VARIATION IN PASSING FOR A NATIVE SPEAKER: ACCENTEDNESS IN SECOND LANGUAGE SPEAKERS OF ENGLISH IN PRODUCTION AND PERCEPTION

A thesis submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in Linguistics in the University of Canterbury

by Ksenia Gnevsheva

University of Canterbury

2015

Table of Contents

List of figures

List of tables

Acknowledgements

Words fail to describe how indebted I am to my supervisors Kevin Watson and Jen Hay. I received support from Kevin on so many levels: as a supervisor, as a mentor, and as a friend he helped me become the more confident researcher, teacher, and writer that I am today. I am also extremely thankful to Jen for helpful discussions, R help, and constructive criticisms that have improved this work.

Of course, it takes a whole village to raise a child. My experience at the University of Canterbury would not have been complete without the encouragement and support of other faculty and students. The socio meetings, the brownbag, the lunches, and outings helped me see the bigger picture as well as kept me sane. I would like to thank Lynn Clark, Una Cunningham, Beth Hume, Lucy Johnston, Heidi Quinn, and especially Pat La Shell for statistics support and Robert Fromont for all the ANNE help. My days were made brighter by Daniel Bürkle, Jiao Dan, Sarah van Eyndhoven, Romain Fiasson, Annalise Fletcher, Andy Gibson, Ahmad Haider, Matthias Heyne, Wakayo Mattingley, Kate Naitoro, Jacqui Nokes, Darcy Rose, Keyi Sun, and Xuan Wang. I would also like to thank the University of Canterbury for funding this research through the Doctoral scholarship, the Thesis in Three award, and several small research grants from the School of Language, Social, and Political Sciences.

I am very grateful to my family and friends outside of academia for all the support they have provided. I give my thanks to Jamie, who moved to the other side of the planet to support me in my academic endeavors and who has been by my side through thick and thin; my parents, who kindled my love for enquiry and knowledge; and the life-long friends that support and cheer me up from thousands of miles away.

Last but not least, the success of this research was totally dependent on the participants' willingness to sacrifice their time, so precious as most of them were in the same PhD boat as me. I am grateful to Al, Amy, BrandiATMuhkuh, Dakota, Emily, Gabriella, Grace, Han, Hesse, Jack, Jack Brown, Jess, Kahui, Lea, Linda, Louisa, M, participant12, Sam, Samoth, Sarah, Seung, Vincent, and Zwerg (you know who you are) for their participation and enthusiasm, which, in turn, inspired me as well.

Abstract

This thesis reports on findings from a study of sociolinguistic variation in second language speakers of English in New Zealand. The study combines quantitative methods of acoustic analysis and experimental design with qualitative methods of semi-structured interviews and content analysis. The study focuses on second language speakers' variation in 'passing for a native speaker', that is, being regarded as a first language speaker.

Variation in passing is explored from the perspectives of variation in production and perception. 18 second language speakers of English (first language Korean and German) and 6 first language speakers of English were recorded in four different settings (family, friends, services, and university). In the production study, the second language speakers' monophthongal vowels are analyzed in comparison with the first language vowels and New Zealand English ones. The speakers were found to style-shift in their production of the first and second formants of certain vowels in different settings: the German speakers were more English-target-like in the services setting and the Korean speakers were more English-target-like in the services setting and less English-target-like in the family setting compared to the university one, exhibiting a continuum of native-likeness in the three settings.

Three perception experiments complement the production analysis. Two of these focus on the effect of setting in accentedness perception and passing for a native speaker, and one explores the effect of social information (namely, ethnicity) on accentedness perception. The speakers were found to receive a different accentedness rating depending on the recording setting and whether or not the listener was aware of their ethnicity. Specifically, some speakers were rated less accented in the services setting and some in the family setting compared to the university one. Also, Asian speakers were rated similarly for accentedness both when the listeners were provided with video input and when they were not, but Caucasian speakers were rated more accented when the video input was available.

Additionally, the thesis addresses passing for a native speaker of different English varieties in an experimental context. It reveals interesting trends in the speakers' variation of passing in different settings and passing for native speakers of different varieties. The family setting was conducive to passing, and some speakers passed for a native speaker of the same

variety more often than for a native speaker of other varieties and some vice versa. Finally, the second language speakers' beliefs about passing and listeners' comments on their decision-making in identifying the origin of the speakers are investigated. The results showed that the speakers believed that first (and short) encounters with strangers were conducive to passing. A variety of linguistic and extralinguistic listener comments was revealed.

Taken together, the results paint a complex picture of variation in second language speakers' production, accentedness perception, and passing for a native speaker. The findings suggest that speakers vary in their production according to audience and in the construction of their identities. The perception experiments highlight the effect of listener expectation on their perception. These results have implications for how we understand sociolinguistic variation in second language speakers.

Vita

2009 Diplom, Theoretical and Applied Linguistics, Saratov State University, Russia

2012 M.A., English as a Second Language, University of Arizona, USA

Publications based on this thesis

See permissions from publishers in Appendix D.

Material covered in Chapter 3 has previously been published in:

Gnevsheva, Ksenia. (2015). Acoustic analysis in Accent of Non-Native English (ANNE) corpus. *International Journal of Learner Corpus Research 1*(2), 256-267.

Gnevsheva, Ksenia. (2015). Style-shifting and intra-speaker variation in the vowel production of non-native speakers of New Zealand English. *Journal of Second Language Pronunciation 1*(2), 135-156.

Material covered in Chapter 4 is under review as:

Gnevsheva, Ksenia. (under review). The audiovisual mismatch effect in accentedness perception of Asian and Caucasian non-native speakers of English.

Gnevsheva, Ksenia. (under review). The effect of recording setting in accentedness perception.

Material covered in Chapter 5 has previously been published in or under review as:

Gnevsheva, Ksenia. (2015). Passing for a native speaker: Production and perception. *EPIP 4 4th International Conference on English Pronunciation proceedings*, 37-40.

Gnevsheva, Ksenia. (accepted). Within-speaker variation in passing for a native speaker. *International Journal of Bilingualism*.

Gnevsheva, Ksenia. (under review). Passing for a native speaker: Production and perception.

Gnevsheva, Ksenia. (under review). Beyond the language: The social cues in accentedness judgments.

Chapter 1 : Introduction

'Passing' is a phenomenon of being regarded as belonging to a group which one is not a part of. It has received a reasonable amount of attention in several domains, including ethnicity, sexuality, and language (e.g., Pattinson, 2010 and references therein). It is intricately connected with recognition, the act of correctly identifying the group of origin of the actor (Williams, Garrett, & Coupland, 1999). Passing occurs when the audience fails to correctly identify the actor's social origin and takes him or her for a representative of another group. Passing, purposeful on the part of the actor or not, is a very complex phenomenon requiring high social, cultural, and sociolinguistic competence. Pattinson (2010), for example, described how British agents, attempting to pass for French nationals during World War II, had to pass linguistically, visually (physical appearance and clothing), and performatively (gait, manners). Pattinson also noted that the three domains worked together in constructing an identity and that linguistic competence did not have to be flawless for an act of passing to occur because an audience's assumptions and expectations helped to co-construct second language (L2) identities.

The work presented in this thesis focuses on phonetic variation in passing for a native speaker (NS) of a language. Most of the studies of variation in passing have been based on self-reports (e.g., Piller, 2002) while quantitative studies often background, if not completely disregard, sociolinguistic variation (e.g., Abrahamsson & Hyltenstam, 2009; see further discussion in Section 2.1). This thesis employs quantitative and qualitative methods and explores how non-native English speakers (NNESs) vary in passing for a native speaker. The main overarching question addressed here is: What is the variation in NNESs' passing for a NS?

Undoubtedly, there are many factors involved in identifying a person as a first language (L1) speaker (e.g., grammar, vocabulary, pronunciation); this thesis focuses on accentedness because being native-like in the pronunciation domain is said to be most important for a passing performance, likely because phonology typically retains non-native features as it is most susceptible to maturational constraints (Bongaerts, 1999).

An accent is a '… cumulative auditory effect of those features of pronunciation which identify where a person is from, regionally or socially' (Vishnevskaya, 2008, p. 235). Every speaker has an accent; however, lay people often believe that they do not have an accent, and only people who speak differently from them do. A difference might come in many forms, such

as when someone meets a non-native speaker (NNS) whose first language is different from their own, they may hear differences in the person's pronunciation in the second language which are a result of the speaker's L1, cumulatively perceived as a 'second language accent'. When they meet a speaker of a shared L1, they will probably regard the speaker as a native speaker of a language without any second language accent; however, if the speaker is a native speaker of a different variety of L1, they will hear 'an accent', and if they are a native speaker of the same variety of L1, they will perhaps believe that the person does not have an accent in that language.

1.1 Variation in production

Variation in passing for a NS may be the result of variation in linguistic production by the speaker, perception by the listener, or both. Linguistic variation in a second language has attracted a considerable amount of attention in the literature from both linguistic and sociolinguistic perspectives, in which the former is largely concerned with the acquisition of native speaker forms (or the acquisition of 'linguistic competence', sometimes called 'Type 1' variation; see e.g., Drummond, 2011, p. 281), and the latter focuses on the acquisition of NS patterns of linguistic variability (or the acquisition of 'sociolinguistic competence', sometimes called 'Type 2' variation).

Much previous work regards variation in foreign accentedness as Type 1 variation and, therefore, denies NNSs agency in accentedness production and ignores the sociolinguistic potential of variation in accentedness; however, it seems plausible that variation between L1 and L2 forms is a tool available to NNSs for style-shifting, as it allows a speaker to align with or distance him/herself from potential membership groups (see Dolgova Jacobsen, 2008, and Rampton, 2011). In this thesis, I view the L2 linguistic system as equivalent to the L1 system in that it may vary synchronically from one situation to another (Tarone, 1979). If NSs can be found to use 'ways of manipulating their pronunciation to clearly signal where their loyalties lie' (Gatbonton, Trofimovich, & Magid, 2005, p. 506), so might non-native speakers. That is, NNSs may not only use the sociolinguistic variation attested in the NS community, but may also use a continuum from 'native-like' to 'non-native-like' for further identity work, which we may want

to call 'Type 3' variation. In an extreme example, a NNS, able to style-shift from the far 'native-like' side of the continuum to 'non-native-like' may vary in his/her passing for a native speaker.

Gluszek and Dovidio (2010) discussed accent variation in L2 speakers in production and, although not empirically tested, they claimed that 50% of their NNS participants believed they could consciously control their accents, and 64% thought their accents changed depending on the communicative situation. A number of previous studies have empirically tested whether L2 speakers' production varies between situations (e.g., Rampton 2011; see further discussion in Section 2.2.2). This thesis aims to add to this body of research, by examining whether adult L2 speakers of English exhibit Type 3 variation (i.e., synchronic variation between 'more non-native-like' and 'more native-like') in pronunciation of vowels in different settings. I first focus on speakers' production, in Chapter 3, where I ask the following questions:

1. Do L2 speakers use differences between L1 and L2 vowel systems for situational style-shifting?
2. Does L2 speakers' style-shifting use the differences between L1 and L2 systems as a continuum as opposed to a binary choice?
3. Do speakers of different language backgrounds style-shift differently?
4. Do male and female L2 speakers style-shift differently?

1.2 Variation in perception

Additionally, this study employs a number of perception experiments which explore variation. Many perception studies have examined the factors that affect the perceived intelligibility, comprehensibility, and accentedness of foreign-accented speech. Intelligibility is an objective measure of how much of a speaker's utterance a listener understands, assessed by the number of words the listener transcribes correctly. Comprehensibility and accentedness are more subjective measures of, respectively, how easy listeners report a speaker is to understand and how 'strong' listeners rate a speaker's accent. Although different measures, they are known to be related (Munro & Derwing, 1995). I focus on variation in accentedness perception by native English speakers (NESs).

In accentedness perception tasks listeners are usually presented with utterances which they are asked to rate on an accentedness scale (e.g., from 'No foreign accent' to 'Strong foreign accent'). Presumably, listeners somehow assess the amount of deviation from their 'native speaker ideal' present in the utterance and assign it a numerical representation. This understanding is supported by Munro & Derwing (1995), who found that the majority of listeners in their experiment exhibited a significant correlation between scores of segmental and grammatical errors and intonation and their perceived accentedness scores, and Munro (1993), who found a relationship between accentedness ratings and acoustic values. If this perceived amount of deviation is negligible and the listener believes that they can perceive no foreign accent, the speaker is believed to be a native speaker of the language. If an L2 speaker is mistaken for a NS, they can be said to 'pass for a native speaker' on that occasion.

Accentedness perception has been shown to be influenced by both speaker-independent and speaker-dependent factors (e.g., Levi, Winters, & Pisoni, 2007; Lindemann & Subtirelu, 2013; see further discussion in Section 2.3.2). This thesis focuses on one of each in chapter 4: the effect of recording communication setting and speaker ethnicity. The exploration of sociolinguistic variation in different settings in accentedness perception aims to complement its study in production, and I ask:

5. Is there an effect of recording setting on perceived accentedness of a NNES?

As for speaker ethnicity, there are two main competing accounts of its effect on accentedness ratings: reverse linguistic stereotyping (i.e., assumed social information influences perceived phonetic information; Rubin, 1992) and audiovisual mismatch (i.e., the mismatch between visual and auditory information influences perceived phonetic information; McGowan, 2011; see further discussion in Section 2.3.2). The following research questions are formulated with the two accounts in mind:

6. What is the effect of availability of visual information for Asian NNESs in an accentedness perception task?

7. What is the effect of availability of visual information for Caucasian NNESs in the same accentedness perception task?

8. Will these effects for Asian and Caucasian NNESs be better predicted by reverse linguistic stereotyping or an audiovisual mismatch?

After discussing the variation in production and perception, I return to variation in passing for a native speaker and address the following specific questions exploring variation in passing for a NS:

9. What is the variation in NNESs' passing for a NS of different English dialects?
10. What are some factors that contribute to a successful passing performance?
11. What are some of the elements that listeners notice in the input when a speaker succeeds or fails at passing?

In order to address the above-mentioned questions, I have employed multiple methods, combining the quantitative methods of acoustic analysis and experimental design with qualitative methods of semi-structured interviews and content analysis. Unlike many studies of second language speakers that view variation in accentedness as a diachronic phenomenon and, as a result, rarely apply sociolinguistic methods and instead employ a battery of tests assessing ultimate attainment in a linguistics laboratory environment, I acknowledge the potential for use of within-speaker Type 3 variation for sociolinguistic positioning by NNSs and apply quantitative and qualitative sociolinguistic methods to the study of NNESs in naturalistic settings.

1.3 Thesis structure

The thesis consists of six chapters. The literature scaffolding this investigation is reviewed in Chapter 2. I start by introducing the qualitative and quantitative studies of passing and ultimate attainment. I discuss the audience design (Bell, 1984) and identity construction (Eckert, 2000) accounts of variation and review a number of studies of NNES intra-speaker variation in production and work investigating speaker- and listener-dependent factors affecting accentedness perception.

Chapter 3 focuses on the acoustic analysis of the NNESs' production of vowels and aims to address research questions 1-4 above . I analyze the style-shifting in several settings and discuss the results in light of audience design (Bell, 1984) and identity construction (Eckert, 2000) frameworks in an attempt to account for the observed variation.

In Chapter 4, I present the methods and results of three perception experiments that were designed to investigate the effect of recording setting and ethnicity on accentedness perception (addressing research questions 5-8). I discuss the results and introduce a model of accentedness perception that includes the factors studied in the experiments. The second experiment additionally explores the effect of setting on within-speaker variation in passing for a native speaker.

The topic of variation in passing for a native speaker is continued in Chapter 5 (addressing research questions 9-11). The chapter discusses the speakers' passing for native speakers of different English varieties, the speakers' beliefs about passing, and the listeners' noticing of linguistic and extralinguistic features in the speakers' speech.

In Chapter 6, I summarize and unify the production and perception results and their predictions of passing and return to the main overarching research question in this thesis. Finally, I discuss the theoretical and practical implications of this study and future avenues for further research.

Chapter 2 : Background

This chapter discusses the existing literature on variation in passing for a native speaker, NNSs' phonetic production, and accentedness perception.

2.1 Variation in passing for a native speaker

Many of the existing explorations of passing for a native speaker of a language have been based on self-reports. Marx (2002) linked the phenomenon of passing with identity negotiation in her first-person account of her experiences in a German L2 environment. The author (L1 English) claimed to first have been identified as an American when speaking German, then developing a French accent in L2 as an unconscious way to step back from her L1. With time, she became able to pass for a native speaker and started to incorporate regional 'other voices' into her L2 accent.

In another collection of self-reports, Piller (2002) interviewed NSs and NNSs of English and German and documented their experiences and attitudes towards passing for a native speaker. She recognized that passing for a native speaker was not always the ultimate goal of NNSs as simply not being immediately recognized for a NS of their L1 may be more appealing. Passing for a native speaker of a different dialect may also be preferable from some of the speakers' perspectives because, if listeners do not notice the speakers' otherness, the speakers' achievement in attaining a high level of proficiency is minimized (Piller, 2002). In other words, if an L2 speaker passes for a NS, the listeners will not be able to know and appreciate the high level of achievement in their L2 and, often, the hard work and dedication that earned it. To support this, second language acquisition research suggests that some speakers may prefer a 'neutral' accent, a hybrid containing a blend of linguistic features from several native varieties with some additions, over a standard one like General American or British English (Rindal & Piercy, 2013). In addition, Piller (2002) noticed that NNSs may aim at passing for a NS of a different dialect as an easier alternative to passing for a NS of the same dialect because listeners can be expected to be less familiar with other dialects and, therefore, more forgiving of deviations because 'dialectal influences are frequently heard as foreign and foreign influences are often heard as dialectal' (Markham, 1997 as cited in Major, 2001). Several listeners in

Hayes-Harb & Watzinger-Tharp (2012) also commented that pronunciation can be misleading in assessing someone's native-likeness because of existence of regional accents. For example, Giles (2001) offers an example of one speaker being recognized for a British speaker in California and an American in Britain; I can relate to this myself as a NNES being often considered a Canadian in the USA and an American in New Zealand.

In the aforementioned cases passing is only considered from the point of view of the speaker, and it is unclear whether the presumed acts of passing were actually successful. That is, just because a speaker believes that he or she has passed for a native speaker, that does not mean the listener perceived the speaker to actually *be* a native speaker. On the other hand, speakers may be underreporting the amount of passing they experience on a daily basis. Perceptual experiments are a way to confirm and quantify cases of passing. Some studies have used passing in perceptual experiments to explore speakers' ultimate attainment in the L2. Ioup, Boustagui, El Tigi, & Moselle (1994), for example, played clips of NSs and NNSs of Egyptian Arabic to NS listeners who were asked to indicate whether the speakers were native Egyptians (NS of the same dialect). Two of the highly proficient L2 speakers were judged as native by 8 out of 13 listeners (62%) while NSs were believed to be native by all the judges. The strict dichotomy between passing for a NS of the same dialect and failing to do so, however, misses the important middle ground where the NNSs pass for NSs of other varieties of the L2. It is possible that the remaining 38% of listeners in Ioup et al. (1994) considered the NNSs to be NSs of a different variety of Arabic.

This matter is considered in a perceptual experiment by Abrahamsson & Hyltenstam (2009), who elicited listeners' judgments along three categories: NS of Swedish from the Stockholm area (NS of the same dialect), NS of Swedish from other areas (NS of a different dialect), and NNS; nevertheless, for further analysis, the first two categories were combined, and the proportion of speakers who passed for a NS of Swedish of the same or different variety was not explored. Also, these quantitative studies used a single clip from each speaker, meaning variability in speakers' performance could not be explored.

Piller (2002) argues that passing is highly variable, and is not meant to be a sustained performance. Certain situations, Piller (2002) argues, such as short service encounters with strangers (e.g., buying a coffee in a coffee shop) and conversations with friends, can be more conducive to passing, perhaps, because there is less need to negotiate one's identity in those

contexts. Variation in passing for a native speaker from context to context is also mentioned by the participants in Magnusson & Stroud (2012). The 20 multilingual Swedish speakers, most of whom were born and raised in Sweden and therefore were not the 'typical L2 speaker', exhibited signs of non-nativeness. They claimed that sometimes they could pass for a native speaker as evidenced by a customer who was relieved to be talking to a Swede although in reality they were speaking to another immigrant, but passing did not happen all the time as is clear from one conflict customer service encounter described in the study. The participants believed that their accentedness varied: sometimes automatically (more accented when talking to other immigrant friends) and sometimes deliberately as a stylization device.

To sum up, little previous research has investigated within-speaker variation in passing quantitatively, and there are many issues left to be tackled. First, if there is intra-speaker variation in passing, recording individual speakers' ability to pass for a native speaker as a dichotomy is simplistic, and the potential situational variation in passing suggested by the self-reports in Piller (2002) has not been studied experimentally. Additionally, despite the importance of the distinction between passing for a native speaker of the same or different dialects and some NNESs' suggested preference for passing for a native speaker of a different dialect, it is unclear how often NNESs pass for a native speaker of the same or different dialects. This thesis aims to contribute to our current understanding of the phenomenon of passing for a native speaker, by investigating passing for a NS of the same or different dialects and comparing the variation in passing found in an experimental setting to speakers' self-reports.

2.2 Variation in speech production

As noted in the Introduction, it is important to remember that communication is a joint performance between a speaker and a listener; therefore, variation in passing for a native speaker may be due to variation in either speaker production, listener perception, or both. As native-likeness in the pronunciation domain has been found to be important for an overall nativeness rating (Hayes-Harb & Watzinger-Tharp, 2012), this thesis explores variation in passing for a native speaker by quantitatively looking at variation in accentedness in NNES production and perception. This section describes variation in speech production followed by a section on speech perception.

2.2.1 Accounts of sociolinguistic variation in L1 speakers

The foundations of sociolinguistics are built on observations of linguistic variation among native speakers of a given language. As early as the 1950s, Fischer (1958) showed that speakers' use of language is not always the same and differs depending on style and context. This idea was of course brought to the fore by Labov's seminal work (see e.g., Labov, 1972), which showed, amongst many other things, that a speaker's linguistic style is systematically conditioned by their social identities (e.g., social class) and by situational contexts (e.g., different levels of formality, in which more prestigious language features are used more often in formal than informal contexts). Labov (e.g., 1972) argued that this sort of stylistic variation can be modelled as attention paid to speech – formal situations encourage speakers to pay attention to language, and this increases their use of linguistic features that are viewed as prestigious by the speech community.

Other views of style, such as the frameworks of Accommodation Theory (Giles & Powesland, 1975) and Audience Design (Bell, 1984), argue that stylistic variation is conditioned by the speaker's assumptions about the listener(s) and the speaker's subconscious attempt to show more solidarity with the listener(s). One of the instances of such accommodation that Bell (1984) explores with examples from Coupland (1984) was that of a travel agent who was found to converge to clients from five different occupational classes in her production of intervocalic (t) voicing. With the setting and formality level being kept relatively constant, sociolinguistic variation on the part of the speaker was attributed to the changes in the audience (the clients) and their linguistic production.

Bell (1984) distinguishes among different types of audience who all have a different role and relationship with the speaker: the *addressee* is known (the speaker knows that the listener is present), ratified (the listener is indirectly addressed), and addressed (the listener is directly addressed); the *auditor* is known and ratified; and the *overhearer* is only known. The audience design account has also been extended as *referee* design, in an approach where the audience can be real, potential, or imagined. In this case, speakers converge to an ideal or absent audience – the *referee*. Bell (1984) cited his earlier study (Bell, 1982) as an example of referee design, where newsreaders on a high status radio station used a higher proportion of higher status linguistic variants, when compared to the same newsreaders on a lower status station. This, Bell

argued, was because speakers have their perceived audience (their *referee*) in mind when they talk. Referee design also attempts to model the effects of topic and/or setting on a speaker's style. As certain topics and settings are often associated with particular (groups of) referee(s), they can trigger style-shifting by reference to a particular audience without its actual presence. For example, a student talking about their university studies in an otherwise informal context may style-shift to a more formal linguistic production in response to an absent referee – the university community. Bell (1984) distinguished between *audience* and *referee* design as responsive and initiative styles. Audience design is responsive because linguistic production reflects a change in audience and situation, and referee design is initiative because a change in linguistic production signals a change in situation. Understanding referee design as an initiative style adds some speaker agency to style-shifting. That is, the speaker may style-shift as a consequence of redefining the relationship with the audience.

The agentive referee design approach is compatible with accounts of sociolinguistic variation in which speakers are said to actively project aspects of their identities (e.g., Eckert, 2000; see below). Style-shifting, then, can be understood as an extension of identity changes. Identity can be defined as '[a] person's place in relation to other people, a person's perspective on the rest of the world, a person's understanding of his or her value to others – all of these are integral to the individual's experience of self, and are constructed in collaboration with others as those others engage in the same construction of themselves' (Eckert, 2000, p. 41). The current sociolinguistic understanding of identity assumes it to be emergent, not pre-existing; composed in accordance with macro-social categories and situational positioning; indexically, relationally, and partially constructed (Bucholtz & Hall, 2005). This means that a person does not choose among a pre-defined set of rigid options but constructs an identity that is potentially unique to a given situation with its audience and topic, among other things.

Linking production models with sociolinguistic variation research is important for our understanding of language variation. This thesis is informed by usage-based models, which offer one way to explain how social information might have an effect on linguistic production. Exemplar theory (Pierrehumbert, 2003), for example, suggests that our brain stores a cloud of representations, exemplars, for a given phoneme; this cloud is updated constantly through the perception-production loop. Sociolinguistic information, such as different speaker characteristics (sex, age, origin, etc.) and contextual information, can be attached to and stored together with

these exemplars and activated when certain exemplars are activated as well as activate certain exemplars when it is accessed (Hay, Nolan, & Drager, 2006).

Most sociolinguistic variation research has been conducted on monolingual speakers, but there has also been research on bidialectal individuals, that lie on the continuum between monolingual and bilingual speakers. Walker (2014) studied topic-effected variation in production of rhoticity, intervocalic /t/, and BATH[1] in bidialectal speakers of British and American English. She found a different degree of the effect of topic for all speaker groups suggesting that bidialectal speakers may use cues available to them through the two dialects for style-shifting. These findings can be extended to bilingual individuals, and it can be hypothesized that they, too, use variants in the two languages for style-shifting.

