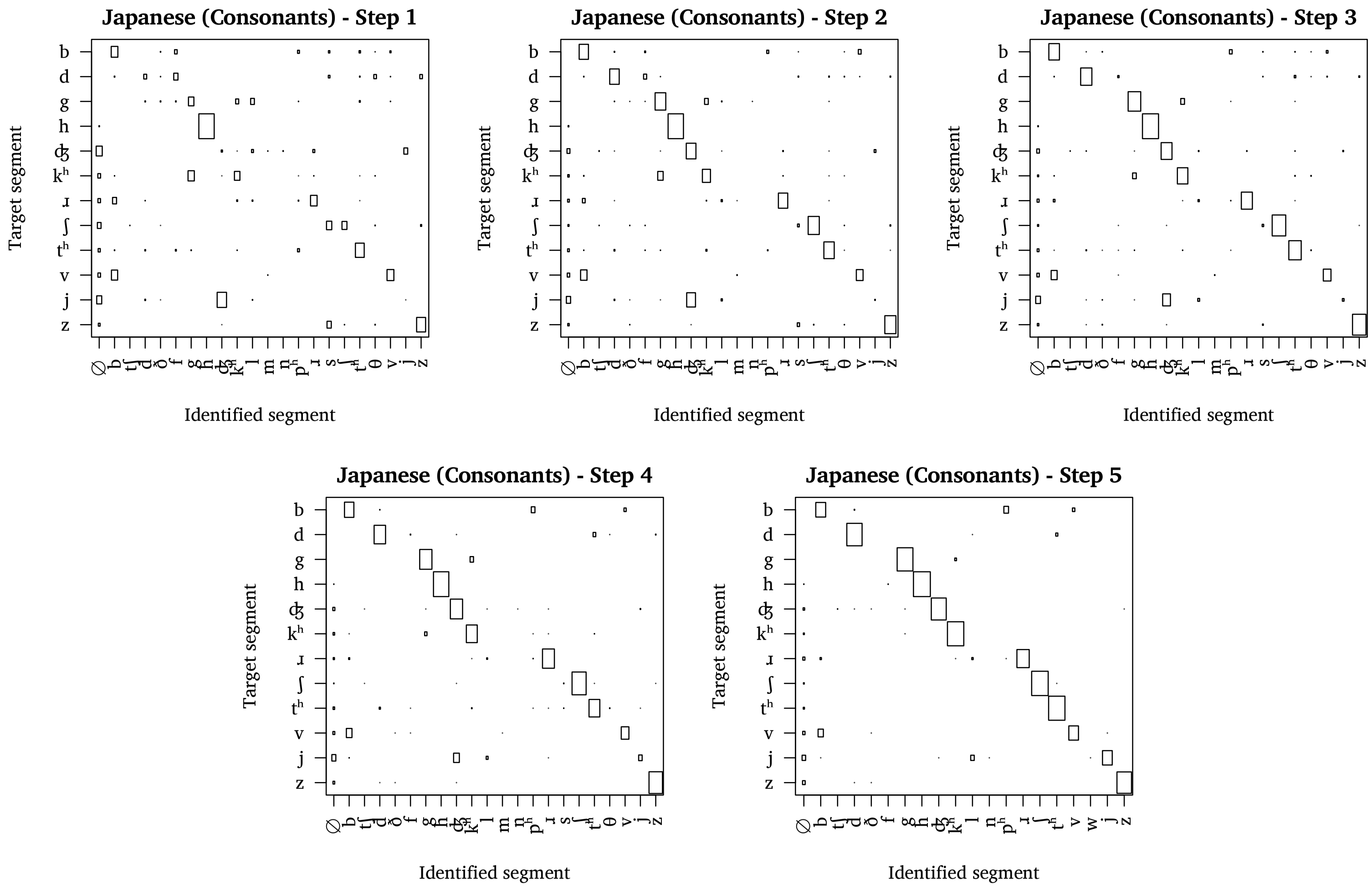


Figure F.7: Word intelligibility confusions of the Japanese cohort for consonants for the word intelligibility task of chapter 7. The area of the squares represents the proportion of responses for each category. The ⊘ category represents the elision of the target segment.

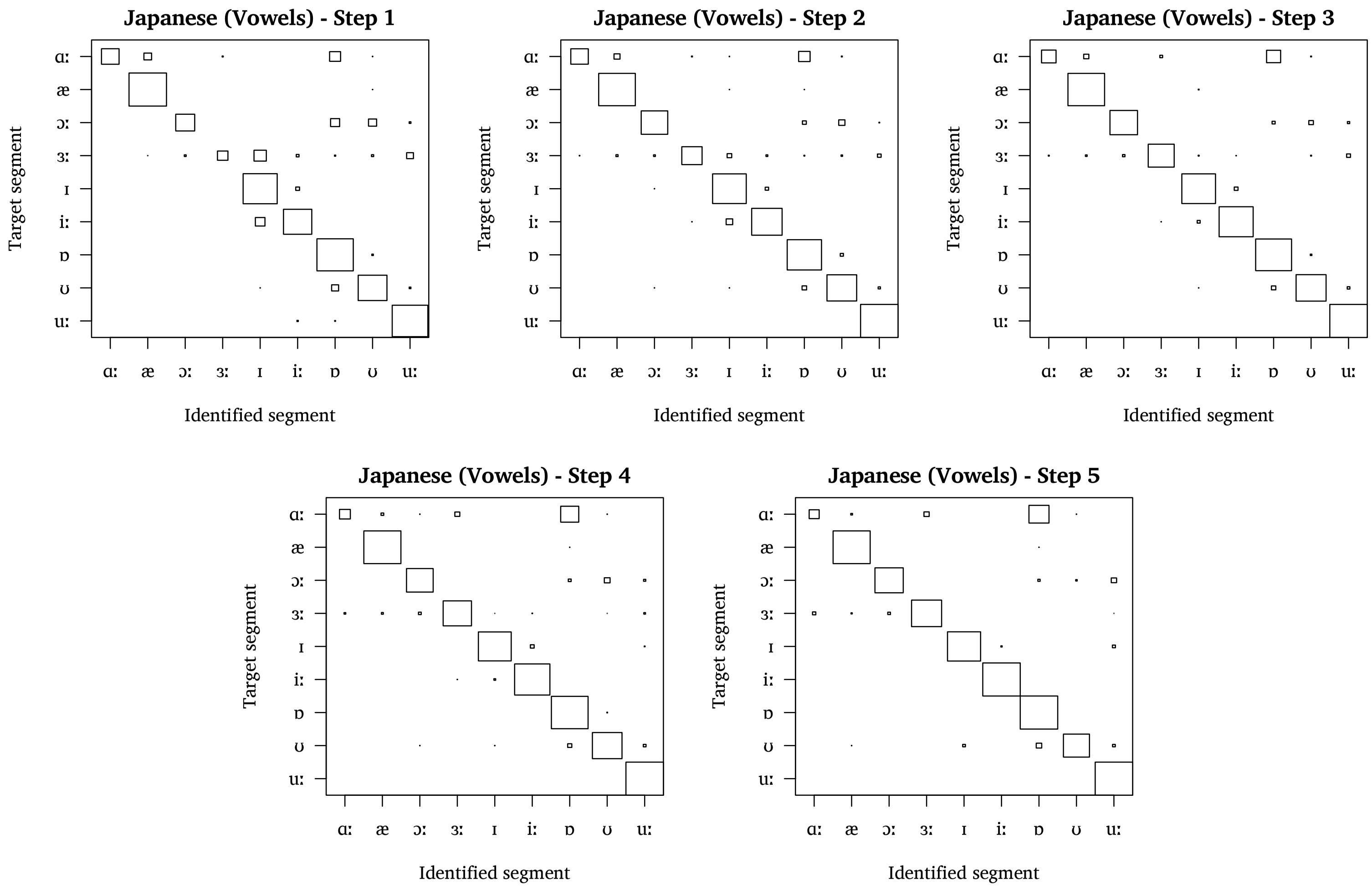


Figure F.8: Word intelligibility confusions of the Japanese cohort for vowels for the word intelligibility task of chapter 7. The area of the squares represents the proportion of responses for each category.

List of Figures

List of Tables

References

A. J. Clark, R., Richmond, K., & King, S. (2004, July). Festival 2 - build your own general-purpose unit selection speech synthesiser. In *5th ISCA Speech Synthesis Workshop* (pp. 147–151).

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992, December). The relationship between native speaker judgments of nonnative pronunciation and deviance in segments, prosody, and syllable structure. *Language Learning*, *42*(4), 529–555. doi: 10.1111/j.1467-1770.1992.tb01043.x

Assefi, M., Liu, G., Wittie, M. P., & Izurieta, C. (2015, October). *An experimental evaluation of Apple Siri and Google speech recognition.*

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Baugh, J. (2003). Linguistic profiling. In *Black linguistics: language, society, and politics in Africa and the Americas.* Psychology Press.

Beep. (2017). *Beep: Phonemic transcriptions of over 250,000 English words. (British English pronunciations).*

Behravan, H., Hautamaki, V., Siniscalchi, S. M., Kinnunen, T., & Lee, C.-H. (2016, January). i-vector modeling of speech attributes for automatic foreign accent recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *24*(1), 29–41. doi: 10.1109/taslp.2015.2489558

Beinhoff, B. (2013). *Perceiving identity through accent.* Peter Lang UK. doi: 10.3726/978-3-0353-0454-1

Bent, T., & Bradlow, A. R. (2003, September). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600–1610. doi: 10.1121/1.1603234

Best, C. T. (1995). A direct realist view of cross-language speech. In C. Goodman & H. Nusbaum (Eds.), *Speech perception and linguistic experience: theoretical and methodological issues* (pp. 171–204). Baltimore: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In M. J. Munro & O.-S. Bohn (Eds.), (p. 13-34). Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer [computer program]. version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/.*

Borsky, S., Tuller, B., & Shapiro, L. P. (1998, May). "how to milk a coat:" the effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, *103*(5), 2670–2676. doi: 10.1121/1.422787

Boyd, S. (2003, June). Foreign-born teachers in the multilingual classroom in Sweden: The role of attitudes to foreign accent. *International Journal of Bilingual Education and Bilingualism*, *6*(3-4), 283–295. doi: 10.1080/13670050308667786

Bradlow, A. R., & Bent, T. (2002, July). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, *112*(1), 272–284. doi: 10.1121/1.1487837

Burda, A. N., Hageman, C. F., Scherz, J. A., & Edwards, H. T. (2003, August). Age and understanding speakers with Spanish or Taiwanese accents. *Perceptual and Motor Skills*, *97*(1), 11–20. doi: 10.2466/pms.2003.97.1.11

Childers, D. G., & Kesler, S. B. (1978). *Modern spectrum analysis* (I. Press, Ed.).

Cho, K., & Harris, J. G. (2006). Towards an automatic foreign accent reduction tool. In *Proc. 3rd intl. conf. on speech prosody.*

Council of Europe. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment.* Cambridge University Press.

Dall, R., Yamagishi, J., & King, S. (2014). Rating naturalness in speech synthesis: The effect of style and expectation. *Interspeech Proceedings*.

del Puerto, F. G., García Lecumberri, M. L., & Lacabex, E. G. (2014, mar). The assessment of foreign accent and its communicative effects by naïve native judges vs. experienced non-native judges. *International Journal of Applied Linguistics*, *25*(2), 202–224. doi: 10.1111/ijal.12063

del Puerto, F. G., Lacabex, E. G., & García Lecumberri, M. L. (2009). Testing the effectiveness of content and language integrated learning in foreign language contexts. In R. M. J. Catalán & Y. R. de Zarobe (Eds.), *Content and language integrated learning: evidence from research in Europe* (pp. 63–80). Multilingual Matters.