2.2.2 Accounts of sociolinguistic variation in L2 speakers

While the work introduced above has focused for the most part on monolingual native speakers of a language, sociolinguistic variation in L2 varieties has also received some attention. Studies of L2 sociolinguistic competence have often shown that non-native speakers adopt variable linguistic patterns used by first language speakers (Adamson & Regan, 1991; Li, 2010; Major, 2004; Regan, 1996; Rehner, Mougeon, & Nadasdi, 2003; Schleef, Meyerhoff, & Clark, 2011). Adamson and Regan (1991), for example, studied the use of the variable (ing) by NNSs of American English and found that female speakers used [iŋ] more frequently than males and even more so in monitored speech, which is similar to the NS pattern. Major (2004) also investigated gender and style and argued that NNESs acquired gender differences more or faster than stylistic differences. Interestingly, Schleef et al. (2011) compared the variation in the use of the variable (ing) in the speech of Polish and locally-born adolescents in the UK and found that migrant teenagers not only adopted some target-like linguistic and social constraints but also introduced novel ones. This suggests that NNSs may not be limited by the sociolinguistic variation present in the L1 community.

[1] I use Wells' (1982) lexical sets for the remainder of the manuscript to represent the target vowel intended by the speaker. Lexical sets, signalled with uppercase letters, are keywords which represent vowel phonemes. For example, the GOOSE lexical set represents the /u:/ vowel, in words like *food*, *hoot* and, of course, *goose*.

Several recent studies have also considered the sociolinguistic potential of non-native-likeness, and identity projection, for example, has been used to explain variation in accentedness in a number of cases. For example, Marx (2002) argued that non-native speakers can (sub-) consciously choose to preserve their foreign/regional accent as a way to signal aspects of their identity changes. She provided a detailed first-person account of how her own L1 (English) accent in L2 German changed diachronically, and recounted her experience of first being identified as an American when speaking L2, then sometimes passing for a native speaker. On her return to an English dominant culture, she reported she had developed an accent in her L1, which had adopted some German influences as a result of L2 transfer. Marx (2002) hypothesized that accent is not always a sign of fossilization and inability to reach native-like proficiency but rather a sign of identity negotiation. A similar example comes from Sancier & Fowler (1997) who examined voice onset time (VOT) in plosives produced by a Brazilian Portuguese L1 speaker of L2 English. Typically, VOT has a longer duration in English than in Brazilian Portuguese, and Sancier & Fowler (1997) reported that after a several months' stay in either Brazil or the USA, the speaker's VOTs in either language shifted in the direction of the language she was most recently exposed to. Marx (2002) and Sancier & Fowler (1997) showed the shifting relationship over time between a speaker's L1 and L2, which could be an inevitable result of long-term contact, where exposure to a given language for a given period of time affects a speaker's production in both of the languages they speak.

Rather than being the result of long-term contact, other studies have argued that synchronic variation in an L2 is the result of style-shifting. In line with Labov's early model of stylistic variation, which equated style with attention to speech because of shifts in formality, Beebe (1980) showed differences in the native-like pronunciation of English /r/ for Thai speakers, with more native-like production in a more formal style (list reading) as opposed to a less formal one (interview) with an interaction with the position of /r/ (in initial /r/, the pattern is reversed). Major (2001) also predicted a more native-like L2 production in formal styles; however, the relationship is more complex for L2 speakers, as this variation in L2 may contrast with variation in L1 resulting in less attention and more accurate production. For example, in informal Japanese some vowel reduction approximates consonant clusters in English (positive L1 transfer) while in formal style the absence of vowel reduction will make consonant clusters more non-native.

Convergence to the interlocutor in L2 speech has been studied by several researchers. Beebe and Zuengler (1983) applied Accommodation theory to two studies of L2 speaker variation and found that bilingual speakers accommodated in syntax and pronunciation to their interlocutors who were different ethnically (bilingual Hispanics in the US and Thai-Chinese) when the topic and setting were controlled for. They argued that interference from the L1 is not the only source of style-shifting for L2 speakers. More recently Zając (2015) discussed the complex patterns of convergence, divergence, and maintenance in the speech of Polish L1 learners of English and found effects of attitudinal factors such that the majority of speakers were found to converge to native English speakers (and not to NNESs) because, she argued, of their preference for target-like pronunciation.

Identity has also been used to explain variation in 'accentedness', meaning divergence from native-like pronunciation, in a number of cases. Dolgova Jacobsen (2008), for example, found a correlation between speech accentedness and identity, operationalized as self-identification with the L1 community (Russian) or the ethnic group in the US (Russian-American). Speakers differed in the amount of target-like pronunciation of the lax [ɪ] depending on the topic in the interview (among other factors, such as self-identification, phonological environment, and stress) with no change of audience.

The studies discussed above linked L2 variation with audience and topic by manipulating one of these variables and keeping the others constant. However, situational style-shifting is multi-faceted and involves a change in topic and audience. Piller (2002) discussed L2 speakers' ability to use native-like linguistic features and 'pass' for a native speaker in some situations. For example, short service encounters (at the post-office, for instance) may be conducive to 'passing' because in brief communication with a stranger, identity is arguably less likely to come to the forefront and being 'native-like' would not be regarded as deceit or an attempt to forge a fake L2 identity. Another situation where speakers believed 'passing' occurred was in communication with friends when the speaker did not feel the need to negotiate their identity. However, Piller's (2002) findings are based on self-reports, and the exact nature and extent of style-shifting is therefore unclear. Rampton (2011) is a more quantitative auditory study of style-shifting, stylization, and register in one adult L2 English speaker of L1 Punjabi - *Mandeep*. Although the tokens of interest are rather few, Rampton focuses on the realizations of [t], dark [l] and diphthongs [ei] and [ou] in several communicative situations: a dialogue with a Punjabi friend at

home, an interview with an Indian interlocutor, and communication with English colleagues at work. Mandeep demonstrated style-shifting with these four contexts: his speech was most 'Anglo' at work and most Punjabi when talking to a Punjabi friend at home. This offers an important foundation for the present study but should be treated cautiously as these findings are based on one male speaker of L1 Punjabi.

Moreover, most of the studies discussed so far did not consider the effect of speaker sex on style-shifting. However, many previous studies of L1 (e.g., Eckert, 2000) and L2 speakers (e.g., Drummond 2011) suggest that males and females often use language differently. The effect of speaker sex on style-shifting was explored in a study of the use of ethnically-marked variants by British-born Asians (Sharma, 2011). Younger female speakers showed more variation and were found to use more Punjabi variants in the family setting compared to males speakers, supporting Eckert's (2000) claim that women exhibit a wider stylistic range than men. Sharma explains this finding by linking the width of ethnolinguistic repertoire with the diversity of social networks. This study investigated style-shifting by male and female second generation immigrants, and it is unclear whether a similar trend will emerge for second language speakers; therefore, this matter is taken up in this thesis.

Some of the work discussed so far suggests that there may be a continuum between native-likeness and non-native-likeness. Both audience design and identity construction accounts may predict different degrees in (non-) native-likeness and in each approach variability is understood as a continuum as opposed to a choice between two extremes. Identity is not a set of options available to a person but a synthesis (Block, 2006), so we would not expect an L2 speaker to only switch between two options of 'L1-ness' on the one hand and 'L2-ness' on the other. There are very likely intermediate forms. Bell (1984) can also be interpreted in a way that predicts this gradation through the different types of audience and audience compositions: addressee, auditor, overhearer, and referee. Similarly, Rampton (2011) showed that Mandeep was most English-like in the 'most English' situation – that is, with English audience who he was used to interact with in English on work-related topics. He was also most Punjabi-like in the most Punjabi situation with a Punjabi audience, which is more likely to be associated with L1 and L1-related topics. The gradation of English-ness/Punjabi-ness suggests that bilingual speakers may exhibit a continuum of target-likeness from a situation with L1 interlocutors interacting on an L1-related topic to an L2-related topic discussed with L2 interlocutors.

However, in cases of situational style-shifting it may be difficult to disentangle the effects of audience and topic as Mandeep's more L2-like pronunciation in the work setting could be the result of either audience or topic separately or a combination of both. An experimental design which examines both audience and topic (see Section 3.1) will help to tease issues like these apart.

To sum up, previous studies have found that audience, topic, and setting may have an effect on style-shifting in NNESs. Studies of situational style-shifting allow us to manipulate more than one variable and examine the cumulative effect and relationship between them. However, existing studies of within-speaker variation have either been based on self-reports or on a single speaker. This has not allowed us to explore differences between males and females and people from different L1 backgrounds with the same methodological tools. The aim of this thesis is to use larger numbers of male and female participants from different L1 backgrounds to help further our understanding of intra-speaker variation in NNESs. I explore style-shifting in Chapter 3.

2.3 Variation in speech perception

2.3.1 The inter-relationship between linguistic and social information

There is a growing body of research on sociophonetic variation in speech perception (see Drager, 2010 for a review). Many studies have shown that the way a person speaks affects listeners' perception of the speaker in terms of a range of social categories, in a form of linguistic stereotyping. For example, Campbell-Kibler (2007) found that two speech samples that differed only in the speaker's production of the (ING) variable were associated with different social categories: *–in* was associated more with lack of education, masculinity, and the country, while *–ing* was perceived to be more educated, gay, and urban.

Reverse linguistic stereotyping has also been attested: perceived phonetic information has been found to be influenced by (assumed) social information, such as geographical region (Hay, Nolan, et al. 2006; Niedzielski, 1999), and the socio-economic status and age of the speaker (Hay, Warren, & Drager, 2006). Niedzielski (1999) found that the information the listeners were given about the origin of the speaker influenced their responses in a perception task. If listeners

were told that a speaker was Canadian, they chose raised-diphthong tokens as best representative of the vowels produced by the speaker in the clip, not because the tokens actually matched the speaker's vowel production, but because those tokens matched most closely with the listeners' expectations of Canadian speech. Hay, Nolan et al. (2006) found a similar effect of mentioning a geographical region with a population of listeners from New Zealand (NZ). Two groups of listeners were asked to choose a synthesized vowel which was most similar to that of the speaker's actual production, and mark it on an answer-sheet which had either 'Australian' or 'New Zealander' written at the top. All listeners heard the same speaker of New Zealand English (NZE) but chose synthesized vowels which were more similar to Australian English if their answer sheet had 'Australian' at the top. Hay & Drager (2010) found the same effect when no region was explicitly mentioned but the listeners were shown stuffed toy kangaroos or koalas, associated with Australia, or stuffed toy kiwis, associated with New Zealand. They argued that once a region is primed, it can have a perceptual effect in the listening task. Similar effects have been found with other social factors, such as socioeconomic class and age. Hay, Warren et al. (2006) manipulated the perceived social class and age of the speakers in a vowel identification task and presented listeners with audio input containing /iə/ and /eə/, which are merged for some speakers of NZE. They found a connection between the assumed social characteristics of the speaker and listener accuracy at identifying the produced vowel.

Hay and colleagues explain their findings with usage-based models of speech perception. Hay, Warren et al. (2006) suggest a relationship between identification accuracy and the difference between expected and actual production. When both the linguistic and sociolinguistic information is available and is congruent, this may facilitate access through more focused activation of representations, resulting in fewest identification errors. An incongruence between the actual production and the expected production, which comes to be expected because of what the listeners are told about the speakers, may lead to higher error rates as the mismatch between the perceived phonetic and social information will result in a more spread-out activation of representations and may inhibit access. One can hypothesize that activation of experience-based representations with conflicting phonetic or social information at the same time may influence a listener's ratings. A usage-based account is potentially insightful when considering studies of L2 variation, including foreign-accentedness rating tasks. Usage-based models would predict that in a foreign accentedness rating task the items that activated the representations most similar to the

ones associated with the listener, if they are a native speaker, would be judged as 'less foreign accented' whereas the items different from them would be judged as 'more foreign accented'. Additionally, for both a native and a non-native listener, representations similar to the ones that have previously been identified as foreign-accented, would be judged as 'more foreign accented' and vice versa. In the next section, I review existing work on foreign accentedness perception and its relationship to social information.

2.3.2 Perceived foreign accentedness

Speaker-independent factors

Accentedness perception is highly variable and is known to be influenced by a number of speaker-dependent and speaker-independent factors. Flege & Fletcher (1992) have found that listeners rated speakers as more foreign accented after they became familiar with the produced sentences. Moreover, orthographical presentation of the stimuli at the same time with the audio may have the same effect, as words presented with their orthography were perceived to be significantly more accented compared to words presented via audio input only (Levi et al, 2007). The same study found an effect of lexical frequency such that words of higher frequency (three groups of frequency) were perceived to be significantly less accented with no significant effect of lexical frequency found in production of the first and the second formants of monophthongal vowels. Familiarity with foreign accents was found to be a significant predictor as listeners who rarely interacted with non-native English speakers perceived a stronger foreign accent (Kraut & Wulff, 2013). Musical ability was explored as one of the factors by Isaacs & Trofimovich (2011) who found that non-music majors assigned higher accentedness scores than music majors. Listener-dependent factors have been found to affect not only accentedness ratings but comprehension of foreign accented speech. For example, teenagers were found to perform better than younger kids, and although not tested statistically, adults were informally assessed to perform better than teenagers (Munro, Derwing, & Holtby, 2011). So, argue Levi et al. (2007), '[a] speaker may therefore only have an "accent" within a specific perceptual framework and listening context' (p. 2337).

Piller (2002) claimed that certain settings (e.g., short service encounters) are conducive to passing for a native speaker. It is possible that such variation in passing is due to variation in

perception of accentedness, and certain settings, or topics, make accentedness more noticeable. Rubin & Smith (1990) explored the interplay between accent, ethnicity, and lecture topic. The same Chinese-accented speakers delivered a 'science' and a 'humanities' lecture. They were perceived to be more 'oriental' in the humanities subject, but no significant effect of topic on perceived accentedness was found. One methodological criticism that could potentially account for this finding is that the humanities lecture was about an Indian classic tale, the Mahabarata, a clearly 'oriental' subject, and the effect of a 'culturally neutral' topic remains unknown. The lexical frequency effect found by (Levi et al., 2007) potentially predicts a higher accentedness rating for the 'science' lecture which contained much technical vocabulary, such as 'growing scarcity of helium supplies' (Rubin & Smith, 1990, p. 342). In this study the audio stimuli were always presented with a Caucasian or an Asian picture, so an effect of ethnicity might have overpowered a potential effect of topic. This line of research will benefit from an experiment studying the effect of setting or topic on perceived accentedness. In chapter 4, I explore the potential effect of setting on accentedness perception in an experimental context (see Sections 4.1 and 4.2).

Speaker-dependent factors: Ethnicity

There are also a number of linguistic and extra-linguistic speaker-dependent factors that influence accentedness perception. Obviously, speakers of different proficiencies may be perceived to have a different degree of accentedness. A speaker may be judged non-native due to several kinds of linguistic differences from native speakers of that language: deviations in grammar, collocation usage, and non-native-like phonology among others (stimulus factors; see Section 2.3.3). However, non-auditory cues pertaining to the speaker may also affect accentedness ratings. For example, articulatory differences in non-native speakers, which result in divergent pronunciation, may be perceived visually and may be a cue in their own right. Some deaf-mute people claim that they are able to perceive a NNS accent by lip-reading (L. Kenn, personal communication, May 14, 2011). Aside from that, extra-linguistic factors, such as physical appearance, clothing, behavior, etc., may influence accentedness perception. For example, Marx (2002) changed her clothing style in order to blend in with the L2 community. Furthermore, Lantolf & Thorne (2006) have found differences in the use of gestures by first

language and second language speakers in terms of gesture amount and type. Additionally, Kraut & Wulff (2013) found that perceived degree of accentedness entered into interaction with sex with female low and intermediate proficiency NNESs receiving lower scores than their male counterparts. Finally, a speaker's ethnicity may be one of the extra-linguistic factors influencing his/her perceived accentedness. Anecdotes of native English speakers of a non-white background being perceived to have an accent are abundant: in Lippi-Green (1997) a monolingual English-speaking woman of Asian Indian decent was asked by a shopkeeper to speak slower because of her 'accent'.

A number of studies have explored the way assumed ethnicity of the speaker influences his/her perceived accentedness and intelligibility. For example, listeners' assumptions of the speaker's ethnicity based on the presented proper name, supposedly representing the speaker, have been found to influence ratings of degree of foreign accent in accentedness perception studies. In Prikhodkine (2012) Swiss listeners were presented with clips recorded by native speakers of French with either majority or minority (Portuguese and Arabic) proper names. The results, which are compatible with the work in the sociophonetic literature discussed in Section 2.3.1, suggest that allusion to a different ethnicity or a potential non-native status of the speaker may have an effect on perceived accentedness and employability ratings.

Reverse linguistic stereotyping based on ethnicity has also been explored by studies using visual stimuli, such as pictures of people of different ethnicities, to represent the speaker in accentedness rating tasks. In Rubin (1992) the same audio-recording of a native speaker of Standard American English (SAE) was presented to students in a class with two different pictures supposedly representing the speaker: a Caucasian and an Asian woman. The students who were presented with a picture of an Asian woman rated the recording as more accented because, Rubin (1992) argues, they *expected* it to be accented. Moreover, comprehension scores of listeners presented with an Asian picture were lower than of those presented with a Caucasian picture. This effect supports the negative bias hypothesis; that is, listeners' negative bias towards Asian faces was said to influence their accentedness rating even when presented with audio stimuli from a native speaker with a SAE accent. This is a persuasive example of what the effect of visual stimuli might be on the perception of native-like linguistic input. However, it remains unclear what the accentedness ratings would have been if the listeners had been presented with Asian and Caucasian faces matched with accented speech rather than SAE.

In an experiment involving foreign- and standard- accented speech, Yi, Phelps, Smiljanic, & Chandrasekaran (2013) collected native English speaker listeners' intelligibility and perceived accentedness ratings of native speakers of SAE and non-native English speakers of Korean L1 in audio only and audiovisual conditions. In the intelligibility experiment, word recognition in noise was better for NESs than NNESs and better in the audiovisual than visual condition; there was also a significant interaction such that the audiovisual benefit was larger for NESs than for NNESs. In the accentedness rating experiment, six NES listeners were presented randomly with and rated on a 9-point Likert scale 40 target sentences, each spoken by four speakers (two NESs + two NNESs) in two conditions (audio and audiovisual), resulting in a total of 320 presentations. In line with predictions of the negative bias hypothesis, the authors argue, the Korean speakers were rated significantly more accented in the audiovisual condition than in the audio only condition, exhibiting an effect of ethnicity. However, it could be that the experiment design may have had an impact on the obtained results because, besides the small number of listener and speaker participants, the use of audio/audiovisual condition as a within-listener factor may have prompted the participants to notice the importance of the visual cue. This interpretation is supported by Yi, Smiljanic, & Chandrasekaran's (2014) finding of a null condition effect in a clarity-rating task with the same stimuli from the Yi et al. (2013) study. Following a similar method as Yi et al. (2013), Yi et al. (2014) found neither a significant effect of condition nor an interaction between condition and speaker group suggesting that listeners found Korean speakers equally comprehensible in both audio only and audiovisual conditions.

McGowan (2011) explored intelligibility and perceived accentedness in SAE and foreign-accented speech. In the intellibility experiment, listeners were presented with foreign-accented speech together with an Asian or a Caucasian photograph or a silhouette. The listeners, who had a task of transcribing Chinese-accented speech, were found to be significantly more accurate when presented with an Asian photograph than a Caucasian or a neutral face (a silhouette), possibly due to a 'mismatch-induced inhibition' in the latter. Psychology literature has well-documented cases of facilitation and interference effects associated with audiovisual integration (Campanella & Belin, 2007 and references therein). Previous studies have found improved intelligibility when congruent visual input is provided. Incongruence between the two lines of input may result in interference as in the McGurk effect in which the production of one phoneme and the visual presentation of a different one resulted in a perceptual illusion of a third phoneme

(McGurk & Macdonald, 1976). According to usage-based models of speech perception, Chinese-accented speech and an Asian picture together would activate a more focused set of experience-based representations enhancing intelligibility while an incongruence between the audio and visual input would result in a mismatch of expectations and spread activation more thinly inhibiting intelligibility. Although this was not an accentedness rating experiment, McGowan (2011) argues against the negative bias hypothesis as he found that socioindexical cues enhanced perception.

It should be clear from the discussion above that most studies of the effect of ethnicity on foreign accentedness perception have looked at Asian speakers, leaving Caucasian non-native speakers an under-studied group. However, in the absence of a negative bias, the use of Caucasian speaker participants allows for the testing of other effects of ethnicity, such as an 'audiovisual mismatch' effect. In an accentedness rating task in which listeners are presented with foreign accented speech either by itself or together with an Asian or a Caucasian face, reverse linguistic stereotyping may predict a lower foreign accentedness rating for Caucasian NNESs when the face of the speaker is presented compared to when it is not in the absence of a negative bias, and a higher foreign accentedness rating for Asian NNESs in the presence of a negative bias. On the other hand, an audiovisual mismatch effect would predict a similar accentedness score for foreign-accented speech presented by itself or with an Asian face and a higher accentedness score for speakers when a Caucasian face is shown. In Chapter 4, section 4.4, I explore the effect of ethnicity for Asian and Caucasian NNESs in an accentedness rating task with the aim of testing the predictions of the reverse linguistic stereotyping and audiovisual mismatch accounts.

2.3.3 Listener cues

Linguistic features

There are many cues that signal a NNS of a language, including a foreign accent, non-target-like grammar and vocabulary. Several linguistic studies have argued that native-likeness in the pronunciation domain is most important for being judged a native speaker of a language. When asked which of the 6 areas of language are most important for being rated native-like (from contextual appropriateness, fluency, morphology, pronunciation, sentence-level grammar, and

vocabulary) half of the 12 listeners in Hayes-Harb & Watzinger-Tharp (2012) rated pronunciation as the most important, while 3 more rated it as 2nd or 3rd most important, though several participants believed that pronunciation is not the most important factor because of the existence of regional accents.

Generally people are very good at noticing a foreign accent from phrases, words, and even phones or in content-masked speech (Munro, 2008; Munro, Derwing, & Burgess, 2010), and Hayes-Harb & Hacking (2015) note that there is generally considerable agreement on accentedness ratings among listeners in perception tasks. Previous research has argued that accentedness ratings are connected with deviations from some notion of a 'standard', relying on the Standard Language Ideology which posits, among other things, that a standard language is 'correct' and internally consistent while deviations from it are 'incorrect' (Lippi-Green, 1997). Munro & Derwing (1995) and Munro (1993) found a significant correlation between segmental and suprasegmental errors in the speech and the speakers' perceived accentedness. A number of studies found that prosody was more strongly correlated with accentedness / intelligibility than segmental structure (Anderson-Hsieh, Johnson, & Koehler, 1992; Childs, 2012; de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012 and references therein).

Several studies calculated the proportions of different cues listeners believed they employed in accentedness and nativeness judgment tasks, and phonological and segmental features in particular seemed to dominate listener judgments. The raters in Derwing & Munro (1997), for example, mentioned segmentals (92%), grammar (46%), enunciation (38%), prosodic features (23%), rate (15%), fluency (8%), and vocabulary (8%). Moyer (2004) also found that raters in a nativeness judgment task believed that they relied mostly (79%) on phonological factors (specific segments (27%), foreign accent (26%), intonation (11%), speed/tempo (7%), syllable stress (5%), hesitation and rhythm (3%)) leaving a modest 21% for non-phonological factors like lexicon/word choice (13%), morphology (5%), and syntax/word order (3%). Ioup et al. (1994) briefly discussed that their raters mentioned one speaker's non-native pronunciation of segments or non-target-like intonation and the other speaker's general accentedness or vowel quality. The authors themselves noticed one of the speaker's native-like use of discourse markers and pause fillers; however, they did not report commentary on the use of these features by the raters themselves.

Munro & Derwing (2015) note that acoustic measures do not always correlate with perception ratings, and sometimes a second language speaker may be believed to be a native speaker by listeners despite some non-target-like production when assessed objectively, a phenomenon Abrahamsson & Hyltenstam (2008) called 'non-perceivable non-nativeness'. Such cases of non-perceivable non-nativeness suggest that the second language speaker's production of all segmental and suprasegmenal features need not be in the native speaker range for passing to occur. Schmid, Gilbers, & Nota (2014) collected L2 speakers' VOTs, vowel formant measurements for /ɜ/ and /æ/, and their ratings on a foreign accentedness scale. Only one L2 speaker out of 20 fell within the range of native speakers in perceived accentedness, but this speaker received a perfect score, meaning that he was perceived to have no foreign accent by any of the judges, and, therefore, passed for a native speaker. The authors scrutinized the speakers as a group and individually and concluded that the two L2 speakers who scored within the NS range on production measures were not the ones judged to be most native-like in perception. On the other hand, the L2 speaker who passed for a NS produced some VOTs and /æ/ which were outside the native range. This may suggest that a certain degree of non-nativeness or non-target production of some (possibly, less salient) elements may still be below the non-nativeness threshold in perception. In fact, the findings of a perceptual dialectology study by Watson, Leach, and Gnevsheva (submitted) suggest that listeners' ability to correctly identify L1 varieties may depend on the presence or absence of (salient) features in the stimulus.

One way to explore the salience of individual features and their effect on passing or perceived accentedness is through acoustic manipulations of second language speech. Magen (1998), for example, found that in their ratings of foreign accent in non-native English speakers of L1 Spanish, listeners were sensitive to such features as vowel quality, consonant manner, and stress, but not to voicing.

An alternative is to collect listener comments and focus on what non-target-like elements they notice in L2 speech. Previous studies have found that listeners often comment on phonological features in general (Derwing & Munro, 1997; Moyer, 2004; see above), but individual segments or suprasegmentals are rarely mentioned. Qualitative studies sometimes discuss individual segments that listeners comment on and which are, therefore, noticeable or salient. For example, in McKenzie (2015), listeners explicitly commented on many non-target-

like pronunciations of consonants: /v/, /l/, /r/, /t/, /d/, /ð/, etc., but non-native-likeness in vowels was only mentioned in general.

Hayes-Harb & Hacking (2015, p. 54) also noted that, while the listeners commented a lot on consonants and even mentioned specific segments (‘ “th” sounds like “d”, sometimes like “t”’), their comments about vowels were more general (‘sounded different’, ‘foreign’, and ‘not English’). To address individual vowels, raters in Hayes-Harb & Hacking (2015, p. 55) often used imitation by providing examples from the NNESs’ speech (e.g., ‘call sounded more like “c[o]ll” ’). Such imitation of speaker features may be reflective of the non-linguist listeners’ lack of terms for description but, at the same time, also reflective of a certain degree of awareness of noticeable differences (Preston, 1996), which suggests that listener imitation of vowels can be used for analyzing salient features.