Derwing, T. M. (2003, June). What do ESL students say about their accents? *Canadian Modern Language Review*, *59*(4), 547–567. doi: 10.3138/cmlr.59.4.547

Derwing, T. M. (2018). Comprehensibility. In *The tesol encyclopedia of english language teaching* (p. 1-6). American Cancer Society. Retrieved from `https://`

onlinelibrary.wiley.com/doi/abs/10.1002/9781118784235.eelt0012 doi: 10.1002/9781118784235.eelt0012

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four l1s. *Studies in Second Language Acquisition*, *19*(1), 1–16.

Doleželová, A. (2009). *Fossilized pronunciation errors in advanced Czech speakers of English: Analysis & strategies [online]* [Bakalářská práce]. Retrieved from <https://is.muni.cz/th/rvj0q/>

Dudley, H., Riesz, R., & Watkins, S. (1939). A synthetic speaker. *Journal of the Franklin Institute*, *227*(6), 739–764.

Duolingo. (2015). https://www.duolingo.com/.

Eimas, P. D. (1963). The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, *6*, 206-217.

Elsa speak. (2018). https://elsaspeak.com/home.

Elzhov, T. V., Mullen, K. M., Spiess, A.-N., & Bolker, B. (2016). minpack.lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=minpack.lm (R package version 1.2-1)

Engen, K. J. V., & Peelle, J. E. (2014, August). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*. doi: 10.3389/fnhum.2014.00577

Evans, B. G., & Iverson, P. (2004, January). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *The Journal of the Acoustical Society of America*, *115*(1), 352–361. doi: 10.1121/1.1635413

Fant, G. (1960). *Acoustic theory of speech production: With calculations based on x-ray studies of Russian articulations* (Vol. 2; W. de Gruyter, Ed.). Mouton.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, *37*(3), 313-326.

Ferguson, S. H., Jongman, A., Sereno, J. A., & Keum, K. A. (2010, March). Intelligibility of foreign-accented speech for older adults with and without hearing loss. *Journal of the American Academy of Audiology*, *21*(3), 153–162. doi: 10.3766/jaaa.21.3.3

Ferragne, E., & Pellegrino, F. (2010, March). Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association*, *40*(01), 1. doi: 10.1017/s0025100309990247

Field, J. (2004, September). An insight into listeners' problems: too much bottom-up or too much top-down? *System*, *32*(3), 363–377. doi: 10.1016/j.system.2004.05.002

Field, J. (2005, September). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*(3), 399. doi: 10.2307/3588487

Flanagan, J. L. (1972). *Speech analysis synthesis and perception.* Springer Berlin Heidelberg. doi: 10.1007/978-3-662-01562-9

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (p. 233-277). Timonium, MD.

Flege, J. E. (2018). It's input that matters most, not age. *Bilingualism: Language and Cognition*, *21*(5), 919–920.

Flege, J. E., & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics*, *15*, 67–83.

Flege, J. E., Frieda, E., & Nozawa, T. (1997, April). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*(2), 169–186. doi: 10.1006/jpho.1996.0040

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995, May). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125–3134. doi: 10.1121/1.413041

Flege, J. E., & Wayland, R. (2019). The role of input in native Spanish late learners' production and perception of English phonetic segments. *Journal of Second Language Studies, 2 (1).*

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology*, *6*(1), 110–125.

García Lecumberri, M. L., Barra-Chicote, R., Pérez-Ramón, R., Yamagishi, J., & Cooke, M. (2014). Generating segmental foreign accent. In *15th annual conference of the International Speech Communication Association (Interspeech 2014).*

García Lecumberri, M. L., Cooke, M., & Cutler, A. (2010, November). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11-12), 864–886. doi: 10.1016/j.specom.2010.08.014

Gardner, R., & Lambert, W. (1959). Motivational variables in second language acquisition. *Canadian Journal of Psychology*, *13*, 191-197.

Gardner, R., & Lambert, W. (1965). Language aptitude, intelligence, and second-language achievement. *Journal of Educational Psychology*, *56*, 191-199.

Gardner, R., & Lambert, W. (1972). Motivational variables in second language acquisition. In R. Gardner & W. Lambert (Eds.), (p. 119-216). Rowley, MA: Newbury House.

Gass, S., & Varonis, E. M. (1984, March). The effect of familiarity on the comprehensibility of non-native speech. *Language Learning*, *34*(1), 65–87. doi: 10.1111/j.1467-1770.1984.tb00996.x

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363-376.

Greene, J., & Wells, E. (1927). *The cause and cure of speech disorders: A textbook for students and teachers on stuttering, stammering and voice conditions*. Macmillan.

Griffen, T. (1980). A non-segmental approach to the teaching of pronunciation. *Revue de Phonétique Appliquée*, *54*, 81–94.

Gupta, H., & Gupta, D. (2016, January). LPC and LPCC method of feature extraction in speech recognition system. In *2016 6th international conference - cloud system and big data engineering (confluence)*. IEEE. doi: 10.1109/confluence.2016.7508171

Hahn, L. D. (2004, July). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, *38*(2), 201. doi: 10.2307/3588378

Hansen, J., & Arslan, L. (1995). Foreign accent classification using source generator based prosodic features. In *1995 international conference on acoustics, speech, and signal processing*. IEEE. doi: 10.1109/icassp.1995.479824

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A. R., Jaitly, N., ... Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal processing magazine*, *29*.

Huang, B. H., & Jun, S.-A. (2014, oct). Age matters, and so may be raters. *Studies in Second Language Acquisition*, *37*(04), 623–650. doi: 10.1017/s0272263114000576

Imai, S., Walley, A., & Flege, J. E. (2005, February). Lexical frequency and neighbourhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *The Journal of the Acoustical Society of America*, *117*(2), 896–907. doi: 10.1121/1.1823291

Ioup, G., Boustagui, E., Tigi, M. E., & Moselle, M. (1994, March). Reexamining the critical period hypothesis. *Studies in Second Language Acquisition*, *16*(01), 73–98. doi: 10.1017/s0272263100012596

Jensen, C., & Thøgersen, J. (2017). Foreign accent, cognitive load and intelligibility of EMI lectures. *Nordic Journal of English Studies*, *16*(3), 107–137.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*.

Kang, O., Rubin, D., & Pickering, L. (2010, oct). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, *94*(4), 554–566. doi: 10.1111/j.1540-4781.2010.01091.x

Kat, L. W., & Fung, P. (1999). Fast accent identification and accented speech recognition. In *1999 IEEE international conference on acoustics, speech, and signal processing. proceedings. ICASSP99 (cat. no.99ch36258).* IEEE. doi: 10.1109/icassp.1999.758102

Kempelen, W. V. (1791). *Mechanismus der menschlichen sprache nebst der beschreibung seiner sprechenden maschine.* J.B. Degen. doi: 10.6083/m4fq9vfp

Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007, October). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *122*(4), 2365–2375. doi: 10.1121/1.2773986

Kim, M.-R. (2011). Native and non-native English speakers' VOT productions of stops. *The Linguistic Association of Korea Journal*.

Kissling, E. M. (2014, August). Phonetics instruction improves learners' perception of L2 sounds. *Language Teaching Research*, *19*(3), 254–275. doi: 10.1177/1362168814541735

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). Springer.

Kuhl, P. K. (2000, oct). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850–11857. doi: 10.1073/pnas.97.22.11850

Lazaridis, A., Potard, B., & Garner, P. N. (2015). DNN-based speech synthesis: Importance of input features and training data. In *Speech and computer* (pp. 193–200). Springer International Publishing. doi: 10.1007/978-3-319-23132-7_24

Lenneberg, E. H. (1967). *Biological foundations of language.* John Wiley and Sons.

Lev-Ari, S., & Keysar, B. (2010, November). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. doi: 10.1016/j.jesp.2010.05.025

Li, F., Edwards, J., & Beckman, M. E. (2009, January). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, *37*(1), 111–124. doi: 10.1016/j.wocn.2008.10.001

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech*, *54*, 175–195.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, *61*, 379–388.

Lombard, E. (1911). Le signe de l'elevation de la voix. *Ann. Mal. de L'Oreille et du Larynx*, 101–119.

Long, M. H. (1990). Maturational constraints on language development. *Studies in second language acquisition*, *12*(3), 251–285.

Lybeck, K. (2002, April). Cultural identification and second language pronunciation of Americans in Norway. *The Modern Language Journal*, *86*(2), 174–191. doi: 10.1111/1540-4781.00143

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*, 369–390.

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology.* Lawrence Erlbaum Associates, Inc., Publishers.

Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, *36*(2), 173. doi: 10.2307/3588329

Martínez-Paricio, V., Koreman, J., Husby, O., Eggesbø Abrahamsen, J., & Bech, O. (2014, September). *Expanding CALST: Multilingual analysis of L1-L2 phonotactics for language teaching.*

Morrison, G. S. (2002). Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: Longitudinal results. In *Proceedings of the north west linguistics conference 2002.*

Moulines, E., & Charpentier, F. (1990, December). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*(5-6), 453–467. doi: 10.1016/0167-6393(90)90021-z

Muñoz, C., Gallardo, P., & Llurda, E. (2000). *The effects of age on rate of acquisition of a foreign language.* Universitat de Lleida.

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition*, *20*(2), 139–154.

Munro, M. J. (2008). Foreign accent and speech intelligibility. In K. de Bot & D. Ayoun (Eds.), (p. 193-218). John Benjamins Publishing Company.

Munro, M. J., & Derwing, T. M. (1995a, March). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (1995b, July). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289–306. doi: 10.1177/002383099503800305

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech. *Studies in Second Language Acquisition*, *23*(4), 451–468.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006, February). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*(01). doi: 10.1017/s0272263106060049

Nelson, C. (1982). Intelligibility and non-native varieties of English. In B. B. Kachru (Ed.), *The other tongue: English across cultures* (pp. 58–73). Urbana: University of Illinois Press.

Nielsen, A. H., Horn, N. T., Sørensen, S. D., McGregor, W. B., & Wallentin, M. (2015, November). Intensive foreign language learning reveals effects on categorical perception of sibilant voicing after only 3 weeks. *i-Perception*, *6*(6), 204166951561367. doi: 10.1177/2041669515613674

Nishi, K., & Kewley-Port, D. (2007, December). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1496–1509. doi: 10.1044/1092-4388(2007/103)

Ohara, Y. (2001). Finding one's voice in Japanese: A study of the pitch levels of L2 users. In A. Pavlenko, A. Blackledge, I. Piller, & M. Teutsch-Dwyer (Eds.), *Multilingualism, second language learning, and gender* (pp. 231–254). Berlin, New York: De Gruyter Mouton. Retrieved from http://www.degruyter.com/view/books/9783110889406/9783110889406.231/9783110889406.231.xml doi: 10.1515/9783110889406.231

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of psycholinguistic research*, *5*(3), 261–283.

Pinet, M., Iverson, P., & Huckvale, M. (2011, September). Second-language experience and speech-in-noise recognition: Effects of talker–listener accent similarity. *The Journal of the Acoustical Society of America*, *130*(3), 1653–1662. doi: 10.1121/1.3613698

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*, 191–215.

Piske, T., & MacKay, R. (1999). Age and L1 use effects on degree of foreign accent in English. In *Proceedings of the 14th International Congress of Phonetics Sciences* (pp. 1433–1436).

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*, 253–260.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustic Society of America*, *55*(2), 328–333.

Poltrock, S., & Nazzi, T. (2015, March). Consonant/vowel asymmetry in early word form recognition. *Journal of Experimental Child Psychology*, *131*, 135–148. doi: 10.1016/j.jecp.2014.11.011

Polyanskaya, L. (2015). *Contribution of prosodic timing patterns into perceived foreign accent* (Unpublished doctoral dissertation).

Polyanskaya, L., Ordin, M., & Busa, M. G. (2016, May). Relative salience of speech rhythm and speech rate on perceived foreign accent in a second language. *Language and Speech*, *60*(3), 333–355. doi: 10.1177/0023830916648720

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1992). *Numerical recipes in Fortran 77: The art of scientific computing.* Cambridge University Press.