Extralinguistic features

The above-mentioned features are all linguistic ones, and they are clearly important when identifying speakers from audio data, but linguistic research in other domains has also highlighted the relationship between linguistic and social information in speech perception (see Sections 2.3.1 and 2.3.2). This has been argued to indicate that listeners rely on social information in such linguistic perception tasks as accentedness rating or dialect recognition. In fact, Williams, Garrett, & Coupland (1999) argued that dialect identification is inseparable from affective and evaluative processes in listeners’ perception, and attitudinal comments often appear in ‘draw a map’ tasks, typical of perceptual dialectology work, where participants are asked to outline and label dialect regions (for example, the South of the USA is thought to be courteous and hospitable; Garrett, 2010).

Related to this, many studies have identified sociolinguistic features alongside linguistic ones in listeners’ comments about speakers (e.g., Hayes-Harb & Hacking, 2015; McKenzie, 2015). In Hayes-Harb & Hacking (2015), 10 listeners rated the same reading passage produced by 10 native and non-native speakers of English on an accentedness scale and were interviewed about the features that influenced their ratings. The listener comments were divided into several linguistic and extralinguistic categories: segments, rhythm, speaking rate, intonation, speaker models, and task effects. The authors noted that many linguistic comments were expected, but

some, pertaining to the socio-cultural background of the speakers, suggested that listeners did not limit themselves to linguistic factors in the task and employed models of speakers inclusive of their sociolinguistic knowledge, which, Hayes-Harb & Hacking (2015) argued, reflected the listeners' tendency to stereotype speech and groups of speakers. The raters often compared the speakers to themselves and 'ideal' (e.g., 'pure American') or stereotypical (e.g., 'accent makes him sound Eastern European') speakers of certain varieties. They were also found to create more detailed portraits of the speakers and refer to social groups as models (e.g., 'Caucasian highschool girls'). The listeners also took the nature of the task into consideration when making their ratings (e.g., 'when I read, I'm sometimes slower, too'). Although this study employed read recordings of the same paragraph, which made the rating task less naturalistic and limited the speaker opportunities for self-expression, the listener reliance on social attitudes in addition to the purely bottom-up analysis still became apparent.

In McKenzie (2015), UK-born listeners were presented with clips produced by speakers of different native and non-native varieties of English and were asked to identify the speaker origin and comment on their decision process. The listeners were found to comment on linguistic features (e.g., 'the TH sound') and several extra-linguistic features (e.g., confidence) as well as speaker models (e.g., 'sounds like HSBC call centre'). Both McKenzie (2015) and Hayes-Harb & Hacking (2015) suggest that despite our heavy reliance on perception tasks, we do not fully understand what linguistic and extra-linguistic cues may underlie listener judgments, how listeners assign social information to perceived speech, or what raters think they notice, because listener qualitative comments are rarely scrutinized.

The analysis presented in Chapter 5 addresses feature salience in passing for a native speaker through a detailed discussion of three L2 speakers' monophthongal vowel production, native listeners' judgments of the speakers' origin in a perception experiment, and qualitative comments on their speech, specifically imitations (examples) of native-like and non-native-like elements. It also expands our understanding of listener reliance on extra-linguistic cues in perception tasks. It is not, however, limited to the construct of accentedness and includes origin / native-likeness judgments as well which allows us to tap deeper into the listeners' sociolinguistic awareness. The use of spontaneous speech in this thesis (1) minimizes the task effects mentioned by listeners in Hayes-Harb & Hacking (2015), (2) allows the speakers to choose their own words, giving them an opportunity to avoid the words whose pronunciation they are not sure

about, making it a more naturalistic task, and finally (3) allows to present different content to raters, again making it a more naturalistic task and avoiding familiarization effects (Flege & Fletcher, 1992).

2.4 Summary

To sum up, this thesis aims to contribute to our understanding of sociolinguistic variation in NNESs. Unlike most studies of ultimate attainment (e.g., Ioup et al. 1992), it considers passing for a native speaker as a variable phenomenon, compares the experimental results to speaker self-reports, and particularly tests Piller's (2002) claims that specific settings are conducive to passing. It also makes a distinction between passing for a native speaker of the same dialect and a different dialect as listener. Another contribution of this thesis is that, unlike most research on passing for a native speaker, it applies quantitative methods to variation in passing.

As variation in passing for a native speaker can be the result of variation in production, perception, or both, the thesis additionally explores sociolinguistic variation in NNESs' production and perception with the focus on pronunciation. Specifically, it employs acoustic analysis of vowels in the investigation of L2 speaker variation on the L1-L2 continuum with the aim of extending our understanding of situational style-shifting in NNESs and adding to the scarce existing research on the matter (e.g., Rampton, 2011). The analysis of speaker variation between settings is complimented by an analysis of variation in accentedness perception between settings. Additionally, the effect of speaker ethnicity on accentedness perception of foreign-accented speech is investigated in order to test the predictions of reverse linguistic stereotyping (Rubin, 1992) and an audiovisual mismatch (McGowan, 2015). In order to add to our understanding of listener reliance on linguistic and extra-linguistic cues (which are only rarely reported on in perception tasks, e.g., Hayes-Harb & Hacking, 2015), the specific listener comments pertaining to speaker linguistic and socio-cultural features are explored.

Chapter 3 : Variation in Speech Production

Previous studies of sociolinguistic variation in L2 speakers have found an effect of audience, topic, and setting; however, studies of situational style-shifting have typically been based on self-reports (Piller, 2002) or a single speaker (Rampton, 2011). Additionally, they rarely employed the same methodological tools, and comparing the behavior of males and females and different L1 groups was not possible. Building on the work discussed in Section 2.2, the research questions that this chapter aims to address are:

- Do L2 speakers use differences between L1 and L2 vowel systems for situational style-shifting?
- Does L2 speakers' style-shifting use the differences between L1 and L2 systems as a continuum as opposed to a binary choice?
- Do speakers of different language backgrounds style-shift differently?
- Do male and female L2 speakers style-shift differently?

To address these questions, I analyze the production of a set of vowels by two groups of NNESs (L1 Korean and L1 German) in three different communicative situations (two interviews and one service encounter) in New Zealand. I focus in the main on the first and/or second vowel formants (F1 and F2 respectively) for KIT (F1 and F2), DRESS (F1 and F2), TRAP (F1), and GOOSE (F2) for both language groups, and additionally STRUT (F2) for Korean and FOOT (F2) for German L1 speakers (see below).

3.1 Methodology

3.1.1 Participants

The speakers in this thesis (see Table 3.1 for details) were 18 highly proficient but non-native speakers of English (9 L1 Korean (K) and 9 L1 German (G)) and 6 L1 speakers of English (2 NZE, 2 SAE, and 2 Standard British English (SBE)). They were recruited in my social circles and through the 'friend of a friend' method (Milroy, 1987). The age, education, socio-economic

class of the participants were comparable to those of the investigator and listeners in the perception experiments, to be discussed in the next chapter, which used recordings from these speakers. The age range of the speakers was 21-34; average age = 25; all were affiliated with the same university in New Zealand at the time of the study (highest academic degree achieved or in progress: 8 Bachelor's, 4 Master's, and 12 PhD). Half of the participants were males and half females. Six participants in each L2 English group (3 males and 3 females) were informally categorized by me to be of higher English proficiency, and three were categorized as lower English proficiency. Under informal observation higher proficiency speakers exhibited greater fluency, fewer phonological mistakes, and native-like or near-native grammar. Indeed, I thought several of them to be a NES in the first few minutes of the first meeting with them. This division was subsequently supported by a significant difference in accentedness rating of the two proficiency groups found in the perception experiments (see Chapter 4). The higher proficiency L2 speakers' production was collected for both production and perception analyses while the lower proficiency speakers were recorded for the purpose of creating a range of proficiencies in the accentedness rating experiments only (see Chapter 4). Therefore, only the production data of the higher proficiency NNESs are discussed further in this chapter.

Table 3.1: Speaker biographical information (Speakers of higher proficiency, whose production is analyzed in this chapter, are highlighted)

Name[2]	Recording date	Age	Sex	Degree	L1	NZ arrival	First exposure to English (age in years, setting)	English-speaking countries visited, NZ and country of birth excluding (name, duration of stay, age in years)
Al	June 2013	23	M	BSc	SBE	Aug. 2012	N/A	
Amy	Sep. 2013	22	F	BA	SBE	2009	N/A	USA, 7 mnths, as adult
BrandiAT Muhkuh	May 2013	26	M	PhD	G	Apr. 2013	elementary school	USA, several weeks, as adult
Dakota	May 2013	26	F	PhD	SAE	Apr.	N/A	Australia, 6 mnths, as

[2] All names of speakers are self-chosen pseudonyms.

						2013		adult Scotland, 1 yr, as adult
Emily	Aug. 2013	21	F	BSc	K	2004	10 yrs, tutor	Australia, 1 mnth, at 11
Gabriella	Nov. 2013	23	F	BA	K	2002	10, school	
Grace	Sep. 2013	27	F	MSc	K	2001	6 yrs, tutor, school	USA, 1 yr, as adult
Han	Jan. 2014	21	M	BA	K	2006	7, school	
Hesse	Mar. 2013	23	F	MA	G	Feb. 2013	11, school	USA, 10 mnths, at 16 Australia, 9 mnths, as adult (18+) India, 2 mnths, as adult
Jack	Apr. 2013	26	M	PhD	G	Sep. 2012	12, school	England, 2-3 weeks, as adult
Jack Brown	July 2013	28	M	PhD	SAE	Mar. 2011	N/A	Canada, 3 yrs, as adult
Jess	July 2013	22	F	BA	K	2006	7 yrs, school	
Kahui	July 2013	23	M	PhD	G	Feb. 2012	9.5, school	England, 1 mnth, as adult
Lea	Apr. 2013	25	F	PhD	G	Sep. 2012	14, school	England, 5 mnths, as adult USA, 2 mnths, as adult
Linda	July 2013	21	F	BSc	G	July 2013	10, school	
Louisa	Apr.2013	31	F	PhD	G	Sep. 2010	11, in school	USA, 1.5 yrs, at 16, as adult Canada, 1 yr, as adult
M	Dec. 2013	23	M	PhD	NZE	N/A	N/A	Canada, 1yr, at 5 USA, 4 mnths, as adult England, several weeks, as adult
participant 12	Apr. 2013	23	M	PhD	K	2001	10, moved to NZ	Australia, 2 wks, as adult
Sam	Aug. 2013	21	M	BE	K	2003	11, moved to NZ	Australia, 4 mnths, as adult
Samoth	June 2013	24	M	MA	G	April 2013	2-3, kindergarten	USA, 2 mnths, at 10, 14 South Africa, 2wks, at 12
Sarah	May 2013	30	F	PhD	NZE	N/A	N/A	Scotland, 1yr, as adult England, Ireland,

								Australia, several weeks, as adult
Seung	May 2013	34	M	PhD	K	2004	13-14, school	
Vincent	Apr. 2013	29	M	PhD	K	July. 2011	12, school	Australia, 4 yrs, at 15, as adult
Zwerg	May 2013	25	F	MSc	G	Jan. 2013	12, school	

NNESs with two L1s with a majority of Asian and Caucasian L1 speakers each were chosen because most native speakers of these languages tend to be visually distinguishable from each other, and this was used as an independent variable in a perception study involving these participants (for elaboration see Section 4.4). As well as this, there was an interest in a comparison of two languages which are different in their typological distance from English, so an Asian and a Germanic language were chosen. Also, I am familiar with both German and Korean through formal study, and this knowledge of the languages and cultures was an asset in conducting the research. Lastly, according to the 2013 New Zealand census, for the birthplace of respondent, Korea and Germany were in the top three countries which did not have English as an official language (Statistics New Zealand, 2013).

The two populations of Korean and German speakers were quite different in many respects. Firstly, there was the obvious difference in the first language and, hence, a different phonological system as a starting point for second language acquisition. In NZE, KIT (/ɪ/) is realized as a central vowel, DRESS (/e/) is half-close, TRAP (/æ/) is half-open, GOOSE (/u/) and STRUT (/ʌ/) are central, and FOOT (/ʊ/) is back (Wells, 1982). The short front vowels KIT, DRESS, TRAP were studied for both the Korean and German groups because they are involved in a push chain shift in New Zealand and are said to be a salient marker of New Zealand identity: TRAP is raised, DRESS is raised and fronted, and KIT is centralized (Hay, Nolan et al., 2006). The German language has a counterpart of the KIT vowel, which is quite similar acoustically and perceptually to American English KIT; however, TRAP does not have a similar counterpart, with German short front vowels being higher than TRAP (Strange, Bohn, Trent, & Nishi, 2004). GOOSE and FOOT were additionally studied for the German group as the German language has respective counterparts, but they are produced further back in the vowel space than the English

ones (Strange et al., 2004). Therefore, in a more NZE-like (native-like) production by a German speaker, I expect to see a centralized KIT, a raised and fronted DRESS, a lowered TRAP, and fronted FOOT and GOOSE. The Korean language has two front vowels: a counterpart of FLEECE (/i/) and a merged mid vowel which seems to occupy a wide area from the position reported to be acoustically similar to American English KIT (Baker & Trofimovich, 2005) to about the position of American English DRESS (Yang, 1996). GOOSE and STRUT were additionally studied for the Korean group as the Korean language has respective counterparts, but they are backer than the English ones (Yang, 1996). Therefore, in a more NZE-like (native-like) production by a Korean speaker, I would expect to see a centralized KIT, a raised and fronted DRESS, a lowered TRAP, and fronted and lowered STRUT and fronted GOOSE. Figure 3.1 and Figure 3.2 offer a schematic representation of the German and Korean vowels of interest compared to NZE vowels; arrows indicate the direction of shift which would represent a move towards a more native-like pronunciation for the two L1 groups.

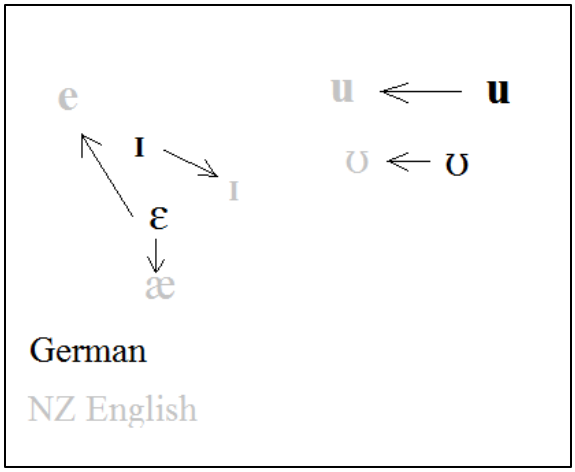

Figure 3.1: A schematic representation of the relative position of German and NZE vowels

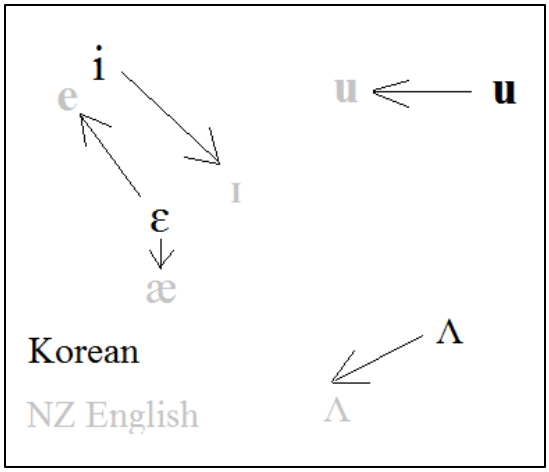

Figure 3.2: A schematic representation of the relative position of Korean and NZE vowels

Additionally, the two language groups were quite different in their history of English acquisition: age of acquisition (AoA) and, for many consequently, length of residence. All participants had an age of acquisition of 10 or higher and, arguably, were slightly past the critical period for pronunciation for the two language groups. Long (1990) argued that the critical period for phonology lasts from 6 until 12 years of age and for morphology and syntax until 15. For Korean L1 speakers of English specifically, a study by Johnson & Newport (1989) set the critical

period at 7-8 years, and Flege, Yeni-Komshian, & Liu (1999) set it at 12 for syntax and 9 for pronunciation. The mean AoA for the higher proficiency Korean group was 11.3 (range 10-13); for the German group, the mean age of first contact in a school setting was 9.8, and mean age of first visit to an English-speaking country was 18.5 (range 10-26), and none of the German group had spent more than 2 months in an English-speaking country before the age of 14. Only two of the German participants had lived in New Zealand for longer than 1 year. For the Korean group AoA coincided with their arrival to New Zealand.

For most Korean L1 speakers, their age of arrival to New Zealand was quite low, while for German L1 speakers it was higher. This is likely a result of a different purpose of arrival. Most of the Korean L1 participants came to New Zealand with their families with an intention to stay for a long time or permanently. All the German L1 participants came to New Zealand alone, of their own accord, to study for a post-graduate degree or participate in an academic exchange. This resulted in a further difference: all Korean L1 speakers had had all of their previous tertiary education conducted in NZE while all German L1 speakers studied previously in German.

3.1.2 Procedure

The participants were recorded with a head-mounted Opus 55.18 MKII beyerdynamic microphone and an H4n Zoom audio-recorder. To ensure participants' interlocutors' privacy, a 0.5 recording level setting on the recorder was used: this way only the speech of the participant was recorded and not that of any of their conversational partners as in line with the Human Ethics Committee approval. I had tested the recording equipment at different recording levels in quiet and noisy environments to ensure a balance between a good quality of the speaker's linguistic production and undecipherability of the speech of the interlocutor.

The participants were recorded speaking English in four different situations of about 15 minutes each, resulting in a total of about one hour of recording per participant. I interviewed all participants about their family and childhood at their home (hereafter referred to as the *family* setting) and about their research and studying on the university campus (the *university* setting). The order of the interviews was counterbalanced. Both were semi-structured interviews eliciting descriptions of the participant's family, a typical childhood day, family vacation, etc. for the *family* setting and descriptions of the participant's subject of study, research, papers written, etc.

for the *university* setting. After the audio-recorded interview at the university, the speakers were additionally video-recorded for about 5 minutes speaking about the applications of their research or study field; these recordings were used in one of the perception experiments (see Section 4.4) and are not analyzed in this chapter.

Between the two interviews, the participants self-recorded conversations in their natural communication with friends (the *friends* setting) and in a minimum of four short service encounters in a public space such as ordering a drink at a coffee shop (the *services* setting). They were instructed to carry the recorder around as they were involved in usual everyday activities and turn it on when they were about to engage in a face-to-face conversation with friends or a service encounter. For the *friends* setting, the participants were instructed to record an everyday conversation with a friend who they normally spoke English to for a minimum of 20 minutes to make sure that at least 10 minutes of their speech would be recorded. There was no limitation on the topic or context of the *friends* setting recording in the instructions given. It took the participants an average of one week to collect the self-recordings. The participants could edit the recordings in order to delete passages that contained personal information that could inadvertently come up while recording. Two participants requested a word and a sentence be edited out in their recordings.

As being aware of the object of study may influence participants' behavior, I could not tell the speakers what I was truly interested in before the recordings were completed. However, since many of the participants knew that my field of study was linguistics, I had to offer them a reasonable explanation that would draw their attention away from their pronunciation, so I told them that I was interested in their choice of words. A debriefing, which explained the true goal of the study and purpose of recordings to the participants, and a questionnaire, which aimed to collect biographical data about the participants and information about the circumstances of the self-recordings, followed the second interview.

The higher proficiency NNES participants, whose linguistic production is analyzed in this section, reported that they believed that the service personnel they had spoken to were native speakers of New Zealand English in more than 90 percent of the encounters. In the *friends* setting, two out of three Korean males and one out of three Korean females reported having spoken to other ethnic Koreans; all German males reported having recorded themselves while speaking to a mix of New Zealanders and foreigners, and two out of three German females to

foreigners, while one reported speaking to a New Zealander and a foreigner. Such a range of conversational partners, which was a methodological oversight that I had not anticipated, made it difficult to make comparisons between speakers. Therefore, the *friends* setting was excluded from production analysis. Future research will benefit from a tighter control over speakers' conversational partners and topics.

The current analysis focuses on the *family*, *university*, and *services* settings, which have different combinations of topic and audience. The immediate audience is the same for the *family* and *university* settings (the author, a female speaker of L1 Russian, a non-native speaker of English); the immediate audience in the *services* setting is, largely, native speakers of New Zealand English; therefore, the expected effect of audience would be different from that of the *services* setting but uniform across the *family* and *university* ones. The settings can also be distinguished according to their orientation towards or away from an 'English' context. That is, although the exact topic of the *university* and *services* settings is different, both are more 'English' focused than the *family* setting because in the latter the participants spoke about their childhood in a foreign country and family members who they communicate with in their L1. In the *university* setting, on the other hand, they spoke about their studies at an English-medium university. For the German L1 group, however, this topic could also trigger memories of their previous German university. Taking into consideration audience design and identity construction accounts, I hypothesized that I would find gradation in native-likeness between the settings: the *services* setting will be the most native-like with English-related topics and interlocutors; the *family* setting will be least native-like, with an L2-related topic; while the *university* setting will occupy an intermediate position, with a non-L2 interlocutor but an L2-related topic (and more so for the Korean L1 group).

3.1.3 Corpus ANNE and formant extraction

LaBB-CAT[3] is a web-browser-based research tool, which stores recordings and transcripts together (Fromont & Hay, 2012). The transcripts can be manually and automatically annotated with the help of CELEX (Baayen, Piepenbrock, & Gulikers, 1995) for, for example, word and lemma frequency and also time-aligned at the level of phoneme with the help of *the Hidden*

3 It can be downloaded from http://labbcat.sourceforge.net/

Markov Model Toolkit (HTK), developed by the University of Cambridge (2014). One advantage that LaBB-CAT has over other similar tools is that it uses a 'train and align' method of alignment where acoustic models are speaker-dependent and created based on the data that is being aligned and not on a set of pre-existing training data. Forced Alignment & Vowel Extraction program suite (Rosenfelder, Fruehwald, Evanini, & Yuan, 2011) uses a corpus of Supreme Court Justices for its acoustic models and has no acoustic models for phonemes other than those found in General American English (MacKenzie & Turton, 2013). This makes LaBB-CAT well suited for alignment of highly idiosyncratic phones, such as those present in L2 idiolects. Time-alignment at the level of segments allows a researcher to search the corpus and extract timing information about segments automatically by simply entering the segment(s) of interest in the search field. Speaker biographical information (age, sex, etc.) can be entered in association with transcripts to allow for filtered search.

The recordings collected for this thesis were orthographically transcribed and time-aligned at the utterance level by hand in Transcriber software (Barras, Geoffrois, Wu, & Liberman, 2001) and then uploaded into the corpus *Accents of Non-Native English* (ANNE), which is an instantiation of LaBB-CAT which was built for the purpose of analyzing stylistic variation in NNESs. ANNE contains over 140,000 word tokens; Table 3.2 shows the word count by speaker for all NNESs. Automatic time-alignment at the level of the word and phoneme was performed as described above. The quality of alignment was checked and manually corrected for 100% of all utterances containing noises (e.g., laughs, loud inhalations and exhalations, coughs, as they are known to hinder accurate alignment) and for 5% of all other utterances for quality control. Of the 5% of checked utterances, the mean percentage of utterances which contained at least one misaligned segment was 15.75% for higher proficiency speakers. Of these the majority contained only one misaligned word.

Table 3.2: Word count for individual NNES participants

L1	Sex	Name	Word Count
German	Male	BrandiATMuhkuh	10992
		Jack	7065
		Kahui	4854
		Samoth	7120

	Female	Hesse	7176
		Lea	7385
		Linda	4657
		Louisa	6069
		Zwerg	9381
Korean	Male	Han	3687
		participant12	6674
		Sam	5635
		Seung	5823
		Vincent	6718
	Female	Emily	4135
		Gabriella	7128
		Grace	7248
		Jess	3846

In its current form ANNE allows researchers to search for grammatical and phonological information which can be accessed through the browser or exported into a spreadsheet facilitating grammatical and auditory analyses. It also allows for acoustic analysis of the segments through direct interaction with Praat (Boersma & Weenink, 2009). The user can open and examine utterances in Praat grids individually (Figure 3.3), or certain information can be extracted for all segments that match the set criteria.

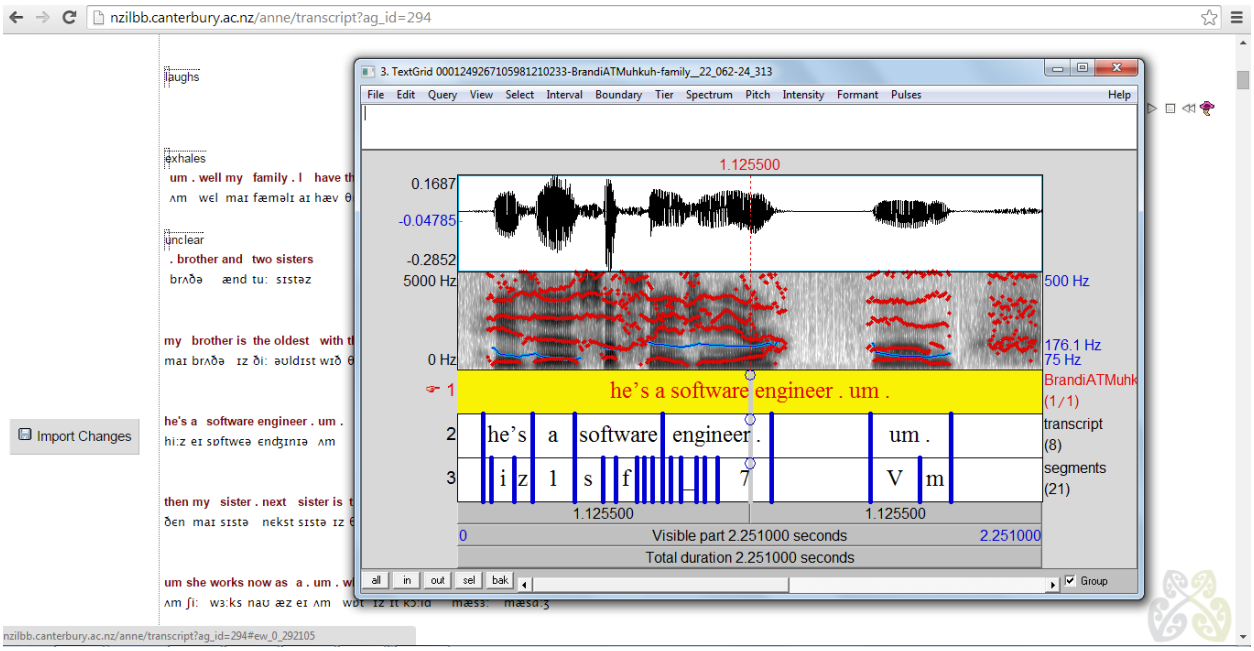

Figure 3.3: LaBB-CAT interface

For example, for the analysis of monophthongs, the corpus can be automatically searched for segments of interest (all or specific monophthongs) by entering them in the search field. Figure 3.4 illustrates how /u/ is searched for on the segments tier in the four transcript types corresponding to the four recording settings. When the segments and their starting and ending points are located, vowel formants for F1, F2, and F3 can be extracted at a set point (e.g., 50%) or number of points (e.g., for diphthongs) and automatically measured using Praat. Currently ANNE contains a total of 54,781 lexically stressed monophthongs (Table 3.3).