Purcell, E. T., & Suter, R. W. (1980, December). Predictors of pronunciation accuracy: a re-examination. *Language Learning*, *30*(2), 271–287. doi: 10.1111/j.1467-1770.1980.tb00319.x

Quilis, A., & Esgueva, M. (1983). Realización de los fonemas vocálicos españoles en posición fonética normal. In M. Esgueva & M. Cantero (Eds.), *Estudios de fonética (collectanea phonetica).* Consejo Superior de Investigaciones Científicas, Instituto "Miguel de Cervantes". Retrieved from `https://www.amazon.com/Estudios-fonetica-Collectanea-phonetica-Spanish/dp/8400053192?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=8400053192`

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rajadurai, J. (2007, February). Intelligibility studies: a consideration of empirical and ideological issues. *World Englishes*, *26*(1), 87–98. doi: 10.1111/j.1467-971x.2007.00490.x

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In *Speech and language* (pp. 243–335). Elsevier. doi: 10.1016/b978-0-12-608610-2.50012-1

Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007, July). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, *35*(3), 439–443. doi: 10.1016/j.wocn.2006.01.002

Rogers, C., Dalby, J., & Nishi, K. (2004, June). Effects of noise and proficiency on intelligibility of Chinese-accented English. *Language and Speech*, *47*(2), 139–154. doi: 10.1177/00238309040470020201

Rogers, J., & Davis, M. (2009, January). Categorical perception of speech without stimulus repetition. *Proceedings of Interspeech, Brighton*, 376-379.

Romportl, M. (1973). On the Czech system of consonants. *Studies in Phonetics*, 105–117.

Rosenberg, A., & Ramabhadran, B. (2017). Bias and statistical significance in evaluating speech synthesis with mean opinion scores. In *Proc. Interspeech 2017* (pp. 3976–3980). Retrieved from `http://dx.doi.org/10.21437/Interspeech.2017-479` doi: 10.21437/Interspeech.2017-479

Ryu, H., Hong, H., Kim, S., & Chung, M. (2016, December). Automatic pronunciation assessment of Korean spoken by L2 learners using best feature set selection. In *2016 Asia-Pacific Signal and Information Processing Association annual summit and conference (APSIPA)*. IEEE. doi: 10.1109/apsipa.2016.7820673

Saito, K., Dewaele, J.-M., & Hanzawa, K. (2017, February). A longitudinal investigation of the relationship between motivation and late second language speech learning in classroom settings. *Language and Speech*, *60*(4), 614–632. doi: 10.1177/0023830916687793

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics.* Springer Netherlands.

Sankowska, J., García Lecumberri, M. L., & Cooke, M. (2011, January). Interaction of intrinsic vowel and consonant durational correlates with foreigner directed speech. *Poznań Studies in Contemporary Linguistics*, *47*. doi: 10.2478/psicl-2011-0009

Schneider, K. (2013). *The German boundary tones: categorical perception, perceptual magnets, and the perceptual reference space.* Universität Stuttgart. doi: 10.18419/opus-2999

Sereno, J., Lammers, L., & Jongman, A. (2015, January). The relative contribution of segments and intonation to the perception of foreign-accented speech. *Applied Psycholinguistics*, *37*(02), 303–322. doi: 10.1017/s0142716414000575

Shin, D.-J. (2018, May). Effect of experience to language on speech-in-noise recognition for Korean L2 speakers. *Studies in English Language & Literature*, *44*(2), 301–319. doi: 10.21559/aellk.2018.44.2.015

Silva, C. C., & Barbosa, P. A. (2018, January). The contribution of prosody to foreign accent: A study of Spanish as a foreign language. *Loquens*, *4*(2), 041. doi: 10.3989/loquens.2017.041

Singleton, D. M., & Ryan, L. (2004). *Language acquisition: The age factor* (Vol. 9). Multilingual Matters.

Skarnitzl, R., & Machač, P. (2011). Principles of phonetic segmentation. *Phonetica*, *68*(3), 198–199. doi: 10.1159/000331902

Skarnitzl, R., & Šturm, P. (2014, June). Assimilation of voicing in Czech speakers of English: The effect of the degree of accentedness. *Research in Language*, *12*(2), 199–208. doi: 10.2478/rela-2014-0007

Skarnitzl, R., & Šturm, P. (2016, mar). Pre-fortis shortening in Czech English: A production and reaction-time study. *Research in Language*, *14*(1), 1–14. doi: 10.1515/rela-2016-0005

Smith, L. E., & Nelson, C. L. (1985, November). International intelligibility of English: directions and resources. *World Englishes*, *4*(3), 333–342. doi: 10.1111/j.1467-971x.1985.tb00423.x

Southwood, M. H., & Flege, J. E. (1999). Scaling foreign accent: direct magnitude estimation versus interval scaling. *Clinical linguistics & phonetics*, *13*(5), 335–349.

Stella. (1985). *Pattern Playback Machine.* Retrieved from `http://www.haskins.yale.edu/featured/heads/SIMULACRA/playback.html`

Stibbard, R. M., & Lee, J.-I. (2006, July). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *The Journal of the Acoustical Society of America*, *120*(1), 433–442. doi: 10.1121/1.2203595

Streetharan, C. S. (2004, January). Students, sarariiman (pl.), and seniors: Japanese men's use of manly speech register. *Language in Society*, *33*(01). doi: 10.1017/s0047404504031045

Stuart-Smith, J. (2008). Scottish English: Phonology. *Varieties of English*, *1*, 48–70.

Suter, R. W. (1976, December). Predictors of pronunciation accuracy in second language learning. *Language Learning*, *26*(2), 233–253. doi: 10.1111/j.1467-1770.1976.tb00275.x

Tajima, K., Port, R., & Dalby, J. (1997, January). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, *25*(1), 1–24. doi: 10.1006/jpho.1996.0031

Taylor, B., Dey, A., Siewiorek, D., & Smailagic, A. (2015). Using physiological sensors to detect levels of user frustration induced by system delays. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing - UbiComp '15.* ACM Press. doi: 10.1145/2750858.2805847

Tokuda, K., Yoshimura, T., Masuko, T., Kobayashi, T., & Kitamura, T. (2000). Speech parameter generation algorithms for HMM-based speech synthesis. In *2000 IEEE international conference on acoustics, speech, and signal processing. proceedings (cat. no.00ch37100).* IEEE. doi: 10.1109/icassp.2000.861820

Tollfree, L. (2014). South East London English: discrete versus continuous modelling of consonantal reduction. In P. Foulkes & G. J. Docherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 163–184). Routledge.

Tomaschek, F., Truckenbrodt, H., & Hertrich, I. (2011). Processing German vowel quantity: Categorical perception or perceptual magnet effect? In *Icphs* (pp. 2002–2005).

Trofimovich, P., & Isaacs, T. (2012, May). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(04), 905–916. doi: 10.1017/s1366728912000168

Van Dommelen, W. (2018, November). The production of Norwegian vowels by French and Russian speakers.

van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002, April). Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America*, *111*(4), 1906–1916. doi: 10.1121/1.1456928

Varonis, E. M., & Gass, S. (1982, March). The comprehensibility of non-native speech. *Studies in Second Language Acquisition*, *4*(02), 114. doi: 10.1017/s027226310000437x

Volín, J., & Skarnitzl, R. (2010, November). The strength of foreign accent in Czech English under adverse listening conditions. *Speech Communication*, *52*(11-12), 1010–1021. doi: 10.1016/j.specom.2010.06.009

Wang, H., & van Heuven, V. J. (2015, November). The interlanguage speech intelligibility benefit as bias toward native-language phonology. *i-Perception*, *6*(6), 204166951561366. doi: 10.1177/2041669515613661

Watt, D., Llamas, C., & Johnson, D. E. (2014). Sociolinguistic variation on the Scottish-English border. In *Sociolinguistics in Scotland* (pp. 79–102). Palgrave Macmillan UK. doi: 10.1057/9781137034717_5

Weil, S. A. (2003). *The impact of perceptual dissimilarity on the perception of foreign accented speech* (Unpublished doctoral dissertation). Ohio State University.

Whitley, M. (2002). *Spanish/English contrasts: A course in Spanish linguistics*. Georgetown University Press. Retrieved from https://books.google.es/books?id=yyqU_tXek1EC

Yusnita, M., Paulraj, M., Yaacob, S., Bakar, S. A., & Saidatul, A. (2011, November). Malaysian English accents identification using LPC and formant analysis. In *2011 IEEE international conference on control system, computing and engineering.* IEEE. doi: 10.1109/iccsce.2011.6190572

Zampini, M., & Green, K. (2001). The voicing contrast in English and Spanish: The relationship between perception and production. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (p. 23-48). Malden, MA: Blackwell.

Zielinski, B. W. (2008, March). The listener: No longer the silent partner in reduced intelligibility. *System*, *36*(1), 69–84. doi: 10.1016/j.system.2007.11.004

Zuengler, J. (1988, February). Identity markers and L2 pronunciation. *Studies in Second Language Acquisition*, *10*(01), 33. doi: 10.1017/s027226310000694x

Resumen de la tesis en castellano

La producción oral de hablantes no nativos suele caracterizarse por la presencia de rasgos que difieren de lo que los hablantes nativos consideran pronunciación estándar, generando lo que se conoce como acento extranjero. El acento extranjero se ha estudiado tradicionalmente desde un punto de vista holístico, es decir, como un fenómeno de habla general y no como la suma de rasgos individuales que pueden variar entre sí en diferente grado (Bradlow, 2002; Southwood, 1999). Los pocos estudios sobre acento extranjero que se han centrado en atributos individuales lo han hecho generalmente poniendo el foco en rasgos suprasegmentales tales como duración (Tajima, Port & Dalby, 1997), entonación y pausas (Kang, Rubin, & Pickering, 2010)2010) o velocidad de elocución (Munro & Derwing, 2001). Sin embargo, la influencia de la realización incorrecta de segmentos en la percepción de acento extranjero no ha sido aún investigada en detalle.

El objetivo de esta tesis es, precisamente, recabar información sobre cómo se percibe el acento extranjero desde un punto de vista segmental. En particular, nuestro objetivo es comprobar cómo influyen distintos grados de acento extranjero español en la percepción de palabras inglesas. Con este fin describimos y comparamos tres técnicas para generar acento extranjero segmental: code-switching, que consiste en cambios voluntarios del locutor para pronunciar un solo segmento con acento extranjero; splicing, que, mediante manipulación acústica, intercambia un segmento de una palabra inglesa por su equivalente acentuado en español; y síntesis de voz, en la que se generan dos modelos estadísticos (uno de la lengua nativa y otro de la lengua extranjera) para sintetizar cada segmento de la palabra con el acento que se le indique.

Mediante estas técnicas generamos un corpus de palabras inglesas en las que solo un segmento fue reemplazado por su correspondiente versión acentuada. Los segmentos del inglés seleccionados fueron [h, ɹ, k^h, t^h, v, w, ð, ʤ, j], cuyas correspondientes versiones con acento español son [x, r, k, t, b, gw, d, j, ʤ] respectivamente. Con el fin de comparar el efecto del acento extranjero segmental en oyentes con distinto perfil lingüístico, tres grupos de oyentes (nativos ingleses, españoles y checos) fueron reclutados para evaluar los estímulos generados en términos de inteligibilidad y grado de acento extranjero. Los resultados de este experimento revelaron que el sistema fonológico de la primera lengua de los oyentes es crucial a la hora de juzgar el acento extranjero. Algunos segmentos como [ʤ]/[j], tienen la particularidad de ser distinguidos fonémicamente en inglés mientras que, en español, son variantes de un mismo fonema /j/. Esto hace que los nativos ingleses perciban especialmente acentuada la realización de [ʤ] como [j], por ejemplo, pero no así los españoles, para quienes estos segmentos aparecen mayoritariamente en distribución libre. Otro hallazgo es que a nivel segmental, y como ya se ha demostrado en palabras o frases (cita Munro), la percepción de un segmento como especialmente acentuado no siempre implica pérdida de inteligibilidad. Particularmente, de entre los segmentos seleccionados para este experimento, las versiones acentuadas de [h] y [ɹ] (i.e. [x] y [r]) fueron evaluadas con un alto grado de acento extranjero por los tres grupos de participantes, pero no acarrearon efectos negativos en la inteligibilidad de palabras acentuadas.