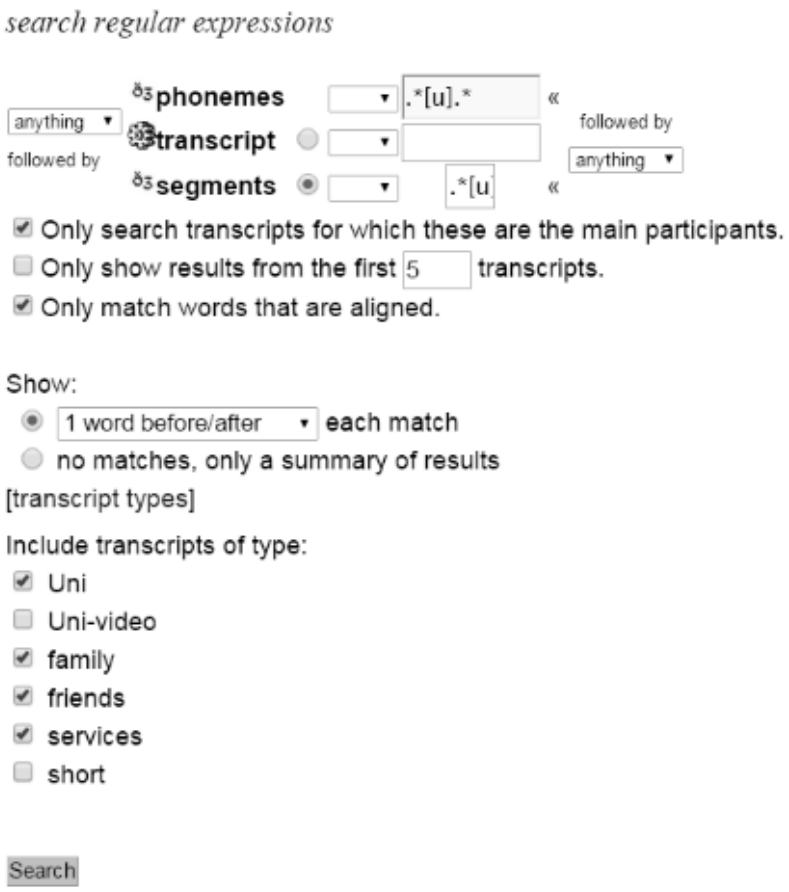

Figure 3.4: The search function in LaBB-CAT

Table 3.3: Stressed monophthong counts for NNESs in ANNE

	German	Korean	Total
DRESS	2682	2412	5094
FLEECE	4372	3347	7719
FOOT	657	391	1048
GOOSE	3008	2741	5749
KIT	5083	3992	9075
LOT	3266	2764	6030
NURSE	1176	1234	2410
START	833	653	1486
STRUT	2610	2318	4928
THOUGHT	1667	1254	2921
TRAP	4681	3640	8321

For the current analysis, vowel formants for F1, F2, and F3 for all stressed monophthongal vowels were extracted at the midpoint and measured using Praat. Figure 3.5 and Figure 3.6 represent the non-normalized vowel spaces of two speakers plotted using R (R Core Team, 2012). The ellipses represent one standard deviation (SD) from the mean. Plotting the vowel spaces may help to visualize as a system what auditory analysis may notice individually, while quantitative formant measurements allow for statistical analysis of points of interest (as has been done in, for example, Gnevsheva, 2013). For instance, from visual analysis of the two vowel spaces, one can easily note certain non-native features in the German L1 speaker BrandiATMuhkuh: for example, his DRESS and TRAP vowels seem to lack a distinction as may be predicted for the L1 vowel system which does not have a counterpart for the TRAP vowel. Also, his DRESS vowel is at the level of the STRUT vowel on the F1 while it is at the level of the FLEECE vowel on the F1 for the L1 speaker of NZE speaker M. For some vowels, however, it is more difficult to make claims based on visual analysis only: the KIT vowel is more centralized for M than for BrandiATMuhkuh, but it is not clear whether this difference is statistically significant.

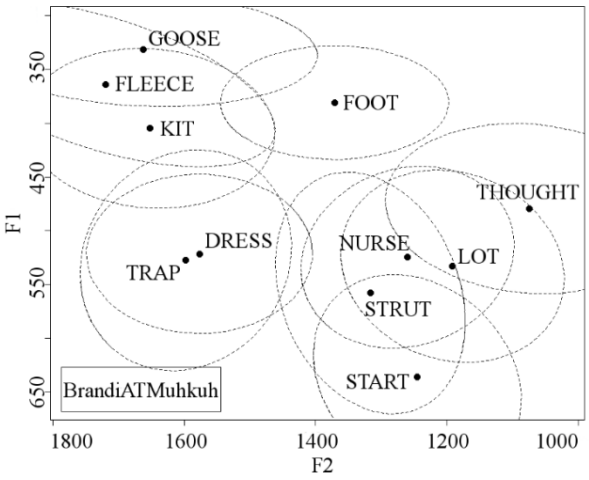

Figure 3.5: The vowel space of a German L1 speaker

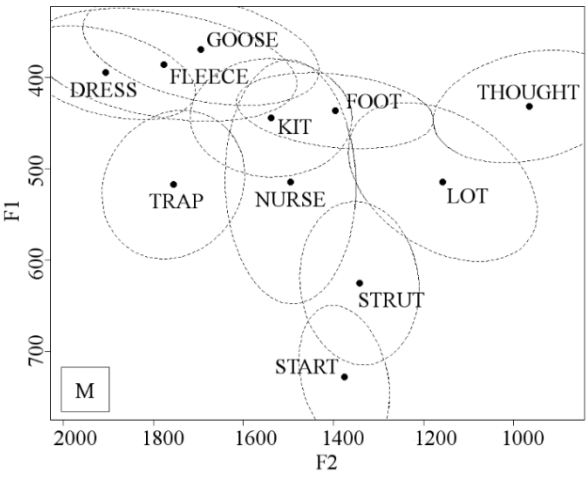

Figure 3.6: The vowel space of a NZE L1 speaker

It is possible to plot vowel means for different settings to explore intraspeaker variation. Figure 3.7, for example, shows the position of the GOOSE vowel in the three settings for a randomly chosen L1 Korean speaker Emily. It can be noted that her production of the vowel

varies on F2 such that it is the most fronted in the services setting (the most L2-oriented setting) and most backed in the family setting (the most L1-oriented). Subsequent statistical analysis of the normalized formant values will help to assess whether the observed differences are statistically significant (see Section 3.2.1).

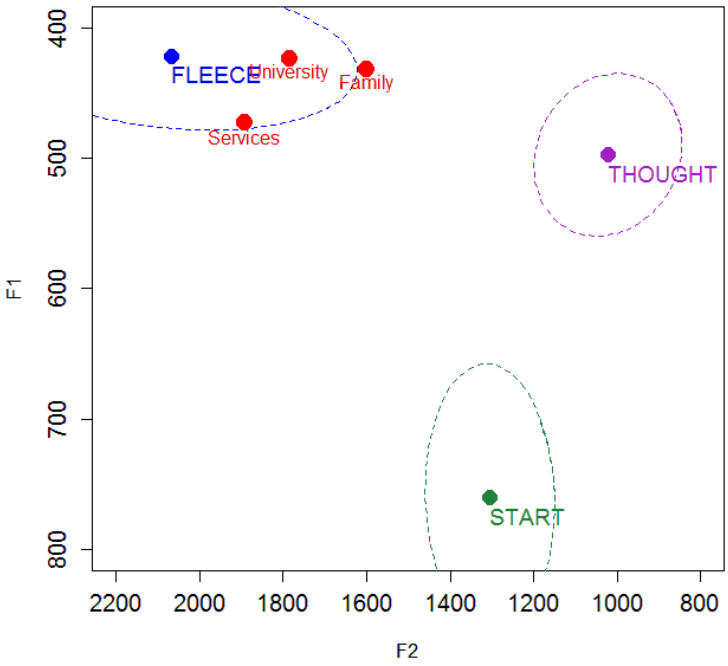

Figure 3.7: Emily's GOOSE vowel in the three settings and surrounding vowel space

When the vowels within two SDs from the mean were plotted for each speaker for preliminary visual analysis, it was noted that none of the vowels' means for any speaker was larger than 810 Hz on the F1 and the maximum standard deviation for the lowest vowel of all speakers was 159.4. As some alignment or measurement error may persist, vowels whose F1 value was larger than 1000 Hz were excluded from analysis as well as vowels whose formant values were not within two SDs from the mean for each vowel in each setting for each participant. The vowels of all the speakers were normalized with the Lobanov normalization method to allow for inter-speaker comparison (Thomas & Kendall, 2007). Table 3.4 and Table 3.5 show the number of analyzed tokens per setting per vowel of interest for the Korean and German L1 speakers.

Table 3.4: Vowels chosen for analysis; number of tokens per vowel per setting in the recordings of Korean L1 speakers

	DRESS	GOOSE	KIT	STRUT	TRAP	**total**
family	509	554	924	562	935	**3484**
services	40	75	49	41	82	**287**
university	522	589	864	449	878	**3302**
total	**1071**	**1218**	**1837**	**1052**	**1895**	**7073**

Table 3.5: Vowels chosen for analysis; number of tokens per vowel per setting in the recordings of German L1 speakers

	DRESS	FOOT	GOOSE	KIT	TRAP	**total**
family	691	161	603	1012	1230	**3697**
services	160	62	196	265	282	**965**
university	396	87	405	764	707	**2359**
total	**1247**	**310**	**1204**	**2041**	**2219**	**7021**

3.1.4 Statistical analysis

The normalized vowel formant measurements for the vowels of interest were analyzed using R (R Core Team, 2014). Linear-mixed effect models allow us to model fixed and random effects on a dependent variable (*normalized vowel formant measurements*, in this case) (Baayen, Davidson, & Bates, 1998). Fixed effects are independent variables which are regarded as non-random, such as the *setting* in the present study. Random effects are explanatory variables that are treated as arising from random causes; it is a source of variance which the model allows us to control for but which we might not be interested in for the particular study (e.g., individual *word* from which the vowel was extracted and *speaker* in the current study). Mixed effects models may include random slopes as well as random intercepts (Barr, Levy, Scheepers, & Tily, 2013). A random intercept for a certain variable allows the effect to vary within that variable: for example, one participant can *in general* have a more centralized KIT vowel than the next participant. Random slopes allow the effect to vary for a given intercept: the difference in KIT production from setting to setting for one participant can be larger or smaller than it is for the next participant.

I ran mixed-effects models for the two L1 groups separately for each vowel (KIT F1 and F2, DRESS F1 and F2, TRAP F1, GOOSE F2 for both groups; additionally, FOOT F2 for the German L1 group and STRUT F2 for the Korean L1 group) with the normalized formant values as the dependent variable. The full model included an interaction between *setting* and *speaker sex*, as well as *preceding* and *following phonological environment*, *duration of the vowel*, *log CELEX frequency*, and *CELEX category of the word* (function vs content; Baayen et al., 1995) as fixed effects, with *word* and *speaker* as random intercepts, and *setting* as a random slope for *speaker*. The university setting was treated as the reference level (Intercept) as it had been hypothesized to behave as an intermediate one in terms of nativeness. If a fixed effect was found to be non-significant, it was removed and the model was re-run, and then the two models were compared with an ANOVA to test whether either of the two models was significantly better. *Preceding* and *following phonological environment* were retained in all models; however, for Korean males the STRUT model did not include *preceding environment* and the FOOT models for German males and females did not include *following environment* for the benefit of model convergence. These models were run separately, with either following or preceding environment retained, and subsequently compared with an ANOVA. In each case, the best model of the two was chosen. Full final models are shown in Appendix B; however, for the benefit of space, phonological environment is excluded from the tables presented in the chapter as indicated with marks of omission.

3.2 Results and discussion

Section 3.2.1 discusses within-speaker variation among settings for one L1 Korean speaker Emily who exhibited the most extreme variation in passing behavior as she passed for a NS of NZE approximately as often as she failed to pass for a NES at all (see more in Chapter 5). The following sections discuss the models including multiple speakers; the final section comprises the general discussion for the chapter.

3.2.1 Case study of within-speaker variation: Emily

Figure 3.7 illustrated how speakers may produce individual vowels with the mean formant value on the continuum between more L1-like and more L2-like using Emily's production of the

GOOSE vowel as an example. This section employs statistical analysis to verify whether the observed differences in this and other vowels were significant. Linear-mixed effects models were fit to the Emily data only, with the normalized formant values of the vowels of interest as dependent variables. The full model included *setting*, *preceding* and *following phonological environment*, *duration of the vowel*, *log CELEX frequency*, and *CELEX category of the word* (function vs content; Baayen et al., 1995) as fixed effects and *word* as a random effect. The models were pruned as described in Section 3.1.4; full models can be seen in Appendix B.

Emily style-shifted between the three settings with significant differences found for KIT F1, DRESS F1, and GOOSE F2. She produced a more native-like KIT in the services and a less native-like KIT, DRESS, and GOOSE in the family setting compared to the university. Table 3.6 represents the final model for KIT exclusive of phonological environment. The *estimate* and the *standard error* columns in the table give us the predicted normalized F1 for the vowel and standard error for a level respectively. So for the Intercept (level *university* of the factor *setting* and the base levels of other variables, such as phonological environment and word frequency), the predicted normalized value is -1.378. To calculate the predicted normalized F1 for a different level of factor setting, the respective value in the *estimate* column is added or subtracted. For example, KIT F1 was 0.207 higher in the family setting and 0.539 lower in the services setting than in the university setting, which was significantly different from the baseline as indicated in the *significance* column. The relationship is illustrated in Figure 3.8. The DRESS vowel was significantly lower (Table 3.7; Figure 3.9) and the GOOSE vowel was significantly more backed Table 3.8; Figure 3.10) in the family setting compared to the university setting. The model of GOOSE F2 shows no significant difference between the family and the university settings despite a visual difference in Figure 3.7, illustrating the importance of statistical analysis.

Table 3.6: Summary for model of KIT F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-1.378	0.692	181.960	-1.991	0.048	*
…						
celex_frequency	0.082	0.036	181.960	2.263	0.025	*
setting_family	-0.207	0.084	181.960	-2.465	0.015	*
setting_services	0.539	0.246	181.960	2.192	0.030	*

Note. * p<0.05; ** p<0.01, *** p<0.001

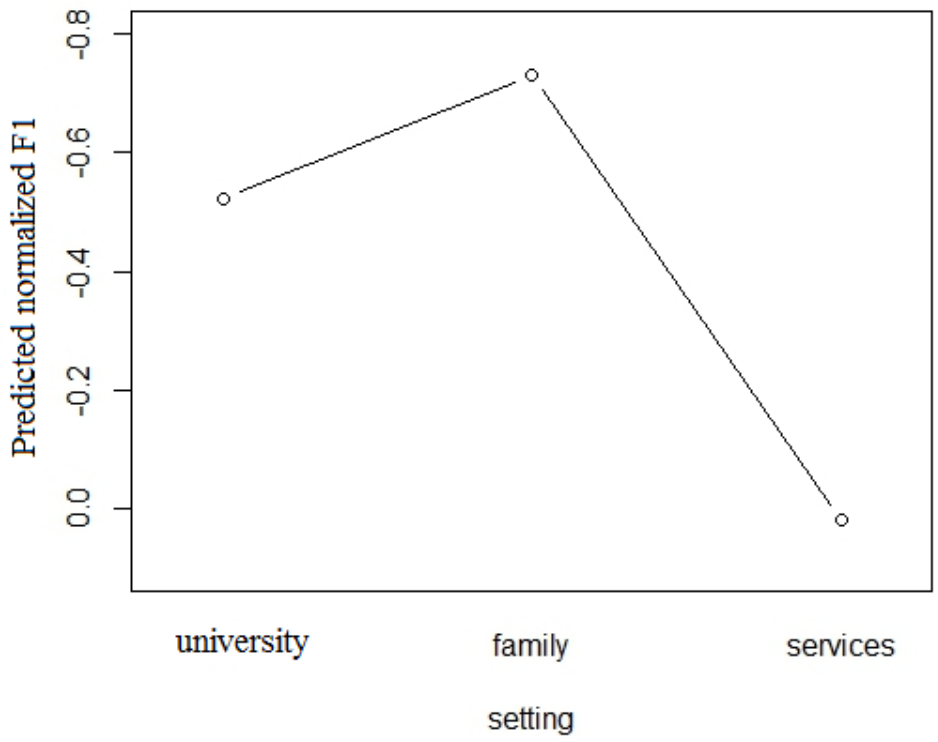

Figure 3.8: Model prediction for KIT F1 produced by Emily in the three different settings (from model in Table 3.6)

Table 3.7: Summary for model of DRESS F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-1.326	0.319	76.980	-4.160	0.000	***
…						
category_function	0.602	0.184	76.980	3.267	0.002	**
setting_family	0.463	0.182	76.980	2.540	0.013	*

Note. * p<0.05; ** p<0.01, *** p<0.001; the services setting is missing from the model because Emily did not produce any DRESS vowels in her services recordings

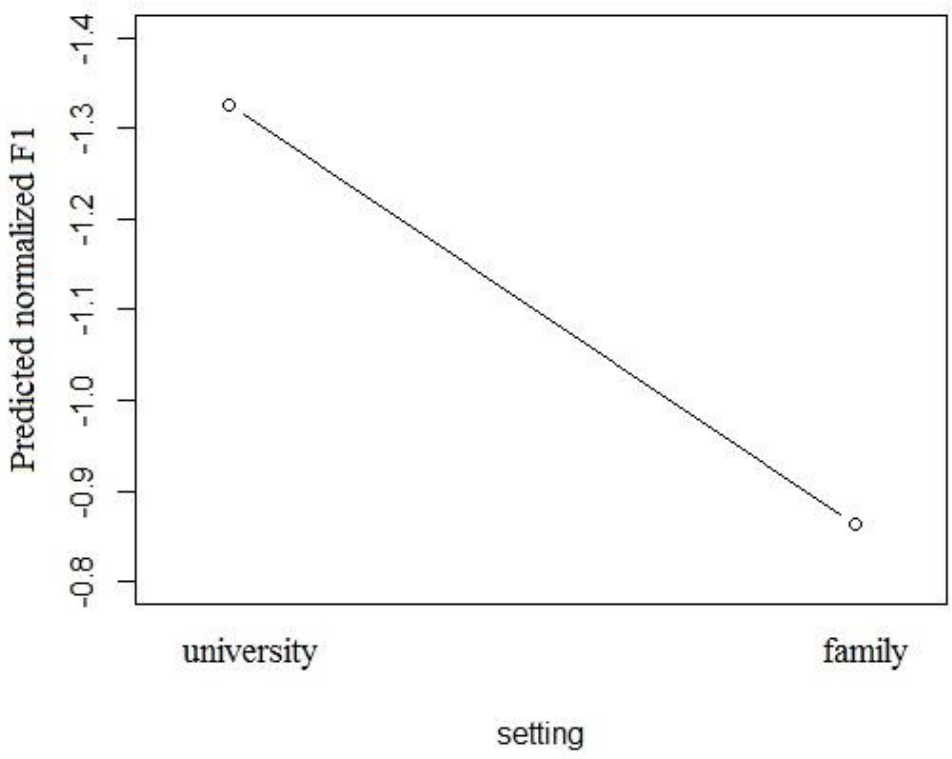

Figure 3.9: Model prediction for DRESS F1 produced by Emily (from model in Table 3.7)

Table 3.8: Summary for model of GOOSE F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	1.637	0.633	127.960	2.584	0.011	*
celex_frequency	-0.136	0.053	127.960	-2.548	0.012	*
setting_family	-0.395	0.076	127.960	-5.192	0.000	***
setting_services	-0.150	0.148	127.960	-1.010	0.314	

Note. * p<0.05; ** p<0.01, *** p<0.001

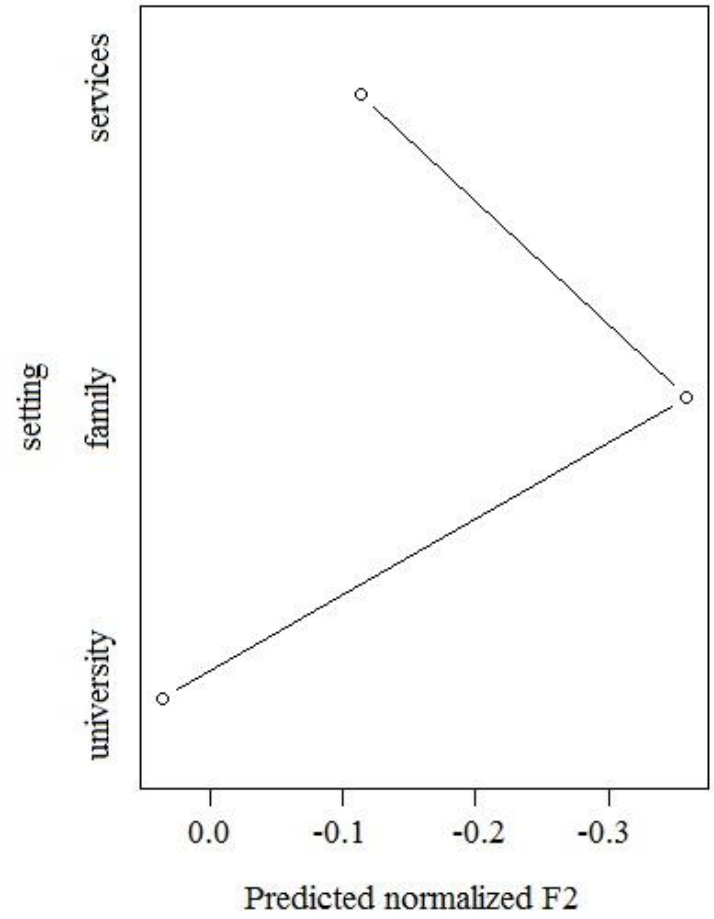

Figure 3.10: Model prediction for GOOSE F2 produced Emily in the three different settings (from model in Table 3.8)

3.2.2 Style-shifting in German speakers

The previous section focused on the variation in one speaker who exhibited an interesting trend in passing behavior by passing for a NS of NZE just a little more often than not passing for a NES at all; in this and the following sections I explore variation in groups of speakers.

The German participants exhibited some variation with regard to the vowels studied, with the *services* setting found to be most native-like. KIT was lower (more NZE-like) in the *services* setting compared to the *university* setting. Table 3.9 represents the final model for KIT exclusive of phonological environment (Figure 3.11). Variation in DRESS, TRAP, FOOT, and GOOSE

did not reach significance. No significant interaction between setting and sex was found in any of the models.

Table 3.9: Summary for model of KIT F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-0.586	0.480	1601.0	-1.222	0.222	
…						
duration	2.448	0.304	2031.0	8.051	0.000	***
setting_family	0.007	0.040	6.1	0.167	0.873	
setting_services	0.294	0.044	10.3	6.673	0.000	***

Note. * p<0.05; ** p<0.01, *** p<0.001

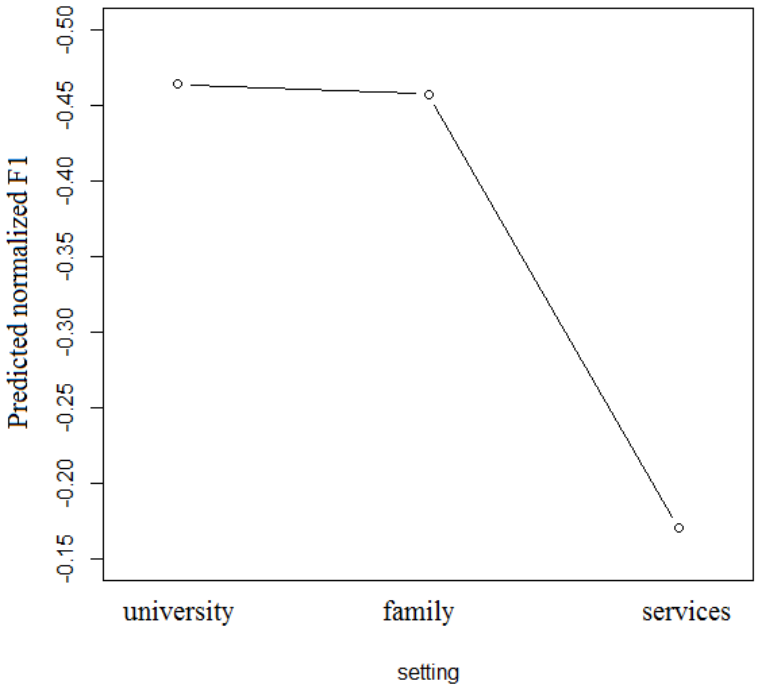

Figure 3.11: Model prediction for KIT F1 produced by German L1 speakers in the three different settings (from model in Table 3.9)

Other variables, such as following and preceding phonological environment, vowel duration, log CELEX word frequency and category, were found to be significant in several of the models (see full model output for all vowels in Appendix B). Unsurprisingly, phonological environment was found to be significant, and some following and preceding phonemes had a

significant effect on the production of vowels. Additionally, KIT, DRESS, and TRAP were significantly lower, front vowels KIT and DRESS were significantly fronter, and the back vowel GOOSE was significantly backer than the baseline in vowels of longer duration. For example, as can be seen in Table 3.9, the coefficient for duration is 2.448, which means that the predicted normalized F1 increases by this estimate for every unit of duration. This could be explained by a centralization effect of shorter vowels. Next, a word category effect was found such that in function words the DRESS vowel was produced higher and backer than in the content words.