Los resultados fueron analizados a la luz del llamado *Matched Interlanguage Intelligibility Benefit*. Esta hipótesis, formulada en Bent & Bradlow (2003), sugiere que los oyentes no nativos expuestos a hablantes no nativos con su misma L1 sufren una menor pérdida de inteligibilidad que los oyentes nativos. Como extensión a este efecto, las autoras describen el *Mismatched Interlanguage Intelligibility Benefit*, que sugiere que los oyentes no nativos con una L1 distinta a la del hablante no nativo también sufrirán menos pérdida de inteligiblidad que los oyentes nativos. La hipótesis del Matched Interlanguage Intelligibility Benefit ha sido cuestionada en Stibbard & Lee (2006), en cuyo estudio no se encontró un beneficio y que incluso mostró un detrimento en ciertos casos en los que la L1 de hablante y oyente no nativos no era la misma. En nuestro experimento, los oyentes con distinta L1 respecto al hablante sufrieron una mayor pérdida de inteligibilidad que los oyentes con la misma L1 que el hablante. Este déficit es altamente dependiente del segmento evaluado y el sistema fonológico del oyente.

La investigación en acento extranjero también ha demostrado que segmentos individuales pueden presentar distintos grados de desviación de la norma (Best, 1995). Es posible que distintos grados de acento extranjero desencadenen respuestas distintas por parte de los oyentes. Por este motivo, hemos desarrollado una técnica para generar distintos grados acústicos de acento extranjero, esto es, una técnica que facilita la generación de estímulos con una desviación acústica específica respecto al estímulo nativo. En el segundo experimento (experimento 2) de esta tesis queremos comprobar si distintos grados de acento extranjero afectan por igual a grupos con distinta L1 (ingleses y españoles) y distinto nivel de competencia en inglés (españoles con competencia alta y españoles con competencia baja).

De entre los segmentos del experimento 1, seis fueron elegidos para comprobar la incidencia de distintos grados de acento extranjero: [h] y [ɹ], por su saliencia en cuanto a acento extranjero con respecto a otros segmentos; [ʤ] y [j] por la distinta representación que tienen estos dos segmentos en el sistema fonológico de ingleses y españoles; y, finalmente, [t^h] y [v] porque acarrearon un grado de acento similar para oyentes ingleses

Sonido nativo	**Sonido extranjero**	**Palabra 1**	**Palabra 2**
ɹ	r	rainbow	reason
ʤ	j	gipsy	gender
j	ʤ	user	yours
h	x	hammer	happen
tʰ	t	tea	type
v	b	veil	vanish

Tabla 1: Corpus empleado en el experimento.

y españoles. Para cada uno de estos seis segmentos se eligieron dos palabras en los que el segmento aparecía en posición inicial (tabla 1). Con estas 12 palabras se generaron sendos continuos de 9 pasos con la técnica de gradación de acento extranjero. Los participantes juzgaron los estímulos en tareas de categorización de acento extranjero (2AFC) y de discriminación (AX) en las que tuvieron que juzgar, respectivamente, si los estímulos sonaban nativos o extranjeros y si pares de estímulos separados en dos pasos eran iguales o distintos.

Los resultados de la tarea de categorización revelaron que el acento extranjero es percibido, de forma general, categorialmente, es decir, los oyentes escuchan dos categorías distintas en los dos extremos del continuo y el cambio en la categoría percibida no es gradual, sino que se produce de forma abrupta. Uno de los principales hallazgos de este experimento es que, pese a que todos los continuos fueron percibidos categorialmente por los oyentes nativos, el punto del continuo en el que se produjo el cambio de categoría percibida varió mucho de unos segmentos a otros. Los segmentos cuya contraparte acentuada es más saliente ([h] y [ɹ]) fueron percibidos como completamente extranjeros con muy poca desviación del segmento nativo; sin embargo, [tʰ] y [v] mostraron más resistencia al cambio de categoría, y solo fueron percibidos como acentuados después de ser sometidos a una gran desviación acústica.

También se encontraron diferencias en la forma de percibir estos continuos por parte de oyentes nativos y no nativos. Específicamente, los oyentes no nativos percibieron el extremo no nativo de los continuos de [ʤ] y [j] mucho menos acentuados que los oyentes nativos. Esto significa que los oyentes no nativos son menos capaces de percibir la diferencia entre estos dos segmentos que los oyentes nativos debido a la representación de ambos en sus respectivos sistemas fonológicos (Flege, 1995; Imai, 2005). Sin embargo, los oyentes españoles con más competencia en inglés mostraron capacidades más similares a la de los nativos que los menos competentes, lo que sugiere que nuevas categorías pueden emerger a través del aprendizaje de segundas lenguas. Esto contrasta con los resultados de la tarea de discriminación, en la que no se encontraron diferencias significativas entre los grupos en ningún continuo. En Pisoni (1973) se sugiere que, mientras que una tarea de categorización activa el procesamiento fonético del oyente (esto es, procesamiento de arriba abajo o top-down), la tarea de discriminación da preferencia al procesamiento auditivo (procesamiento de abajo arriba o bottom-up). A la luz de nuestros resultados, podemos afirmar que los oyentes tienen una capacidad auditiva similar, pero la capacidad fonética depende en alto grado de su perfil lingüístico.

En el siguiente experimento (experimento 3) el repertorio de segmentos se amplía considerablemente y se presenta una adaptación de la técnica de gradación para poder generar también diversos grados de acento extranjero con vocales. Aparte de los seis segmentos experimentales del experimento 2, se añaden tres consonantes más en posición inicial, tres consonantes en posición final y 11 vocales (tablas 2 y 3). Para cada uno de estos segmentos se seleccionaron cuatro palabras y se generaron continuos de 21 pasos.