A significant log CELEX word frequency effect was found for the KIT vowel such that the higher the word frequency the backer (more NZE-like) the production of the KIT vowel. This is largely in line with the usage-based models of speech production: the words that are more frequent have been perceived and produced more in the NZE accent allowing for a more NZE-like production of more frequent words. Additionally, the higher was the word frequency the higher was the production of the TRAP vowel. It can be argued that there are two opposing processes influencing the TRAP vowel in these L2 speakers at the same time. On the one hand, a category absent in their L1 is being created through distinguishing DRESS and TRAP, which may result in the raising of DRESS and lowering of TRAP in German NNESs as visualized in Figure 3.1: A schematic representation of the relative position of German and NZE vowels. On the other hand, the word frequency effect, as predicted by usage-based models, may result in the raising of TRAP in higher frequency words due to exposure to a raised TRAP in frequent, compared to infrequent words in native speech, the manifestation of which we can see in this model (Table 3.10). These factors are not the main focus of this thesis, and so are not further discussed in Section 3.2.4.

Table 3.10: Summary for model of TRAP F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	0.379	0.497	837.100	0.762	0.446	
…						
celex_frequency	-0.043	0.015	235.400	-2.867	0.005	**
duration	3.859	0.246	2172.000	15.702	0.000	***
setting_family	-0.029	0.110	6.000	-0.262	0.802	
setting_services	0.089	0.065	12.400	1.367	0.196	

Note. * p<0.05; ** p<0.01, *** p<0.001

3.2.3 Style-shifting in Korean speakers

Korean L1 speakers showed more variation than the German speakers and were more NZE-like in their production of the vowels of interest in the *services* setting and less NZE-like in the *family* setting compared to the *university* setting (see

Table 3.11). The speakers had a significantly lower (more NZE-like) production of the KIT and TRAP vowels in the *services* setting compared to the *university* setting (Table 3.12, Figure 3.12 and Table 3.13, Figure 3.13 respectively). They also produced a backer (less NZE-like) GOOSE in the *family* setting compared to the *university* setting (Table 3.14, Figure 3.14). Variation in DRESS and STRUT failed to reach significance. No significant interaction between setting and sex was found in any of the models.

Table 3.11: Significant differences from the university setting for Korean L1 speakers

Family	Services
GOOSE backer (less NZE-like)	KIT lower (more NZE-like) TRAP lower (more NZE-like)

Table 3.12: Summary for model of KIT F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-1.113	0.496	1437	-2.246	0.025	*
…						
duration	1.722	0.312	1826	5.528	0.000	***
setting_family	-0.041	0.047	6	-0.870	0.416	
setting_services	0.414	0.108	4	3.848	0.018	***

Note. * p<0.05; ** p<0.01, *** p<0.001

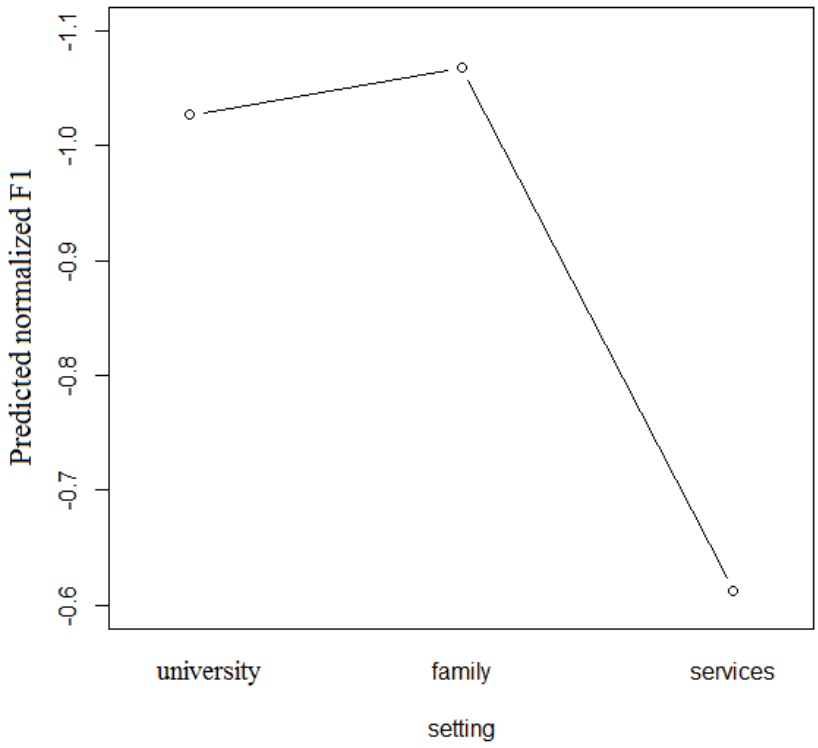

Figure 3.12: Model prediction for KIT F1 produced by Korean L1 speakers in the three different settings (from model in Table 3.12)

Table 3.13: Summary for model of TRAP F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	0.287	0.378	345.2	0.759	0.449	
…						
celex_frequency	-0.063	0.014	165.5	-4.565	0.000	***
duration	2.921	0.208	1880.0	14.026	0.000	***
setting_family	-0.076	0.063	5.7	-1.197	0.279	
setting_services	0.420	0.105	5.7	3.889	0.009	**

Note. * p<0.05; ** p<0.01, *** p<0.001

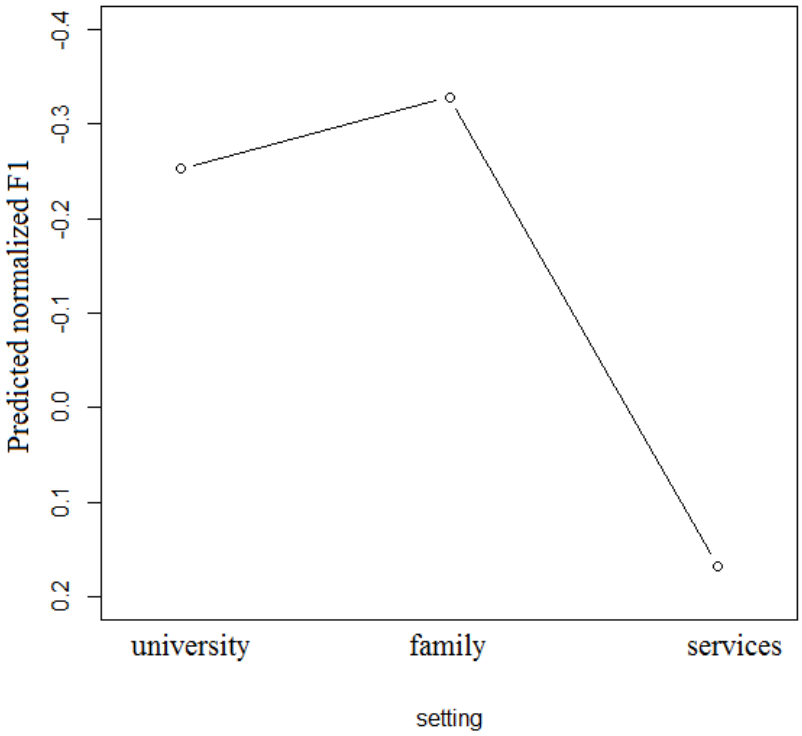

Figure 3.13: Model prediction for TRAP F1 produced by Korean L1 speakers in the three different settings (from model in Table 3.13)

Table 3.14: Summary for model of GOOSE F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	0.605	0.182	19.8	3.321	0.003	**
…						
duration	-2.276	0.205	1203.0	-11.124	0.000	***
setting_family	-0.177	0.068	5.7	-2.582	0.044	*
setting_services	-0.074	0.102	1.5	-0.715	0.571	

Note. * p<0.05; ** p<0.01, *** p<0.001

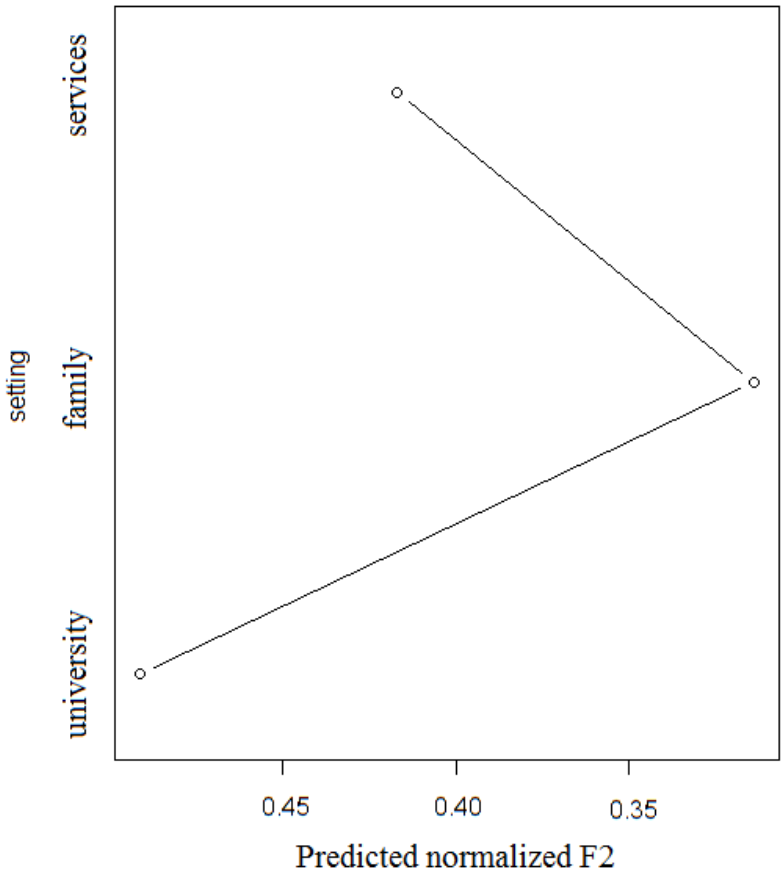

Figure 3.14: Model prediction for GOOSE F2 produced by Korean L1 speakers in the three different settings (from model in Table 3.14)

The other variables, such as phonological environment, vowel duration, and log CELEX word frequency were found to be significant in several of the models (see model output in Appendix B). For Korean L1 speakers, KIT, DRESS, and TRAP were significantly lower, front vowels KIT and DRESS were significantly fronter, and back vowels STRUT and GOOSE were significantly backer than the baseline in vowels of longer duration, which could be explained by a centralization effect of shorter vowels. It should be noted that a significant log CELEX word frequency effect was found for the TRAP vowel such that the higher the word frequency the higher the production of the TRAP vowel which is similar to the effect found for the German L1 speakers as discussed in Section 3.2.2. However, these factors are outside the scope of this thesis and will not be discussed further.

3.2.4 General discussion

At the beginning of this chapter, I asked four questions, to which I now return.

- Do L2 speakers use differences between L1 and L2 vowel systems for situational style-shifting?

The finding that the participants in both language groups exhibited some sort of variation in their production of vowels between the settings lends support to the hypothesis that L2 speakers may use the differences between L1 and L2 vowel systems for situational style-shifting in production. This result is in line with the findings by Rampton (2011) and further suggests that this sort of variation is neither idiosyncratic nor L1-specific despite some differences between the German and Korean L1 groups.

- Does L2 speakers' style-shifting use the differences between L1 and L2 systems as a continuum as opposed to a binary choice?

The findings suggest that some speakers may use the accentedness continuum between more L1-like and more L2-like as opposed to two extremes. The L1 German group showed binary differentiation in style-shifting between the settings in KIT F1 as I found a significant difference in production between the university and services settings only. For the L1 Korean speakers, there was a significant difference between the university and services settings in their production of KIT F1 and TRAP F1 and a significant difference between the university and family settings in their production of GOOSE F2. Additionally, there was a significant difference in Emily's production of the KIT vowel between the family and university settings and between the services and university settings; the other models showed a binary differentiation. The found differences among the three settings for the Korean speakers suggest the existence of an accentedness continuum that NNSs navigate. On the one hand, these findings may be reflective of the speakers' variation on a phonetic continuum, using variants which are more L1-like at one extreme and more L2-like at the other. On the other hand, these results may be due to a change in probabilities of L1 and L2 forms in different settings. The mixed effects models used in this study do not allow us to distinguish between these two possibilities; however, in any case the

found variation between more L1-like and more L2-like mean productions on the accentedness continuum suggests that speakers may be using a gradation of NZE-likeness rather than two extremes.

The services setting with an L2-related audience and topic was most native-like for both language groups. This finding is in line with the audience design (Bell, 1984) account of style-shifting because the participants were most NZE-like as they accommodated to a NS audience in the service encounters as opposed to the family and university settings with a NNS of New Zealand English interviewing them. The identity construction account (Eckert, 2000) also predicted services to be the most native-like topic of the three as there was less need or incentive to express one's identity (Piller, 2002).

The difference found for the L1 Korean group between the family and university settings (with the same addressee - the interviewer) highlights the importance of referee design and/or identity construction. The participants were used to a more Korean audience at home creating a more Korean referee for the topic, and there was more incentive to bring more of the L1 influences to the forefront in construction of their identity when talking about their family and childhood. For the L1 German group, however, there was no significant difference between the family and university settings. As all of them had been educated in Germany for most of their past and 4 out of 6 had moved to New Zealand only in the preceding year, it is possible that their mental representation of the university setting had not yet acquired many NZE influences and was more German-like, resulting in a German-related referee for style-shifting. Identity construction for both topics was probably rather similar because of their relatively short stay in the L2 community. Future studies with a more complex combination of audience and topic types will help to explore the extent of possible variation among settings.

- Do speakers of different language backgrounds style-shift differently?

Overall, the Korean L1 participants showed a greater number of significant differences between the settings than the German L1 participants did, suggesting that speakers of different L1 backgrounds may style-shift differently. The current data do not offer a clear explanation for this finding; however, the two groups are different in a number of ways, as outlined in Section 3.1.1, which could have produced such an effect and could be tested in future research: L1, AoA, and length of residence among others. For example, we might hypothesize that a longer length of

residence correlates with the breadth of sociolinguistic variation as speakers' exposure to different settings and audiences is enriched and the L2 identity is further developed.

- Do male and female L2 speakers style-shift differently?

The male and female L2 speakers in this study style-shifted similarly along the L1-L2 continuum which was perhaps an unexpected finding based on the results of previous gender-related linguistic research. The interaction between speaker sex and setting was not found to be significant; moreover, there was no main effect of speaker sex. The male and female participants in the two language groups did not differ significantly in either their production of the vowels or variation in that production from setting to setting. This is different from Sharma (2011) who found that younger female speakers style-shifted more and also exhibited variation by setting. She explained this through the females' more diverse social networks. I did not collect social network information about the speakers in my study, so it is not possible to rule out that it was similar for male and female participants. Additionally, Sharma's finding is based on second generation speakers, and it may be possible that male and female second language speakers do not style-shift differently. Finally, as there were only three participants in each sex by L1 group, there may not have been enough statistical power to detect such a difference.

Summary

To sum up, this production study has found within-speaker variation among settings in NNESs. This suggests that L2 speakers are not limited to sociolinguistic variation exhibited by native speakers of a language and are creative users who can adapt and employ (consciously or unconsciously) the L1 and L2 resources available to them for style-shifting. I previously called this Type 3 variation. In this light the differences exhibited by L2 speakers in target sociolinguistic variation (as in, for example, Schleef et al., 2011) can be seen as sociolinguistic variation in an L2 variety.

Moreover, the accounts that have been successfully applied to variation in L1 varieties, namely audience design (Bell, 1984) and identity construction (Eckert, 2000), have proved useful in understanding variation on the L1 – L2 continuum for L2 speakers which suggests that many processes underlying sociolinguistic variation in L1 and L2 varieties are universal. The

attested gradation on the accentedness continuum from more L1-like to more L2-like suggests that there is a complex set of factors influencing the speakers' production at any given moment and L2 speakers reflect this in their speech. Such factors include but are not limited to audience, topic, and speaker linguistic background.

However, communication is a joint performance, so is identity construction, and reliance on production only will not paint the full picture. If the signs used by the speaker for style-shifting are not noticed or interpreted by the listener in the way intended by the speaker, the understanding of identity or meaning will not be full. The perception studies in the next chapter aim to clarify the variation in perception of the speakers.

Chapter 4 : Variation in Speech Perception

This chapter describes three perception experiments that explored variation in accentedness. In the first and second sections several listener-dependent and listener-independent factors are explored in Experiments 1 and 2: recording setting, speaker sex, speaker proficiency, speaker L1, listener sex, listener age, etc. The focus, however, is on recording setting with the main research question formulated in the following way:

- Is there an effect of recording setting on perceived accentedness of a NNES?

To address this question, NES listeners were presented with clips recorded by NNESs in different settings in two accentedness rating experiments. In the first experiment, the listeners were presented with NNES clips only, four clips from the same speaker at a time. In the second experiment, listeners were presented with one clip at a time with both NES- and NNES-produced clips. There were also differences in the rating scales used: in Experiment 1 the speakers were rated on a scale which read 'I can hear a very strong foreign accent' and 'I cannot hear a foreign accent at all' at the two extremes and in Experiment 2 the scale read 'Definitely a first language speaker of English' and 'Definitely a second language speaker of English'. Thus, in the first experiment any deviation from the NZE target (even native to other varieties of English such as American English) would be considered a manifestation of an accent while in the second experiment only deviations due to an assumed L1 interference would result in a lower score. Consequently, Experiment 1 explores variation in NZE-accentedness (with implications for passing for a NS of the same dialect as listeners) and Experiment 2 explores variation in English-native-likeness (with implications for passing for a NS of any English variety; see Section 4.3 comparing the different scales in respect to the results). The results of the experiments are followed by a consideration of similarities and differences between the various experimental procedures employed in this thesis for researching of variation in accentedness.

The perception experiments employed clips from the four settings in which the speakers were recorded (as described in Section 3.1.2). As noted above, there was much more variation in the friends setting as the speakers reported more than 50% of their interlocutors being L2 speakers. The friends setting varied most on topic as well whereas the other ones were more

uniform. Thus, the results pertaining to the friends setting should be interpreted keeping this variation in mind.

In Section 4.4, I explore the effect of speaker ethnicity on accentedness perception and report on a third accentedness rating experiment in which two groups of NNESs of Asian and Caucasian ethnicities were presented to listeners in three conditions: audio track of the recording only, video track only, and audiovisual (audio and video tracks of the recording together). As discussed in Section 2.3.2, reverse linguistic stereotyping (Rubin, 1992) and audiovisual mismatch effect (McGowan, 2015) have different predictions as to the accentedness rating foreign-accented clips may receive when presented with a Caucasian or an Asian face (see Table 4.1).

Table 4.1: Two accounts' predictions for Asian and Caucasian non-native English speakers' (NNESs) accentedness ratings in two conditions

	Asian NNES	Caucasian NNES
Reverse linguistic stereotyping	Audiovisual > Audio	Audiovisual < Audio
Audiovisual mismatch effect	Audiovisual = Audio	Audiovisual > Audio

The research questions motivated by the literature discussed in Section 2.3.2 are as follows:

- What is the effect of availability of visual information for Asian NNESs in an accentedness perception task?
- What is the effect of availability of visual information for Caucasian NNESs in the same accentedness perception task?
- Will these effects for Asian and Caucasian NNESs be better predicted by reverse linguistic stereotyping or an audiovisual mismatch?

The descriptions of studies are followed by a general discussion summarizing the main findings.

4.1 Experiment 1: Effect of setting on accentedness perception

4.1.1 Method

Stimuli (speakers)

The audio stimuli in Experiments 1 and 2 were the same short clips extracted from the recordings of the 24 speakers in the four different settings (family, friends, services, and university; see Sections 3.1.1 and 3.1.2 for details about the speakers and recording procedure). By way of reminder, the speakers were interviewed about their family at home in the family setting, about their studies on campus in the university setting; they also self-recorded themselves in short service encounters (the services setting) and when talking to friends (the friends setting).

Most of the clips from the family, friends, and university settings were extracted after the initial 5 minutes of recording where the speaker might have been adjusting and were uninterrupted; however, in most of the services settings the time between the speakers' turns was edited out and, therefore, the clips were sometimes interrupted by short periods of silence. Three clips per setting were extracted for each speaker in order to lessen the effect of individual recordings, resulting in a pool of 288 clips (24 speakers * 4 settings * 3 extractions). The clips contained a minimum of 25 words in the family, friends, and university settings; because stopping the clips mid-phrase could have an effect on the listeners' perception, the exact number of words per clip was allowed to vary. Also, as service encounters can be quite brief, some services clips were shorter than 25 words: mean length 22.2 words and 10.11 seconds. The mean length for all clips in the four settings was 26.6 words and 13 seconds. The recordings were normalized to remove variation in volume. Because grammatical inaccuracies and disfluencies can influence judges' ratings of accentedness, I made an attempt to choose clips without errors and hesitations; but for some lower proficiency speakers it proved impossible to find such a passage. The clips did not contain names of persons, geographical locations, or any other extra-linguistic information that might draw attention to the speakers' foreignness. Experiment 1 employed the 216 clips from the NNESs only.

Participants (Listeners)

The listeners in Experiment 1 were 25 native speakers of New Zealand English who were recruited through announcements posted around the University of Canterbury campus and the friend-of-friend method (Milroy, 1987). There were 16 females and 9 males. The age, education, socio-economic class of the participants were comparable to those of the investigator and speakers: age range 18-69, mean age 27, median age 22. All had achieved or were studying towards a Bachelor's degree or above at the time of the study. Ten claimed no knowledge of a foreign language.

Procedure

The listeners were seated individually in a quiet lab in front of the computer with head-phones. Stimuli were presented electronically using the E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2012). Before starting the actual task, the listeners read the instructions on the screen (Appendix C), completed a practice trial with comparable clips from a male NS of New Zealand English and if needed, adjusted the volume and clarified the procedure with the research assistant (the author). After that, the listeners were presented with 18 sets of four clips, each set corresponding to a speaker with a random combination of clips from the four settings. In the task, the listeners were instructed to rate the presented clips and place them on a scale which read 'Very strong foreign accent' and 'No foreign accent at all' at the two extremes (Figure 4.1). At the top of the screen there were four symbols associated with the four clips recorded in different settings. The listeners played the clips by clicking on each of the four symbols one at a time and indicated their accentedness rating by clicking on the scale below where an identical symbol then appeared (Figure 4.2). They could replay the clips for a given speaker and change the position of symbols on the scale until they moved on to the next speaker by clicking on 'Done'.

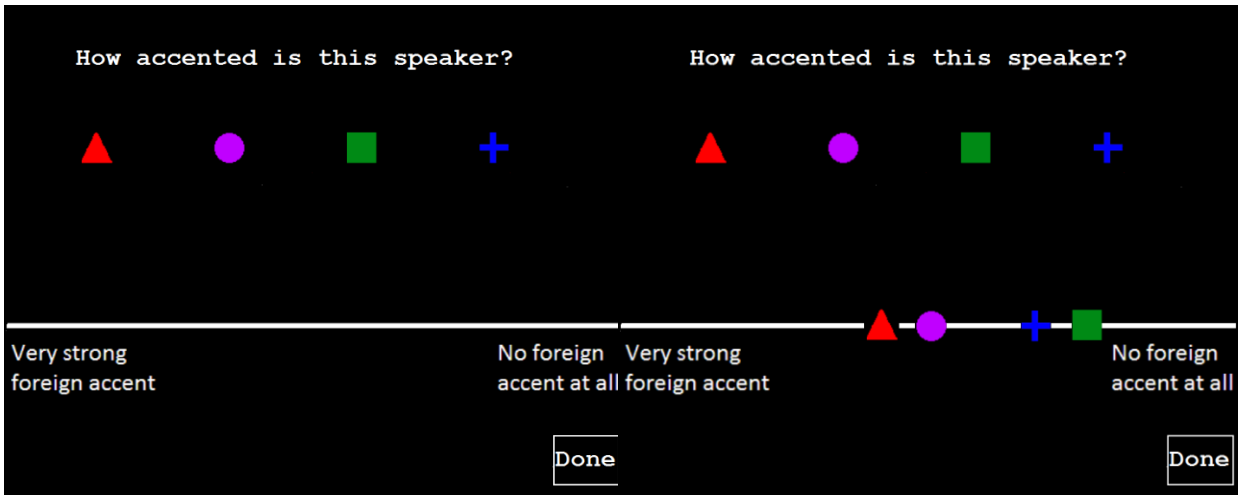

Figure 4.1: Slide presented to listeners in Experiment 1

Figure 4.2: Slide with listener response in Experiment 1

The order of the 18 speakers in the experiment, the extractions (i.e. the choice of one of the three clips for each speaker in each setting), the relative position of the four symbols and the four settings on the screen were randomized. The task was self-paced and took about 30 minutes to complete. At the end, the listeners completed a short biographical questionnaire (see Production study Post-Recordings Questionnaire in Appendix A). They were given a $10 coffee voucher for completing the task. The research was reviewed and approved by the University of Canterbury Human Ethics Committee.

4.1.2 Results and discussion

The position of the symbols on the scale was recorded as an accentedness rating from 1 (Very strong foreign accent) to 100 (No foreign accent at all) which was subsequently analyzed using R (R Core Team, 2014). A linear mixed-effects model was fit to the data with the perceived accentedness rating as the dependent variable. The fixed effects in the full model included two-way interactions between *setting* and each of the other variables as well as their main effects: *speaker L1*, *speaker proficiency*, *speaker sex*, *listener age*, *listener sex*, *listener L2 knowledge* (binary)[4], *mean log CELEX frequency of CELEX content words in the clip* (Baayen et al., 1995),

[4] The listeners were asked whether they spoke any language besides English and how well. Because some participants only listed the languages and did not comment on their proficiency, it was impossible to make more

clip length in seconds, *clip length in words*, *speech rate* (words per second), *progression* in the experiment (1 through 18; to control for a potential familiarization or fatigue effect). *Speaker*, *clip*, and *listener* were included as random intercepts. *Setting* was introduced as a random slope for *listener* (Barr et al. 2013). If an interaction or a fixed effect was found to be non-significant, I simplified the model by excluding the interaction or the variable from the model and then compared the previous and the current models with an ANOVA. The significantly better or simpler model was kept.

Table 4.2 represents the final model. The higher proficiency speakers in the university setting were chosen as the reference level because this setting was chosen as the base level in the production models. For the Intercept (the higher proficiency speakers in the university setting), the predicted accentedness rating is 45.696. The higher proficiency speakers received a rating 5.081 higher in the services setting than in the university setting; this difference was significant. The difference in accentedness ratings between the friends and the university settings was not found to be significant. The difference between the family and the university settings did not reach the level of significance p=0.05; however, the trend was in the direction of the family setting being judged less accented.

Table 4.2: Model summary for accentedness ratings of NNESs in different settings

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	45.696	4.951	34	9.230	0.000	-
setting_family	2.456	1.307	169	1.879	0.062	
setting_friends	0.600	1.325	1567	0.451	0.652	
setting_services	5.081	1.309	169	3.881	0.000	***
proficiency_lower	-19.789	7.147	19	-2.769	0.012	*
progression	-0.132	0.067	1717	-1.972	0.049	*
setting_family :	-0.368	2.263	169	-0.162	0.872	

minute distinctions. Listener L2 knowledge in this thesis divides participants into two groups: those who reported any L2 languages and those that did not.

proficiency_lower						
setting_friends : proficiency_lower	0.697	2.267	166	0.307	0.759	
setting_services : proficiency_lower	-4.561	2.264	169	-2.015	0.046	*

Note. * p<0.05; ** p<0.01, *** p<0.001

This finding suggests that listeners' accentedness perception may vary by setting of recording. As predicted by Piller (2002), who found that L2 speakers believed they passed for a native speaker more commonly in short service encounters, the clips in the services setting in this experiment were rated as less accented (see a detailed discussion in Section 4.5). However, contrary to her prediction, the friends setting was not among the less accented ones. Admittedly, the friends setting is very broad as the only variable defining it is audience, and there is much potential for variation, for example, in topic, which undoubtedly could have an effect on both production and perception.

There was also an effect of proficiency such that the speakers who I had assigned to the lower proficiency group received a stronger accentedness score compared to the higher proficiency group which lent support to my division of the speakers into two groups. Furthermore, I found a significant interaction between setting and proficiency, such that lower proficiency speakers in the services setting did not receive the advantage that higher proficiency speakers did and were judged more accented in that particular setting. This interaction is plotted in Figure 4.3. It suggests that it is only the higher proficiency speakers who receive a setting advantage and are judged less accented in the services setting. When the model was re-run with levels of proficiency re-leveled and lower proficiency as the Intercept, no significant main effect of setting was found. This means that lower proficiency speakers were judged similarly foreign-accented irrespective of setting. Although I did not study the lower proficiency speakers' production in Chapter 3, I interpret the result as an indication of no variation by setting in either production or perception for lower proficiency speakers. Drummond (2011) claimed that before NNSs exhibit variation which mirrors variation in the L1 community (that is, sociolinguistic variation), they have to reach a certain level of attainment in the L2 (that is, be of a certain

proficiency level). This claim is supported by Young's (1988) finding that only high proficiency learners converged to their interlocutor and Zając (2015) who argued that phonetic convergence may be conditioned by the feature's stage of acquisition. My interpretation of no significant difference by setting for lower proficiency speakers offers further support to Drummond's (2011) claim.

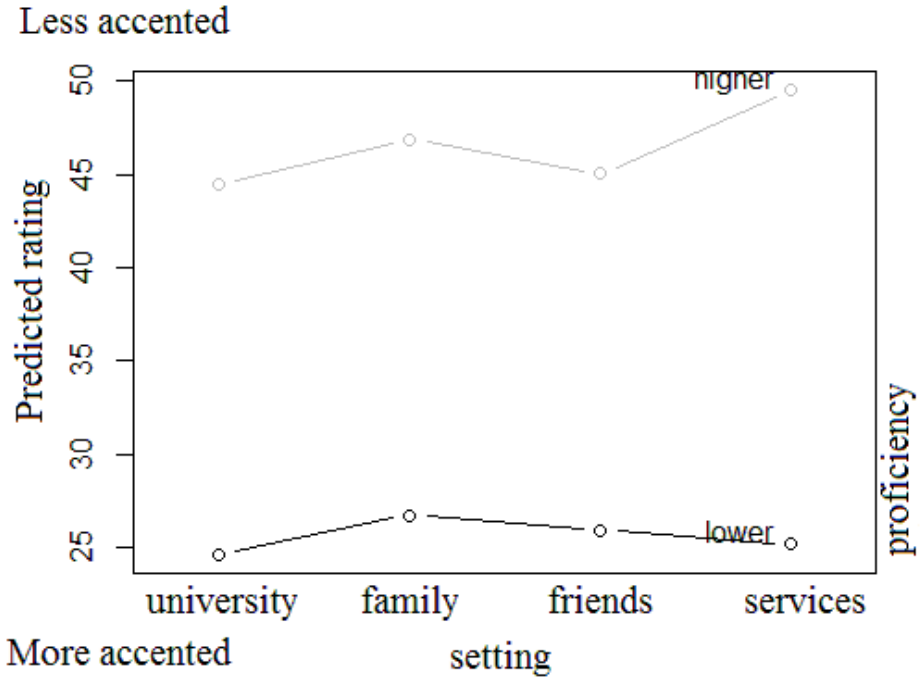

Figure 4.3: Model prediction for accentedness rating in the four settings (from model in Table 4.2)

Finally, there was also a significant effect of progression in the experiment such that the clips presented to the listeners later in the experiment were judged to be more foreign-accented and received a stronger accentedness score. This is compatible with the listeners' getting used to and expecting to hear foreign-accented speech as the experiment progressed.

4.2 Experiment 2: Effect of setting on native-likeness perception and passing for a native speaker

4.2.1 Method

The stimuli used in this experiment were the same 288 clips (24 speakers x 4 settings x 3 extractions) described in Section 4.1.1, plus the stimuli produced by NESs. The listeners were 30 native speakers of New Zealand English, with the age range 18-50, age mean 24, 14 females and 16 males. Fourteen listeners claimed no knowledge of a foreign language. The majority had achieved or were studying towards a Bachelor's degree.

The participants were seated individually (with the exception of two pairs of participants (15 and 16; 29 and 30) who completed the task at the same time on different machines) in a quiet room in front of a computer with E-Prime 2.0 (Schneider et al., 2012). The audio stimuli were presented via head-phones, the instructions – on the screen (see Appendix C); the listeners entered their responses on the keyboard. In the task each listener was presented with 96 clips (24 speakers x 4 settings x 1 extraction) with a break ¼, ½, and ¾ of the way though. The listeners were not warned that they would hear the same speaker more than once. The same speaker did not appear in a quarter more than once, so two clips from the same speaker never appeared next to each other unless separated by a break. They were not asked in the debriefing whether they noticed it, but none commented on it which leads me to believe that, with so many speakers and different topics of recordings, the listeners did not notice that they heard the same speaker more than once. The order of speakers in a quarter, settings, and extractions was randomized. After a clip presentation, the listener was first asked to rate the speaker on an accentedness scale from 1 (Definitely a first language speaker of English) to 7 (Definitely a second language speaker of English). Next, they were asked to name the country or region where they thought the speaker was from. Lastly, they were asked to explain why they had responded the way they did to the two previous questions. The task was self-paced, took about one hour to complete, and was followed by a biographical questionnaire (see Perception study Questionnaire in Appendix A). They received a $10 coffee voucher for their time. The research was reviewed and approved by the University of Canterbury Human Ethics Committee. This section focuses on the setting

variation in listeners' responses to the first question (perceived accentedness) and the second question (geographical origin); Chapter 5 discusses the responses to the second and third questions further.

4.2.2 Native-likeness ratings

Results and discussion

A mixed-effects regression model was fit to the NNES data obtained in the task described above using R (R Core Team, 2014) with perceived accentedness rating as the dependent variable. As setting was the primary variable of interest, the fixed effects were its two-way interactions with all the other variables and their main effects: *speaker sex*, *speaker L1*, *speaker proficiency*, *listener sex*, *listener age*, *listener L2 knowledge* (binary), *length of clip in words*, *length of clip in seconds*, *rate of speech*, *average log CELEX frequency of CELEX content words in the clip* (Baayen et al., 1995; referred to as *word frequency* below), *progression* in the experiment (1-96; to control for a potential familiarization or fatigue effect). *Speaker*, *listener*, and *clip* were included as random intercepts. *Setting* was introduced as a random slope for *listener*, but the model did not converge and the random slope was excluded (Barr et al. 2013). If an interaction or a main effect did not reach significance, the model was re-run without it, and the older and the newer models were compared with an ANOVA. The better or the simpler model was kept, and the process was repeated with the remaining interactions and main effects. The final model in Table 4.3 contains fixed effects which were significant or improved model fit.

Table 4.3: Model summary for accentedness ratings of NNESs in different settings

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	3.781	0.389	43.0	9.709	0.000	***
setting_family	-0.522	0.221	264.1	-2.359	0.019	*
setting_friends	0.005	0.221	263.5	0.023	0.982	
setting_services	-0.473	0.222	263.0	-2.133	0.034	*
S.sex_M	-0.843	0.433	25.6	-1.947	0.063	
L2_Y	0.208	0.297	38.7	0.698	0.489	

prof_lower	2.143	0.398	17.9	5.384	0.000	***
setting_family:S.sex_M	0.622	0.284	188.4	2.194	0.030	*
setting_friends:S.sex_M	0.170	0.283	187.3	0.600	0.550	
setting_services:S.sex_M	0.430	0.284	187.9	1.514	0.132	
setting_family:L2_Y	-0.305	0.171	1977.0	-1.788	0.074	
setting_friends:L2_Y	-0.440	0.171	2003.0	-2.569	0.010	*
setting_services:L2_Y	-0.377	0.171	1996.0	-2.202	0.028	*

Note. * $p<0.05$; ** $p<0.01$, *** $p<0.001$

The reference level (Intercept) is the clips recorded in the university setting by females of higher proficiency judged by listeners with no L2 knowledge. The estimate rating for this level is 3.781. Their counterparts in the family or services setting were judged significantly less accented. This finding partially supports Piller's (2002) claims about different settings and is aligned with the first experiment's results which investigated accentedness with a slightly different method (Section 4.1; see general discussion in Section 4.5). Another statistically significant main effect was speaker proficiency. Unsurprisingly, speakers of lower proficiency were judged as more accented which shows that NSs of New Zealand English generally agreed with my assignment of the speakers to the two proficiency groups.

Speaker sex did not reach significance at the level of $p=0.05$; however, the trend was in the direction of male speakers being rated less accented. These results are reminiscent of the finding by Kraut & Wulff (2013) of some groups of female NNESs receiving a higher accentedness score compared to male speakers of the same proficiency. Speaker sex also participated in a significant interaction with setting such that male speakers in the family setting were judged significantly more accented compared to the university setting in comparison to how less accented women were judged in the family setting compared to the university setting (Figure 4.4). When I re-ran the model with male speakers as the reference level, no significant difference was found between the settings.

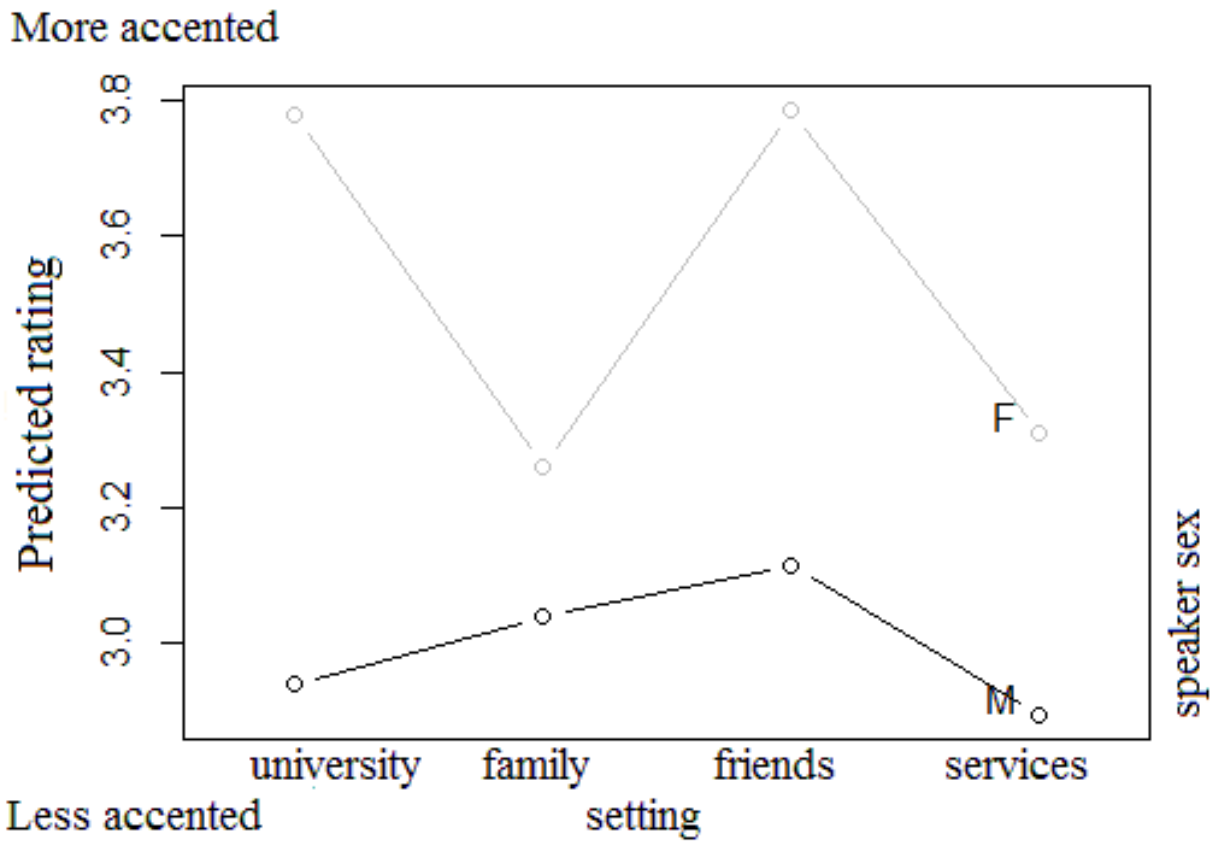

Figure 4.4: Model prediction for accentedness rating of male and female speakers in the four settings (from model in Table 4.3)

On the one hand, this may be reflective of variation in speaker speech by setting; on the other, listeners may react differently to male and female speakers in different settings. The production study (Chapter 3) did not find variation by speaker sex in the production of monophthongal vowels, but only a thorough investigation of male and female production of other features would be able to ascertain that. Although future research will be needed to explore this further, I can tentatively suggest that listeners perceive a different degree of accent when men and women speak on different topics. Psychology literature has shown that different sexes elicit a different expectation of expertise and women are often perceived to be less knowledgeable in a male-associated task (Thomas-Hunt & Phillips, 2004 and references therein). As will be discussed in the next section, the family and university settings differed in formality and technicality of the vocabulary used. My data were not gathered to test this hypothesis specifically, but I tentatively suggest that scientific or professional vocabulary produced in a

female voice was perceived to be more 'foreign' than that produced in a male voice because of listener expectation of male expertise in a professional environment.

The main effect of listener L2 knowledge did not reach significance, but it participated in an interaction with setting such that listeners with L2 knowledge judged clips in the friends and services setting as less accented compared to the university setting (Figure 4.5). The interaction with the family setting did not reach significance at $p=0.05$ level, but the trend was in the same direction ($p<0.1$). This experiment was not designed to explain this finding, but it is possible that listeners with L2 knowledge use cues that are different from those used by listeners with no L2 knowledge. Previous research has found that listeners who rarely interacted with NNESs perceived a stronger accent in an accentedness rating task (Kraut & Wulff, 2013). My finding may be reflective of a similar influence as I expect that listeners with L2 knowledge are more likely to interact with NNESs on a daily basis; however, I did not collect the listeners' social network information and cannot be certain of that. Listener experience with different linguistic varieties through exposure to accented speech or additional language learning can make him/her more tolerant of deviations. Both of these effects, listener L2 knowledge and social networks, are in line with usage-based models that predict that exposure to different exemplars affects speech production and perception (Pierrehumbert, 2003).

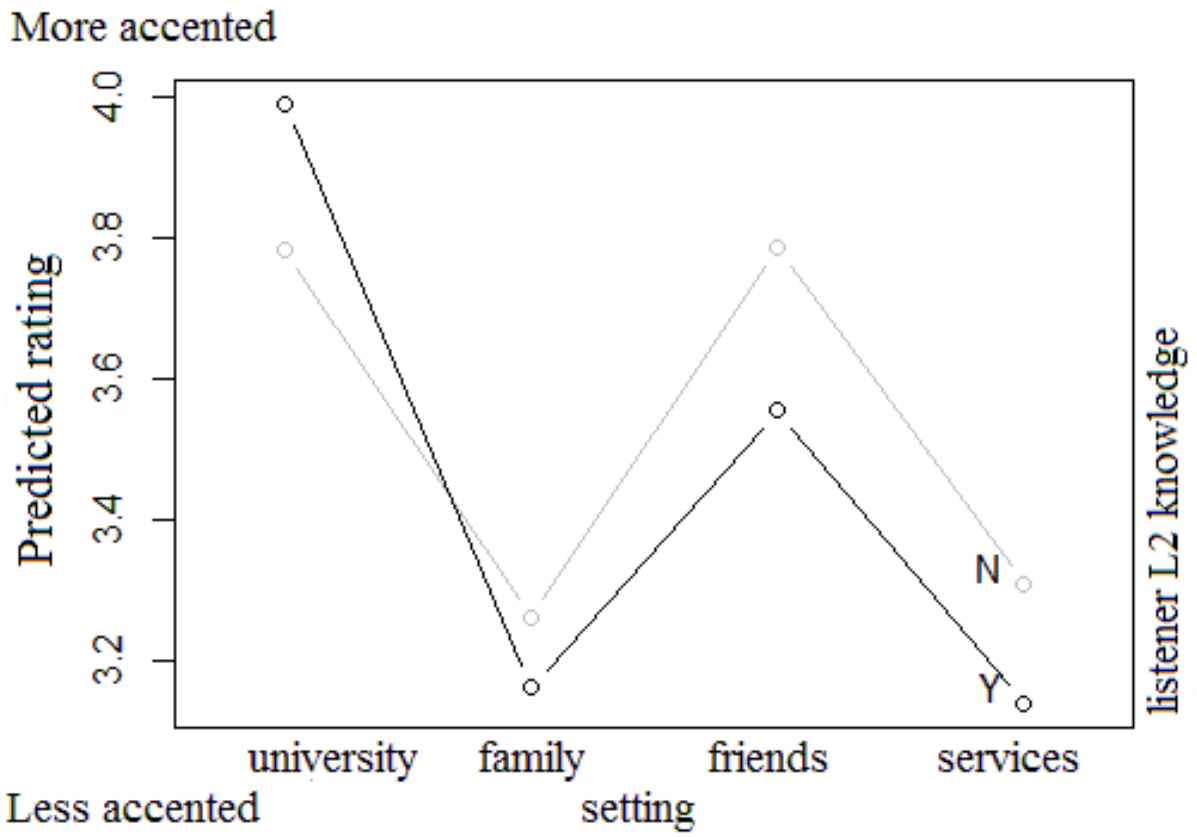

Figure 4.5: Model prediction for accentedness rating of listeners with and without L2 knowledge in the four settings (from model in Table 4.3)

4.2.3 Passing for a native speaker

Data analysis

The participants' answers to the second question about identifying the origin of the speakers in Experiment 2 were systematized: spelling mistakes/typing errors, such as 'Gertmany' for 'Germany' or 'Inida' for 'India' were corrected; if a participant offered two responses such as 'New Zealand or Australia', the first one was recorded as it was assumed to be the first reaction; 'not sure' and the like were equalled to 'no response'. Next, the answers were categorized by region: Africa, Asia, Australia, Europe, North America, New Zealand, Pacific islands, South America, and UK & Ireland. All the cities, countries, or broader areas, such as 'Eastern Europe', which are geographically located within those regions were assigned to the categories. Answers with ethnicity details were subsumed under the country category so that 'African American' or

'Asian American' were added to the North America category and Māori NZ to the New Zealand one. 'North America' included Canada and the USA while Mexico was added to the 'South America' category as the main distinction in the perception of the accent from those countries that is relevant to this study is native-likeness. 'Europe' included continental Europe and Russia. 'UK & Ireland' included Great Britain and Ireland. If the response was too ambiguous to be placed unequivocally, such as 'northern hemisphere' or 'western country', it was added to the 'no response' category.

For the purposes of this study, Africa, Asia, Europe, Pacific islands, and South America were assigned the label NNES; Australia, North America, and UK & Ireland - the label NS of a different dialect; and New Zealand – NS of the same dialect. South Africa was subsumed under 'Africa'.[5] Countries of the outer circle (Kachru, 1992), where English has an official status, such as Singapore, Hong Kong, or India, were not given a separate category. For the purpose of this analysis, only assignments to countries within the inner circle (Australia, NZ, North America, UK & Ireland) were considered examples of passing for a native English speaker.

Results and discussion

To address the matter of variation in passing in different situations experimentally, the four different settings were compared in terms of the passing performance. The categories of passing for a NS of the same dialect, passing for a NS of a different dialect, not passing, and no response for all the NNESs are plotted in Figure 4.6, which suggests that the speakers passed for a NS most often in the services setting and least often in the university setting.

[5] South Africa has 11 official languages including English, so it is impossible to know whether the listeners meant that the speaker is a NNES or a NES of a different variety. However, as under 10% of South Africans speak English as a first language (Statistics South Africa, 2012), I add it to the NNES category.

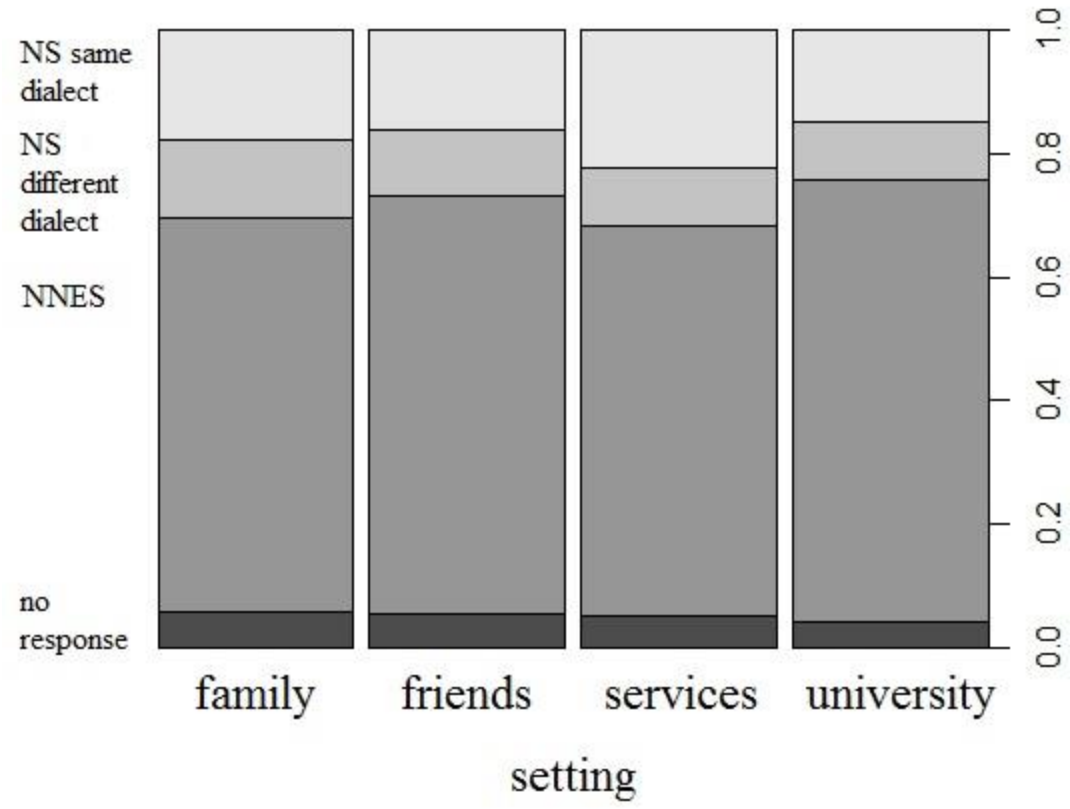

Figure 4.6: Passing in different settings

Fitting a statistical model to these data allows us to check the observed difference for significance while controlling for other factors. To do this, the 'no response' data were excluded, and a binomial mixed effects model was fit to the data with passing (inclusive of NS of the same or different dialect; following Abrahamsson & Hyltenstam, 2009) vs not passing as the binary dependent variable. Fixed effects were of three types: those pertaining to the individual clips, the speakers, and the listeners. The independent variables pertaining to the individual clip included *setting of recording*, *length of clip in words* (Nwords), *length of clip in seconds*, *rate of speech* (words per second), *mean log CELEX frequency of CELEX content words in the clip* (Baayen, Piepenbrock, & Gulikers, 1995), and *progression* through the experiment (from 1 to 96; to control for a potential familiarization or fatigue effect). The independent variables pertaining to the speakers were *sex*, *L1*, *proficiency*, *age of acquisition* (age at which they first lived in an English-speaking country for a minimum of 6 months), and *length of residence* in an English-

speaking country (rounded up to the next full year). As for the listeners, the independent variables were *age* and *sex*.

Additionally, because the setting variable was of the main interest for this study, two-way interactions between setting and all other variables were included into the model as fixed effects. A significant interaction between setting and another variable would mean that there is variation in how different settings affect passing performance in conjunction with other variables. Speaker, listener, and individual clip were introduced as random effects, and setting was introduced as a random slope for listener (Barr et al. 2013). If a fixed effect or an interaction was found non-significant, the model was re-run without it and the new and the previous models were compared with an ANOVA. The significantly better or simpler model was kept, and the pruning cycle was repeated.

Table 4.4 represents the final model. The university setting for higher proficiency speakers was chosen as the reference level (the Intercept). The estimate column in the table represents the log odds of the dependent variable being one factor rather than the other. Positive values in the column mean a higher chance of passing under a certain condition while negative values mean a lower chance of passing. For example, the estimate for the lower proficiency speakers in the university setting is negative at -3.955 and is significantly different from the baseline, as indicated in the Significance column, which means that, unsurprisingly, lower proficiency speakers were less likely to pass for a native speaker, and as the interaction between proficiency and setting was not significant, it was uniformly so across the settings.