Continuo	Palabras			
[x]→[h]	help	hide	hole	house
[r]→[ɹ]	red	rent	rhyme	risk
[k]→[kʰ]	cast	code	cold	kiss
[t]→[tʰ]	town	tile	toad	tone
[b]→[v]	van	veil	valve	view
[d]→[ð]	than	that	this	thus
[s]→[ʃ]	shoe	short	shape	sharp
[s]→[z]	zap	zone	zoo	zoom
[j]→[ʤ]	jam	jaw	june	jest
[ʤ]→[j]	yak	yoke	years	youth
[_f]→[_b]	cab	nib	rib	crab
[_θ]→[_d]	god	load	dad	food
[_x]→[_g]	dog	frog	leg	smog

Tabla 2: Continuos generados para consonantes. Las palabras en gris son aquellas cuyas versiones nativa y acentuada dan lugar a un par mínimo.

Un grupo de nativos ingleses evaluó una muestra representativa de 5 pasos equidistantes de cada continuo (pasos 1, 6, 11, 16 y 21) en una tarea de Mean Opinion Score (MOS) en la que fueron preguntados por la calidad de la señal en una escala de 5 puntos. Además, los participantes también juzgaron todos los continuos en una tarea 2AFC de categorización de acento extranjero, similar a la del experimento 2.

En general, la calidad de los continuos generados fue juzgada como buena (es decir, ningún ítem fue juzgado por debajo de 4 puntos sobre 5). Podemos determinar, por lo tanto, que los resultados extraídos de los continuos generados con la técnica de gradación son medidas fiables de acento extranjero y no están afectados por una reducción de calidad debido a las manipulaciones acústicas.

Continuo	Palabras			
[e]→[ɜː]	bird	burn	firm	learn
[o]→[ɜː]	word	world	worm	worse
[a]→[æ]	back	cat	clap	pact
[a]→[ɑː]	fast	raft	shark	stark
[a]→[ʌ]	cut	drum	gun	nut
[i]→[ɪ]	clip	mist	pick	sin
[i]→[iː]	beam	seem	steam	team
[o]→[ɒ]	cost	dot	pot	spot
[o]→[ɔː]	clause	fall	orb	storm
[u]→[ʊ]	look	nook	put	should
[u]→[uː]	choose	mood	moon	spoon

Tabla 3: Continuos generados para vocales.

Los resultados de la tarea de percepción categorial arrojaron resultados complementarios a los hallazgos del experimento 2. Descubrimos que no todos los segmentos no nativos son necesariamente evaluados como extremadamente acentuados. De hecho, muchos extremos no nativos fueron evaluados por encima del 25% de acento extranjero, y algunos incluso no fueron percibidos como extranjero en absoluto (e.g. [d]→[ð]). Además, no todos los continuos fueron percibidos categorialmente. Repp (1984) sugiere que la percepción categorial parece más ligada a las consonantes, mientras que el cambio de una vocal a otra se percibe como más gradual. Nuestros resultados respaldan esta afirmación. Igual que en el experimento 2, se detectó un alto grado de variación entre segmentos en lo que respecta al paso del continuo en el que se produce el cambio en categoría percibida, lo que sugiere que los resultados de estudios holísticos son consecuencia de varios eventos simultáneos con efectos diversos.

Como el objetivo de esta tesis es comparar el efecto del acento extranjero en oyentes con diferentes perfiles lingüísticos, es conveniente normalizar la percepción de continuos acústicos usando el juicio de los oyentes ingleses para eliminar variación inter-segmental y facilitar la comparación entre grupos de oyentes. Para hacer esta normalización, aplicamos la técnica de gradación de acento extranjero de tal modo que los continuos generados hasta ahora, compuestos por desviaciones acústicas de igual magnitud respecto del prototipo nativo (continuos iso-acústicos), son transformados en continuos compuestos por desviaciones perceptivas (continuos iso-perceptivos).

Para generar los continuos iso-perceptivos, oyentes ingleses nativos juzgaron en una tarea de percepción categorial continuos de 16 pasos generados a partir de la voz de cuatro hablantes bilingües (experimento 4). Los resultados de esta tarea sirvieron para elegir al hablante que más se ajustaba a los requerimientos de la técnica de gradación y para generar un grupo de continuos iso-perceptivos de 5 pasos.

En el último experimento de la tesis (experimento 5), cuatro grupos con distinta L1 (ingleses, españoles, checos y japoneses) evaluaron perceptivamente los continuos iso-perceptivos en cuatro tareas: inteligibilidad de palabras, discriminación (AX), categorización de acento extranjero (2AFC) e identificación de segmentos en sílabas aisladas. La inclusión de esta última tarea responde al deseo de controlar lo máximo posible el efecto del nivel léxico en la percepción de acento extranjero segmental.

El principal hallazgo de este experimento es que la percepción de acento extranjero segmental es en gran medida dependiente del sistema fonológico del oyente. Según la representación fonológica del segmento nativo, el segmento acentuado y la relación entre ambos, el acento extranjero tendrá más o menos repercusiones. Por ejemplo, en un continuo como [i]→[iː] en el que solo varía la duración, los oyentes españoles y japoneses perciben menos acento extranjero que los oyentes ingleses y checos, para quienes la cantidad vocálica es un rasgo distintivo.

La comparación de la tarea de inteligibilidad de palabras con la de identificación de segmentos en sílabas aisladas reveló que existe un efecto léxico, es decir, un cambio en la identificación del segmento acentuado según esté o no en contexto léxico. En general, los ingleses fueron los más beneficiados por este contexto léxico, lo cuál es indicativo de que la disponibilidad del léxico ayuda en gran medida a desencriptar segmentos pronunciados con acento extranjero.

Por último, en este experimento encontramos que las consonantes y las vocales acentuadas no contribuyen de igual manera a la percepción de acento extranjero. De acuerdo con los resultados, los oyentes (especialmente los ingleses) se apoyan más en las vocales que en las consonantes para comprender un mensaje, es decir, las pronunciaciones acentuadas de vocales afectan a la inteligibilidad más que las de consonantes.

Los resultados de esta tesis revelan que el acento extranjero segmental tiene efectos que han pasado desapercibidos en estudios holísticos hasta la fecha. La metodología de investigación propuesta en nuestro trabajo puede ayudar a facilitar la integración de personas en comunidades distintas a la suya, rodeados de personas con un acento que les puede resultar extraño.