Table 4.4: Model summary for accentedness ratings of NNESs in different settings

	Estimate	Standard error	z value	Pr(>|z|)	Significance
(Intercept)	-0.525	0.668	-0.787	0.431	
proficiency_lower	-3.955	0.912	-4.337	0.000	***
setting_family	1.010	0.386	2.616	0.009	**
setting_friends	0.436	0.373	1.170	0.242	
setting_services	0.739	0.402	1.840	0.066	
Nwords	0.676	0.356	1.898	0.058	
progression	0.311	0.152	2.041	0.041	*
settingfam:Nwords	-0.962	0.457	-2.108	0.035	*

settingfr:Nwords	-0.145	0.509	-0.285	0.775	
settingser:Nwords	-0.884	0.397	-2.229	0.026	*
setting_family:progression	-0.481	0.212	-2.272	0.023	*
setting_friends:progression	-0.223	0.210	-1.063	0.288	
setting_services:progression	-0.133	0.207	-0.643	0.521	

Note. * p<0.05; ** p<0.01, *** p<0.001

Setting was found to be a significant predictor so that clips recorded in the family setting were more likely to be judged native-like than the ones in the university setting. It may be that being recorded in a comfortable environment speaking informally on a familiar topic created favorable conditions for passing. The services and friends settings were predicted to be conducive to passing by Piller (2002), but the friends setting was not significantly different from the university setting at all, and the services setting only exhibited a trend (p=0.066) towards being more helpful to passing. This offers only weak experimental support to the claims made in self-reports about variation in passing for a NS (in Piller, 2002 and Section 5.2).

There was a trend for longer clips in the university setting being conducive to passing for a native speaker more often (p=0.058); however, it participated in a significant interaction and longer clips in the family and services settings were less likely to pass for a native speaker. Progression into the perceptual test was significant such that clips further along in the experiment were more likely to be judged native-like, possibly as the listeners relaxed their criteria after hearing more and more different accents from several varieties. It also participated in a significant interaction such that the clips in the family setting further along in the experiment were significantly less likely to be judged native-like. This means that the setting effect for the family setting was waning with more exposure to accented speech.

To sum up, as suggested by some previous literature, the perception experiment revealed within-speaker variation in passing by setting. However, Piller (2002) suggested that the services setting would be favorable to passing, but in the experiment the services setting only exhibited a trend towards being more native-like. Additionally, Piller (2002) argued that communication with friends can exert a positive effect, yet this setting did not reach significance here. The family setting was the only one to reach significance. I leave it to future research to fully corroborate or refute Piller's (2002) claims.

4.3 Relationships between experimental tasks

This section compares listeners' behavior in Experiments 1 and 2 and discusses the differences and similarities in participants' responses.

Figure 4.7 represents the accentedness ratings the 18 NNESs received from the listeners in Experiment 1. The two speakers with the highest score are higher proficiency speakers Kahui and Sam, with the mean ratings of 76.3 and 84.1 respectively. The lower proficiency speaker Vincent stands out as well, with a low mean rating of 12.6. The remaining speakers fall in between these two extremes. In Figure 4.8 we can see the ratings for all the 24 NESs and NNESs in Experiment 2. Despite receiving a few higher judgments (less native-like), the median rating for all native English speakers was 1 (Definitely a first language speaker of English). The median ratings of non-native English speakers varied dramatically from 1 for Kahui and Sam to 7 for Vincent with the majority of rating medians falling between 3 and 6. The use of the whole continua by the participants in accentedness judgment tasks in Experiments 1 and 2 suggests that the speakers were able to hear a degree of accent in the presented clips.

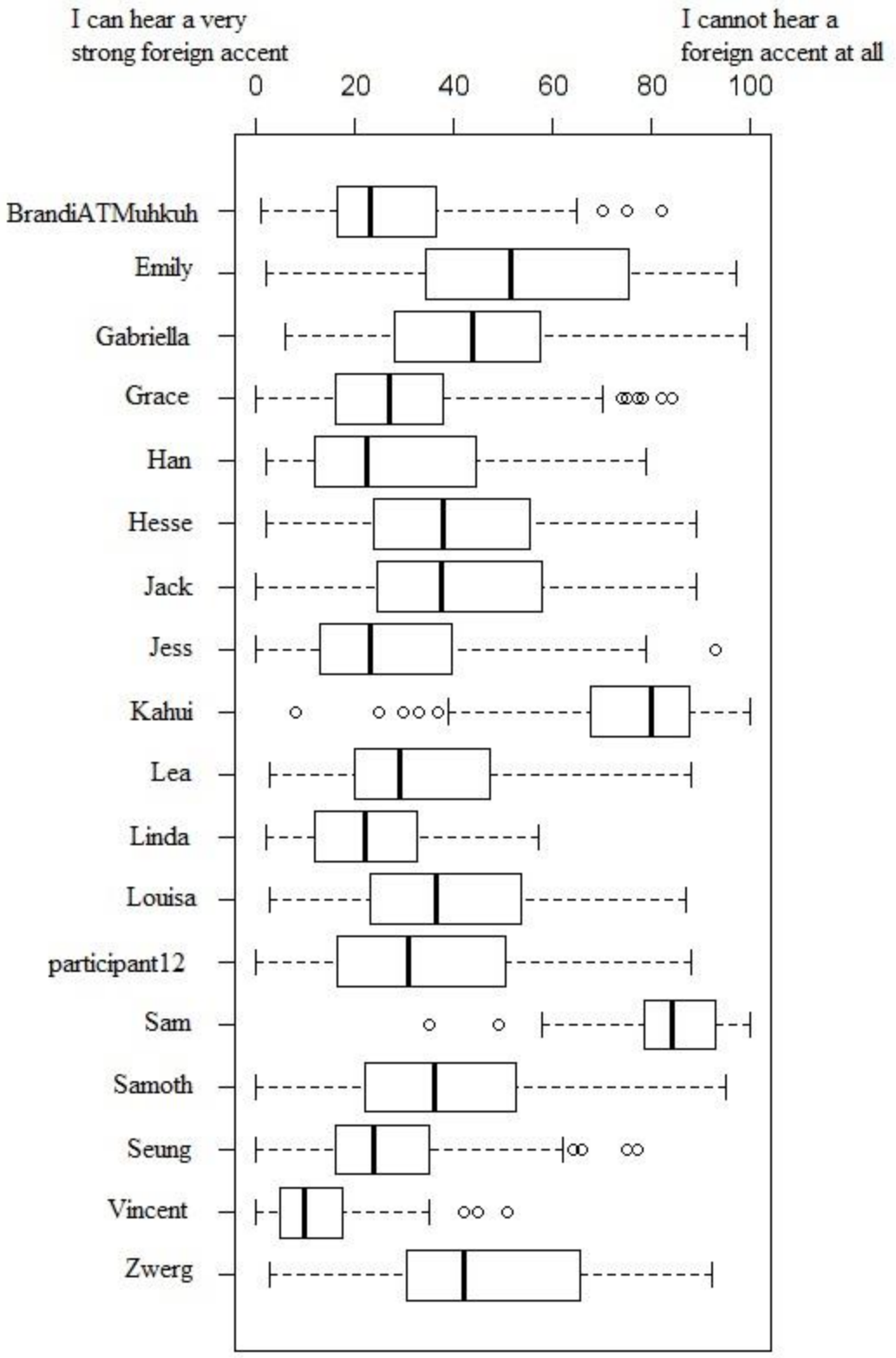

Figure 4.7: The NNESs' accentedness ratings in Experiment 1

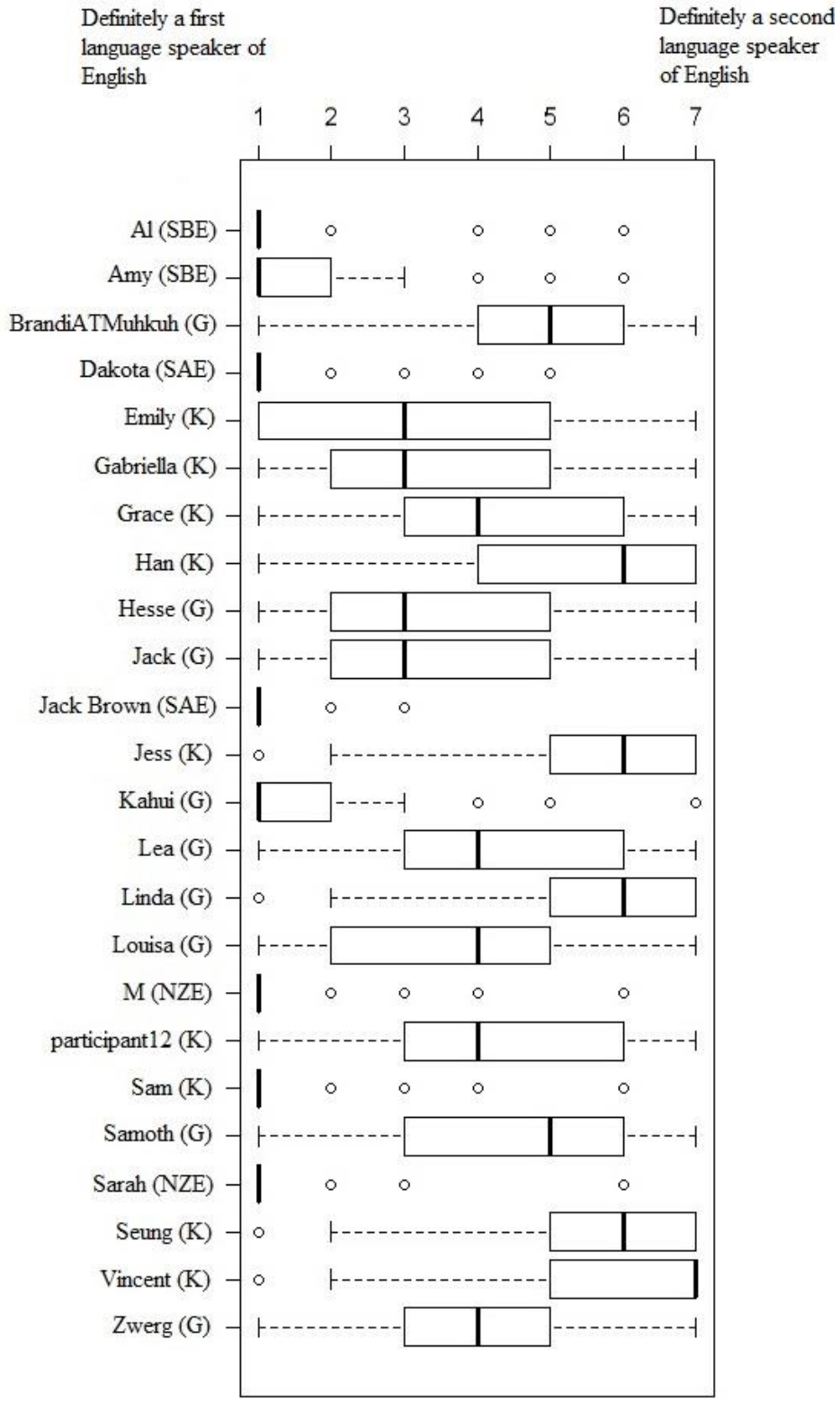

Figure 4.8: The speakers' native-likeness ratings in Experiment 2

The research questions that the two tasks were addressing were similar but slightly different. Experiment 1 asked about foreign accentedness, so, presumably, the listeners were comparing the speakers against an internal ideal of NZE. The NESs were not used for stimuli in this experiment, but it would be expected that NESs of NZE would receive a higher score than other NESs. In Experiment 2 the focus was on native-likeness, which is a much broader concept. It was specifically clarified in the instructions that it included first language speakers of any variety of English. As expected, all NESs had the same median of received ratings (1; Definitely a first language speaker of English).

The listeners' responses in the two experiments were directly comparable. The mean ratings NNESs received in Experiment 1 were predictive of the mean ratings they received in Experiment 2 (Figure 4.9). On an individual level, the same two NNESs Kahui and Sam received the less accented score and Vincent received the most accented score in both tasks.

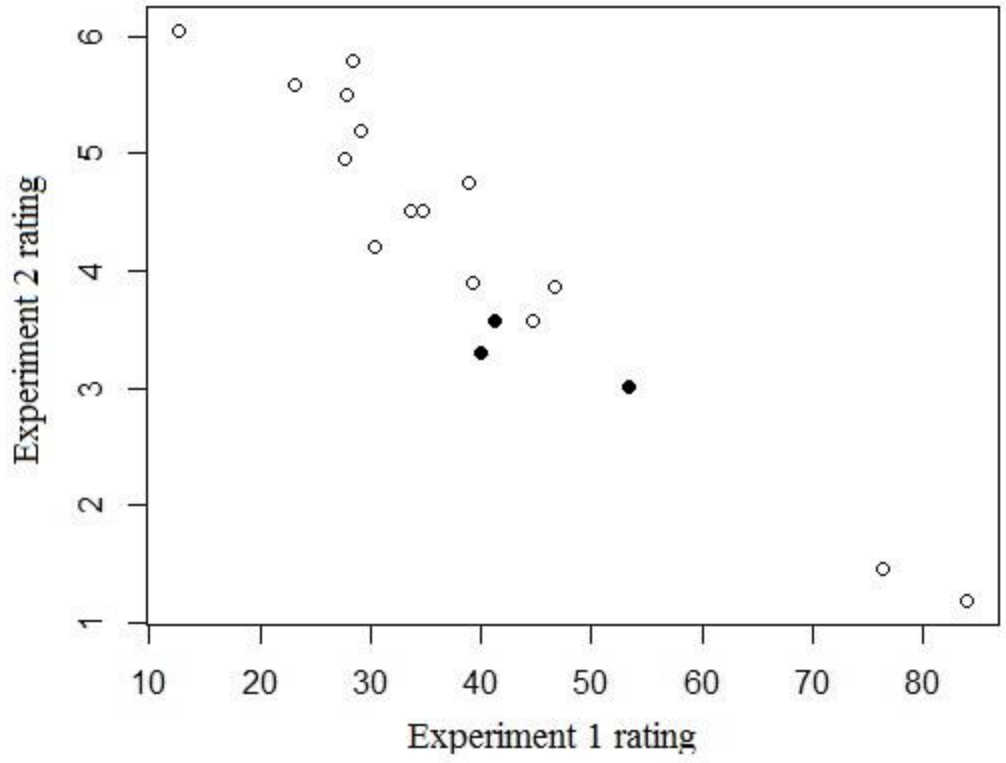

Figure 4.9: Individual NNESs' ratings in the 2 accentedness perception experiments; Emily, Hesse, and Jack filled

I compared the relative accentedness ratings of Emily, Hesse, and Jack in the two experiments to make sure that the listeners actually answered different questions in the two experiments. In the answers to the second question of Experiment 2, Emily, Hesse, and Jack were regarded to be a NNES over 40% of the time; however, their passing performance is quite different: Emily passed for a NS of NZE over 40% of the time and for a NS of other varieties – less than 5%; Hesse and Jack, on the other hand, passed for a NS of another variety over 30% of the time and for a NS of NZE – less than 6% (Figure 4.10). These three speakers would then be expected to behave differently in regards to the accentedness ratings in the two experiments (see Figure 4.9). Emily should receive a higher accentedness score compared to Jack and Hesse in the first experiment where the question asked about foreign accentedness, but they may be expected to receive a similar score on native-likeness in Experiment 2. In fact, Emily, Hesse, and Jack's median native-likeness ratings in Experiment 2 were the same (3; Figure 4.8). Welch's t-test was used for pairwise comparisons of the ratings, and it confirmed that the mean ratings were not statistically different for Emily and Hesse, but they differed for Emily and Jack ($p=0.02$). This means that, when asked about native-likeness, Emily and Hesse were judged similarly native-like and Emily and Jack differently. The same test was used to compare the mean ratings received by these speakers in Experiment 1, and it found a statistical difference for both pairwise comparisons Emily and Hesse ($p=0.00005$) and Emily and Jack ($p=0.0003$). This means that, when asked about foreign accentedness, Emily and Hesse and Emily and Jack were judged differently, and the difference for Emily and Jack in Experiment 1 was larger than that in Experiment 2. So, as expected, the speakers with differences in passing (majority NS of the same variety vs majority NS of a different variety) were rated slightly differently in the two experiments.

Figure 4.10 illustrates the percentage of time the speakers were assigned to different nativeness categories in terms of origin in Experiment 2. It can be seen that NESs all received a majority NES assignment. Kahui and Sam, who were rated most native-like on the accentedness scales, were also believed to be from New Zealand most of the time (see Section 5.1 for details). Emily received a majority New Zealand geographical assignment but got over 40% in the NNES category; this is reflected in her accentedness score, which had a much wider range compared to Kahui and Sam. Naturally, because these two questions were part of the same experiment, dramatically contrasting responses would not be expected. Having two different questions in the

same experiment also allowed me to compare the second language accentedness scale to open-ended responses directly.

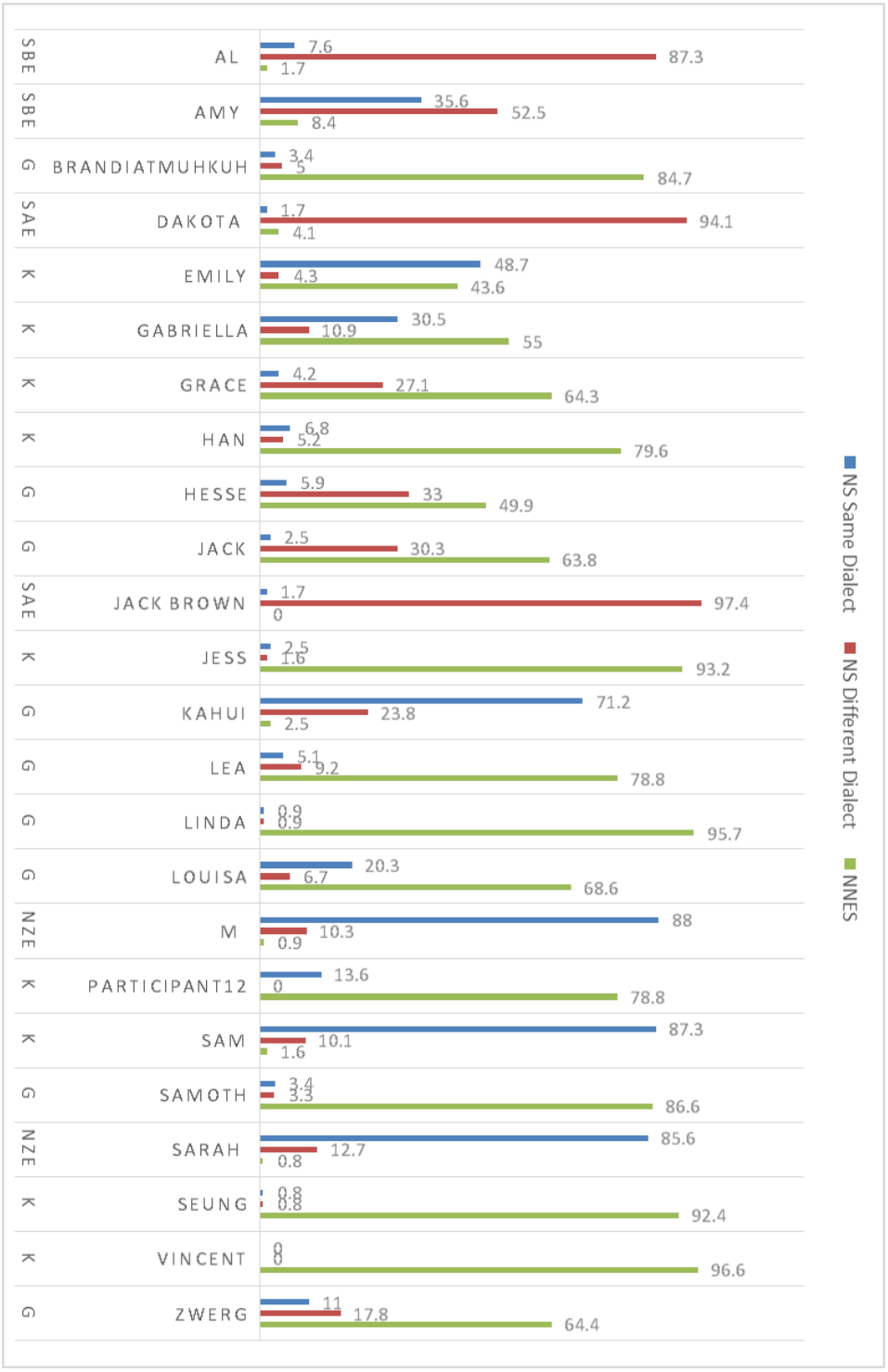

Figure 4.10: Guesses about the speakers' origin in Experiment 2; 'no response' is omitted

Figure 4.11 shows the overall correspondences between the NNESs' nativeness ratings and their geographical assignments in Experiment 2. It can be clearly seen that the speakers who were thought to be from English-speaking regions (Australia, North American, NZ, and UK & Ireland) received accentedness ratings on the 'Definitely an L1 speaker of English' side of the scale (e.g., median rating for Australia is 1). Predominantly non-English-speaking regions, on the other hand, were associated with a higher second language accentedness score (e.g., 5 for Europe). However, one can also notice that this was not a clear-cut correspondence as some English-speaking regions received a few higher accentedness scores (up to 7 for all but Australia) and most non-English-speaking areas were associated with lower accentedness scores at least once. Undoubtedly, some of this variation can be explained through participant error, but it may also reflect the continuous nature of accentedness and native-likeness in perception as listeners take into consideration the speaker's linguistic and social background (e.g., where the speaker was born, what language they speak at home, etc.; see more in Section 5.4).

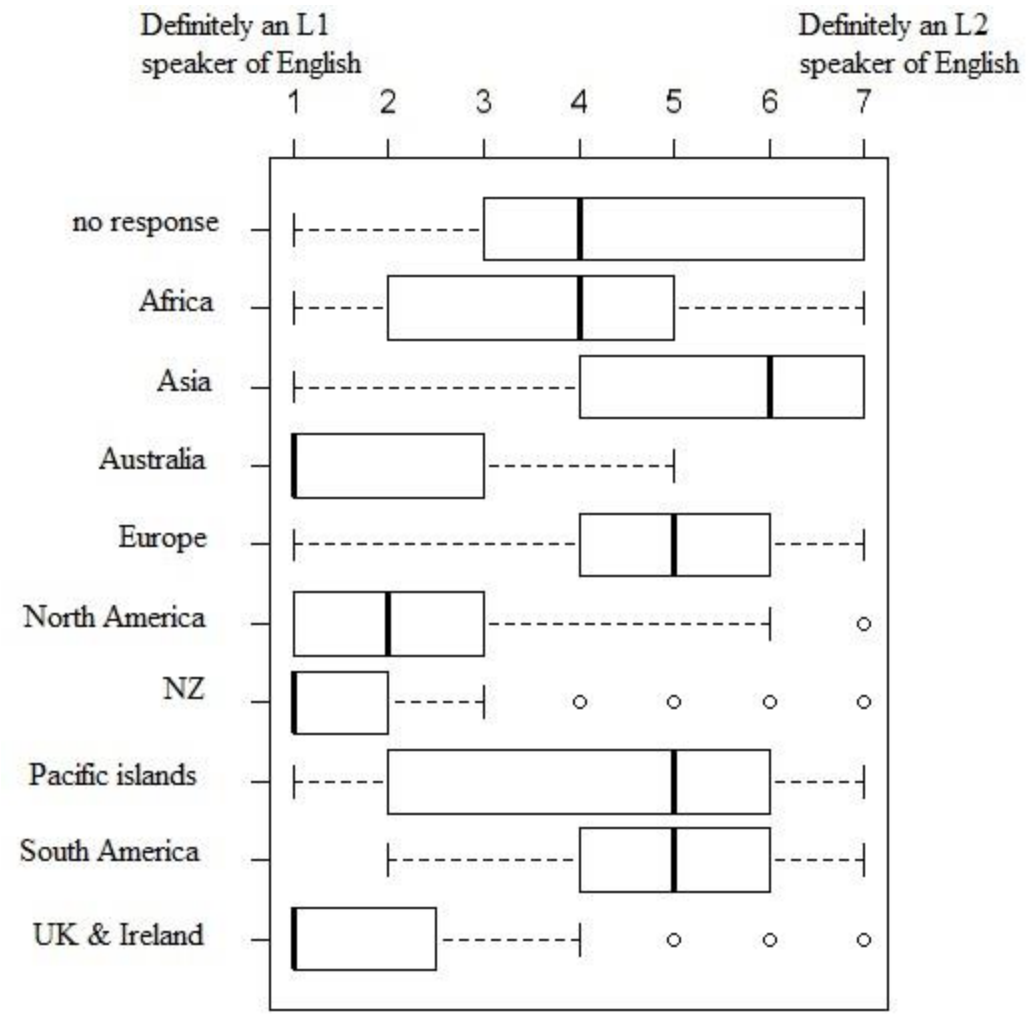

Figure 4.11: The accentedness ratings - region guesses correspondences for NNESs

In their comments (Section 5.4) the listeners indicated that they were not splitting speakers into clear social categories of origin but rather were aware of the mixed nature of accent and its relationship with individual speaker histories. For example, listener 30 in Experiment 2 gave the L1 Korean speaker Gabriella a native-likeness score of 2, judged her to be from New Zealand, and commented that she has a 'Māori or Pacific accent, but sounds kiwi'. The listener probably assessed the presence of target- and non-target-like features in her speech and explained the mixed nature of her accent through sociolinguistic variation. Her accent was attributed to an ethnolect, and its non-standardness resulted in a higher accentedness score, which resembles Beinhoff's (2008) finding that the concept of 'native speaker' is highly connected to 'standard language' and Scottish English, for example, was rated to be less 'native'.

4.4 Experiment 3: Effect of ethnicity

In this section I explore the role of perceived ethnicity in accentedness perception in an experiment involving conditions in which listeners are presented with foreign-accented speech with or without visual input.

4.4.1 Method

Stimuli

As outlined earlier, the 24 native and non-native English speakers were interviewed by me about their university studies in a quiet room at the university (see Section 3.1). To elicit spontaneous speech, the speakers were asked to tell the interviewer about the applications of their research or study field. They were recorded with the use of a lapel Opus 55.18 MKII beyerdynamic microphone and an H4n Zoom audio-recorder and a Sony video-recorder. The speakers were also video-recorded against a plain background with the recorder positioned at their eye-level and the frame including their upper body (see Figure 4.12). The audio tracks recorded by the audio-recorder were used in the experiment as the quality of the recording by the lapel microphone was much better than that of the video-recorder microphone, positioned about two meters away from the speaker. I audio- and video-recorded a clap before commencing the interview for ease of audio and video synchronization. The audio tracks were synchronized with the respective video tracks in Adobe Premiere Pro software, so that listeners heard the same track in both the 'audio only' and the 'audiovisual' conditions (see below). The intensity was scaled to remove variation in volume of the audio-recordings. Short clips of a minimum of 30 words were extracted from the recordings as stimuli. Because stopping the clips mid-phrase could have an effect on the listeners' perception, complete phrases were used and the exact number of words per clip was allowed to vary (mean length in seconds = 15; range = 8-22). The clips did not contain proper nouns.

Figure 4.12: A snapshot from the video track of a Korean (left panel) and a German (right panel) L1 speaking participants

Listeners

The listeners were 45 Caucasian native speakers of New Zealand English who were recruited through announcements posted around the university campus and via the friend of a friend method (Milroy, 1987). 48 people participated in the experiment originally, but three participants were excluded from the analysis as they indicated that they had met one or more of the speakers in the experiment. Of the remaining 45, 27 were females and 18 were males with the mean age of 25.47. The listeners were assigned to one of the three conditions before arriving at the lab: *audio only*, *audiovisual*, and *video only*, – with 15 participants in each (I elaborate on these conditions below). Ten listeners in the audio condition, two in the audiovisual condition, and eleven in the video condition claimed no knowledge of a foreign language.

Procedure

The listeners were seated individually in a quiet lab in front of a computer. Stimuli were presented electronically using the E-Prime 2.0 software (Schneider et al., 2012). The audio stimuli were presented through head-phones; the video stimuli were presented on the computer screen. Before starting the actual task, the listeners read the instructions on the screen (Appendix C), completed a practice trial with a non-linguistic clip which allowed them to adjust the volume

(in the audio only and audiovisual conditions), and if needed, clarified the procedure with the research assistant (the author). After that, the listeners were presented with 24 clips (one from each of the 24 speakers) in random order. In the 'audio only' condition, they were presented with the audio clips with a black screen and a fixation point; in the 'audiovisual' condition, they were presented with both the video and the audio signal; in the 'video only' condition, they saw the video recordings but did not hear anything. In the task, the listeners were instructed to rate the presented clips on a scale which read 'No foreign accent at all' and 'Very strong foreign accent' at the two extremes using number keys 1 through 7. The listeners could not re-play the clips. At the end, the listeners completed a short biographical questionnaire (see Perception study Questionnaire in Appendix A). The task was self-paced and took up to 30 minutes to complete. They were given a $10 coffee voucher for completing the task. The research was reviewed and approved by the University of Canterbury Human Ethics Committee.

4.4.2 Results and discussion

Results

The accentedness ratings of the NNSs were analyzed in R (R Core Team, 2014). A linear mixed-effects model was fit to the NNS data with the perceived accentedness rating as the dependent variable. The full model included an interaction of *condition* and *L1*, as well as *speaker proficiency*, *speaker sex*, *listener age*, *listener sex*, *listener L2 knowledge* (binary), *clip length in seconds*, *clip length in words*, *speech rate* (words per second), *progression* in the experiment (1 through 24; to control for a potential familiarization or fatigue effect) as fixed effects. *Speaker*, nested within *L1* group, and *listener* were included as random intercepts. If a fixed effect was found to be non-significant, I excluded that variable from the model and then compared the models with and without the variable with an ANOVA. The significantly better or simpler model was kept.

Table 4.5 represents the final model. The higher proficiency Korean L1 speakers in the audio condition were chosen as the reference level. For the base level (the higher proficiency Korean L1 speakers in the audio condition), the predicted accentedness rating was 5.388. The higher proficiency Korean L1 speakers received a rating 0.052 higher in the audiovisual condition and 0.156 lower in the video condition than in the audio condition. These differences

were not significant as indicated in the *significance* column. This means that the ratings of Korean L1 speakers between the conditions were not significantly different. In the audio condition German and Korean L1 speakers of higher proficiency were not rated to be significantly different from each other.

Table 4.5: Summary for model of accentedness rating

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	5.388	0.714	20.3	7.542	0.000	-
condition_audiovisual	0.052	0.268	68.4	0.193	0.847	
condition_video	-0.156	0.268	68.4	-0.58	0.564	
L1_German	-0.496	0.349	25	-1.422	0.167	
proficiency_lower	1.130	0.356	17.9	3.171	0.005	**
rate of speech	-0.556	0.251	17.9	-2.219	0.040	*
condition_audiovisual : L1_German	0.511	0.237	747.8	2.16	0.031	*
condition_video : L1_Germany	-0.785	0.237	747.8	-3.318	0.001	***

Note. * p<0.05; ** p<0.01, *** p<0.001

The accentedness ratings of German L1 speakers, however, were significantly different between conditions. They were rated significantly more accented in the audiovisual condition but less accented in the video condition compared to audio only. This interaction is plotted in Figure 4.13. I also found significant effects of proficiency and rate of speech, so that, unsurprisingly, lower proficiency speakers were rated as more accented and the higher the rate of speech the more native-like the accentedness rating.

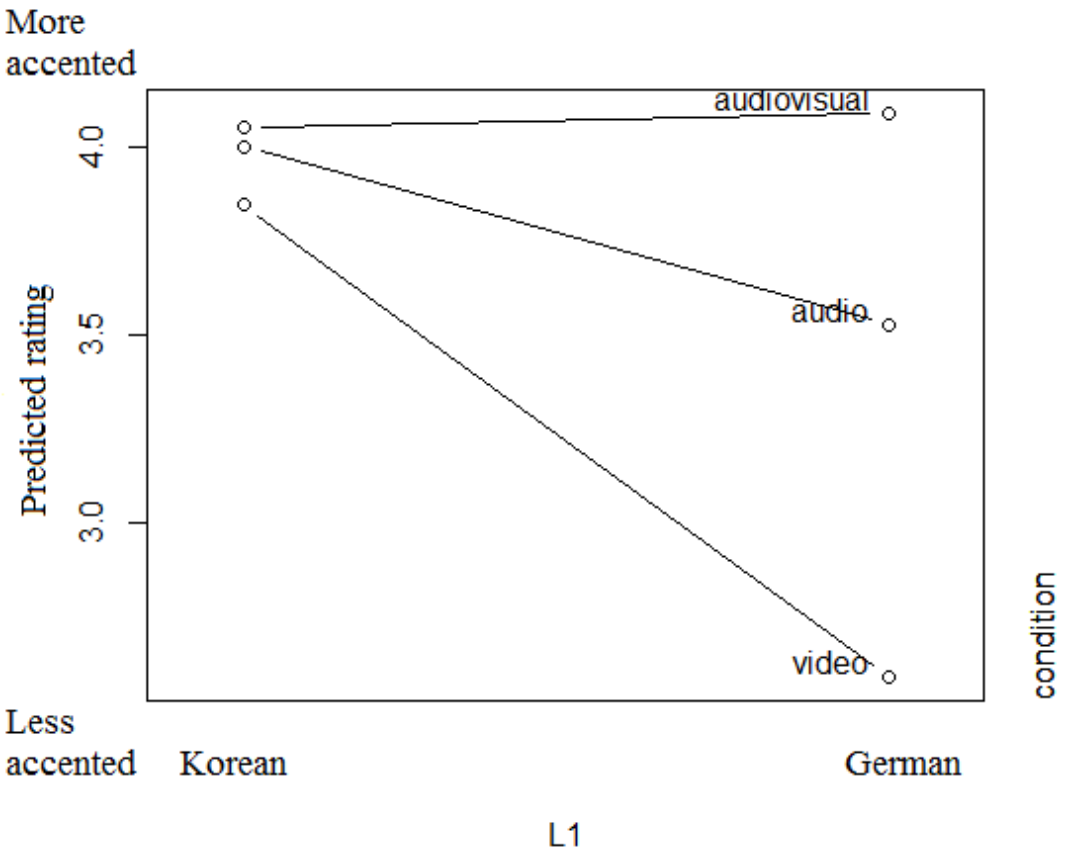

Figure 4.13: Model prediction for accentedness ratings of Korean and German speakers in the three conditions (from model in Table 4.5)

As a group, the L1 German speakers were rated more accented in the audiovisual condition and less accented in the video condition compared to the audio one, supposedly reflecting the difference in listener expectation (based on the video input) and perceived production (see discussion below). Such an effect may be found for individual speakers as well, resulting in a more accented score in the audiovisual condition for speakers with a less accented score in the video condition. To test whether the same NNESs who got a lower score in the video condition also received a higher score in the audiovisual condition compared to audio only, I calculated the mean ratings for each speaker in each condition, then for each speaker subtracted the audio mean from the audiovisual mean, obtaining the individual 'audiovisual enhancement' score, and the video mean from the audio mean, resulting in the individual 'visual accentedness predictability' score. The smaller the audiovisual enhancement score, the more of the visual benefit is found and the less accented the speaker is rated when the visual input is available

compared to when it is not. The larger the visual accentedness predictability score, the more 'accentless' the speaker looks compared to how he or she sounds. For example, German L1 speaker Lea's mean score across all the listeners in the audiovisual condition was 5.80, in the audio condition 5.00, and in the video condition 3.27. Lea's audiovisual enhancement score is 5.80-5.00=0.80, and the visual accentedness predictability score is 5.00-3.27=1.73. The positive audiovisual enhancement score means that Lea is perceived to be more accented in the audiovisual condition than in the audio only one. The positive visual accentedness predictability score means that Lea is perceived to be more accented in the audio condition than in the video only one. Calculated in the same fashion, another German L1 speaker, Linda's, audiovisual enhancement score is 0.93 and visual accentedness predictability score is 1.87. Both of these scores are higher for Linda than for Lea, suggesting that they may be correlated.

To see whether the difference between the audio and the video conditions is predictive of the difference between the audiovisual and audio conditions, I fit a linear regression model to the NNES data with the *audiovisual enhancement score* as the dependent variable and an interaction between *L1* and the *visual accentedness predictability score* as predictors. However, the interaction was not found to be significant and L1 did not improve model fit, so the final model includes only the visual accentedness predictability score as an independent variable. In Table 4.6 we can see that there is a significant positive correlation between the audiovisual enhancement score and the visual accentedness predictability score such that the less accented a speaker was rated in the video condition compared to the audio condition the more accented that speaker was rated in the audiovisual condition compared to the audio condition. In other words, the less accented a speaker looks, the more accented he/she is perceived to be when the video input is available compared to when it is not. This relationship is represented in Figure 4.14.

Table 4.6: Summary for model of the audiovisual enhancement score (audiovisual - audio) in accentedness ratings

	Estimate	Standard error	t value	Pr(>|t|)	Significance
(Intercept)	0.204	0.103	1.983	0.065	-
visual accentedness predictability score (audio – video)	0.189	0.072	2.637	0.018	*

Note. * p<0.05; ** p<0.01, *** p<0.001

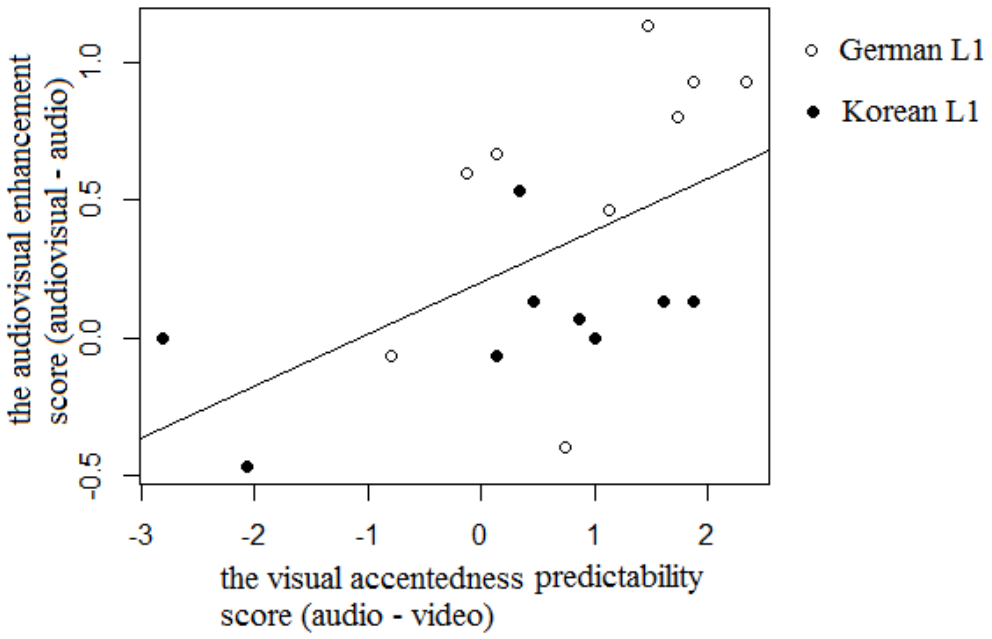

Figure 4.14: Model prediction for the relationship between the audiovisual enhancement score and the visual accentedness predictability score (from model in Table 4.6)

I ran a second linear mixed-effects model to explore the difference in accentedness ratings between German and NZE L1 speakers in the video condition to check whether the Caucasian non-native speakers were rated more accented than the NSs based on visual cues only. The model was fit to the German and NZE L1 speaker video and audio conditions data with the perceived accentedness rating as the dependent variable. The full model included an interaction of *condition* and *L1*, as well as *listener age* and *sex*, *speaker sex*, *clip length in seconds*, *clip length in words*, *speech rate* (words per second), and *progression* in the experiment (to control for a potential familiarization or fatigue effect) as fixed effects. *Listener* and *speaker* were included as random intercepts. If a fixed effect was found to be non-significant, I excluded that variable from the model and then compared the models with and without the variable with an ANOVA. The significantly better or simpler model was kept.

Table 4.7 represents the final model. The model illustrates that in the video condition, there was no significant difference between the two language groups. NZE L1 speakers were rated significantly less foreign accented in the audio condition compared with the video condition, and German L1 speakers in the audio condition were judged to be significantly more

accented than NZE L1 speakers, which is not surprising. This suggests that the listeners were not able to infer the foreign accent based on the video input only. There were also significant effects of clip length and progression in experiment with listeners tending to give higher accentedness scores to longer clips and those later in the experiment.

Table 4.7: Model summary for accentedness ratings of German and New Zealand English first language speakers

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-0.041	0.677	12.63	-0.061	0.953	-
L1_German	-0.169	0.286	10.62	-0.592	0.567	
condition_audio	-2.129	0.685	12.46	-3.106	0.009	**
progression	0.023	0.011	296.93	2.09	0.038	*
length in sec.	0.188	0.038	10.74	5.004	0.000	***
L1_German : condition_audio	3.063	0.733	11.02	4.178	0.002	**

Note. * p<0.05; ** p<0.01, *** p<0.001

Discussion

- What is the effect of availability of visual information for Asian NNESs in an accentedness perception task?

The accentedness ratings of Korean L1 speakers in the audio condition were not significantly different from the other two conditions, which is different from the findings of Yi et al. (2013). No difference between the audio and the video conditions suggests that the degree of accentedness that the listeners heard in the audio condition was similar to the degree of accentedness they expected to hear from the Asian speakers in the video only condition, this expectation created by 'iconic associations between language, nationality, and race such that each category can effectively stand in for the others' (Shuck, 2006 as cited in Subtirelu, 2015). When the video and audio inputs were congruent in the audiovisual condition, as per listeners' expectations, there was no additional effect of ethnicity and the rating in the audiovisual condition was not significantly different from audio only. In other words, experience-based

representations with similar accentedness information attached to them were activated in the three different conditions. The negative bias hypothesis, as interpreted by Yi et al. (2013), was not supported as Korean L1 speakers were not rated significantly more accented in the audiovisual condition compared to the audio one. The effect found by Yi et al. (2013) may be due to the experimental design in which listeners were presented with the same sentence and same speakers multiple times.

- What is the effect of availability of visual information for Caucasian NNESs in the same accentedness perception task?

Based on the results from previous studies, I have predicted that the ratings of German L1 speakers in the audiovisual condition would be either higher or lower than those in the audio condition. The results show that the audiovisual ratings were higher in accentedness than the audio only ones. This suggests that reverse linguistic stereotyping did not play the leading role here as a lower accentedness score would be expected, but, as the video and audio inputs were unexpected (a Caucasian speaking with a foreign accent), that could have constituted an audiovisual mismatch effect. The listeners were not expecting to hear a foreign accent when they saw a Caucasian speaker, but when they did, the accent stood out even more, resulting in a higher accentedness score. This interpretation is supported by the significant positive correlation between the difference in the ratings between the audiovisual and audio conditions and between the audio and the video conditions. This means that the more 'accentless' the speaker looked and was rated in the video condition compared to the audio condition, the more there was of a mismatch effect and the accent 'stood out' to the listeners in the audiovisual condition compared to the audio only one.

In accordance with other accounts of reverse linguistic stereotyping (Rubin, 1992; Yi et al., 2013), German L1 speakers in the video only condition were rated significantly less accented when the listeners could not hear the speakers as in the audio condition when the accent was actually heard. Moreover, no significant difference was found in the ratings of German and NZE L1 speakers in the video condition, which means that the listeners could not tell the difference between Caucasian L1 and L2 speakers of NZE based on the video input only. This suggests that the German L1 speakers' physical appearance, clothes, gesturing, and other visual cues were not

different or salient enough in comparison with their NZE counterparts' to notice their non-native status.

- Will these effects for Asian and Caucasian NNESs be better predicted by reverse linguistic stereotyping or an audiovisual mismatch?

To sum up, I found that Asian NNESs received similar foreign accentedness ratings in the audio and audiovisual conditions while Caucasian NNESs received a higher rating in the audiovisual condition as in line with the predictions of the audiovisual mismatch effect but contradicting the reverse linguistic stereotyping account. These findings are reminiscent of the role of socioindexical expectation described in McGowan (2011), who found that listeners were better at transcribing Chinese-accented speech in noise when presented with a Chinese photograph than a Caucasian one or a silhouette. This and my findings suggest that reverse linguistic stereotyping may not be the only explanation for an ethnicity effect, but rather a perceived alignment between the audio and the video inputs may have a facilitatory effect while perceived mismatch or misalignment may result in inhibition as the visual and the audio input may be activating conflicting experience-based representations, as in the aforementioned finding of Hay, Warren et al. (2006) for L1 speakers, in which listeners were more likely to make errors in vowel identification when there was a mismatch between actual production and their expected production (as per assumed social class of the speaker). When listeners see an Asian speaker, 'accented' representations are more likely to be activated, and hearing accented speech reinforces their activation facilitating easier access and retrieval. However, when listeners see a Caucasian speaker, 'accentless' representations are more likely to be activated, but hearing accented speech activates other representations spreading overall activation thinner and inhibiting access and retrieval.

4.5 General discussion: Variation in perceived accentedness

A number of variables were tested in the three accentedness perception experiments described above. The experiments differed in their methodology and immediate research questions at hand (see Table 4.8). By way of reminder, Experiments 1 and 2 had setting as their main variable of interest. In Experiment 1 the listeners were presented with four clips at a time (one clip from

each setting for NNESs only) and were asked to place them on a foreign accentedness scale. In Experiment 2 the participants listened to one clip at a time with four clips from an individual speaker (the four settings from NESs and NNESs) and rated them on a second language accentedness scale. Experiment 3 investigated the effect of condition (audio, video, audiovisual) in a foreign accentedness rating task in which the listeners were randomly presented with one clip per speaker from NESs and NNESs. This section aims to consolidate the results from the three experiments.

Table 4.8: Details about the three experiments

	Experiment 1	Experiment 2	Experiment 3
Dependent variable	Foreign accentedness rating	Second language accentedness rating	Foreign accentedness rating
Main independent variable of interest	Setting (family, friends, services, university)	Setting (family, friends, services, university)	Condition (audio, video, audiovisual)
Stimulus presentation	4 at a time	1 at a time	1 at a time
Number of clips from the same speaker	4	4	1
Speakers included	NNESs	NESs and NNESs	NESs and NNESs

Despite these methodological differences between the three experiments, a number of similar trends emerged in the results. First, the effect of proficiency was found to be significant in all three experiments, and the speakers who I had assigned to the lower proficiency group were rated significantly more accented in all three experiments. This supports my division of the NNESs into two groups by proficiency.

Setting, which was the main independent variable of interest in Experiments 1 and 2, was found to be significant as a main effect and in interactions. The results of this quantitative study support some of the Piller's (2002) claims about passing based on qualitative data. According to her, NNESs are more likely to pass for a native speaker in short service encounters or in communication with friends. If extrapolated to the four settings in my experiments, this may predict a less accented rating in the friends and services settings. Both experiments found that the clips in the services setting were judged significantly less accented compared to the university setting (for both males and females in the 1st experiment and for females only in the 2nd

experiment). Additionally, the family setting was found to be rated significantly less accented than the university setting for female speakers in Experiment 2; it failed to reach significance in Experiment 1, but the trend was in the same direction with $p<0.1$. My findings support Piller's claims about the services setting; however, the friends setting was not among the less accented ones. I acknowledge that the clips in the friends setting varied dramatically in terms of topic and conversation flow, so it is problematic to make such a generalization about all communication with friends. What the participants in Piller (2002) might be noticing is their increased confidence in communication with people and topics that they are familiar and comfortable with. Under this interpretation my speakers' communication with the interviewer in the family setting may fit this criterion despite it being an interview: the speakers were speaking on an informal topic to someone they knew as part of their social circle in a comfortable environment (their home).

As noted above, communication is a joint performance of the speaker and the listener, so the setting effect in perception may be reflective of variation in production or may be limited to perception. Here I focus on the explanations for the difference between the university and family settings and university and services settings from the perception angle. The obvious difference between the family and services settings and the university setting is the topic, which was more professional, dealing with research and innovation and requiring the speakers to use more technical language, terms, and jargon in the university setting while in the family and services settings the topic was more mundane and the language was less industry-specific as in the following examples.

(1) *all cellulose composites are monocomposites and a monocomposite just um means that the fibre and the matrix are made of the same material which means they're chemically very very similar if not identical* (Jack. University interview. Perception clip 2)

(2) *I would say that I'm closest to my mom because she's the one I I can have the most personal um discussions or conversations with* (Jack. Family interview. Perception clip 2)

(3) *er no er the light is is fine I only need one but I need the bulb and the bulb that fits in there wasn't there do you have any other shelf or* (Jack. Services self-recording. Perception clip 3)

The topic of the university setting is less familiar and more 'foreign' to a non-specialized audience which could lead to a stronger accentedness perception. If we assume a usage-based

account of accentedness rating, we presuppose that listeners compare the accent in the clip to an ideal representation of a 'native accent' based on their multiple experiences with other NSs. A more technical topic would be more novel than a family-related account resulting in stronger accentedness ratings. In a similar fashion, previous research has found an effect of word frequency on accentedness ratings such that the higher the word frequency the more accented the speaker was rated (Levi et al., 2007). Applying the word frequency hypothesis to the two interviews in this experiment, one could hypothesize that the university setting clips with their technicalities could produce a similar effect and attract higher accentedness ratings. For that reason, the mean word frequency in the clips was calculated and entered into the model. The effect did not reach significance or improve model fit and was dropped from the final model, but it is possible that this measure did not capture the word frequency effect well and a different one may prove a better prediction.

Furthermore, Experiment 1 found a significant interaction between setting and proficiency such that lower proficiency speakers in the services setting were not rated significantly less accented compared to the university setting. As opposed to Experiment 1, Experiment 2 revealed significant interactions of setting with listener L2 knowledge and speaker sex. The different interactions in the two experiments are possibly due to differences in methods used. In Experiment 2 the listeners were presented with one clip at a time, so four clips from the same speaker were rated independently exhibiting an effect of setting in perception; in Experiment 1, however, the four clips were presented at the same time, and a listener could make direct comparisons between the settings. These differences may be a task effect. On the one hand, phonetic perception may vary in different environment as listeners have been shown to behave differently in discrimination and identification perception tasks. For example, Waylard (2007) found discriminability of non-native contrasts (Korean and Thai stop consonants) was better predicted by identification than discrimination data. On the other hand, it is possible that the accentedness ratings in Experiment 1 are more representative of the variation inherent in the clips than Experiment 2 where listener-dependent factors may play a larger role. Therefore, the significant interactions with speaker sex and L2 knowledge emerge when the listeners do not realize that they listen to the same speaker more than once and their stereotypes and expectations play a larger role in the assignment of accentedness ratings.

Progression was found to have a significant effect on perceived accentedness in Experiment 1. The listeners were more likely to perceive a stronger foreign accent closer to the end of the experiment. I argued that this effect is due to the listeners' expectation to hear accented speech. Experiment 1 was the only experiment of the three which did not employ NESs as controls, so the listeners may have noticed that the range of accentedness they heard was from light to strong and realized that the experiment included NNESs only. This may have brought them to expect to hear foreign accented speech and rate it in accordance with their expectations.

The effect of ethnicity was investigated in Experiment 3 which included an audio, video, and audiovisual condition. The significant condition by ethnicity interaction suggests that Asian speakers received a similar accentedness score regardless of whether the listeners saw them or not; Caucasian speakers, however, were perceived to be significantly more accented when the listeners saw them. I explained this finding as an audiovisual mismatch effect: because listeners did not expect to hear an accent when they saw a Caucasian speaker, the foreign accent that they heard stood out more, effecting a higher accentedness score in the audiovisual condition.

Lastly, Experiment 3 discovered a significant effect of the rate of speech, so the faster the speakers were talking, the less accented they were rated. Here, we observe the relationship between the rate of speech, fluency, and accentedness (also see e.g., Kang, Rubin, & Pickering, 2010). Naturally, the more fluent speakers were probably less accented, and they were judged accordingly. As there was only one clip per speaker in Experiment 3, this is probably an additional between-speaker effect. It may be that such an effect may surface as a within-speaker effect as well; however, Experiments 1 and 2 did not find a significant effect of the rate of speech. With multiple clips per speaker recorded in different settings, presumably, with different rates of speech, the effect was not significant.

To sum up, based on the results of the three experiments and the factors tested (proficiency and rate of speech exclusive), a NNES is expected to be rated least foreign/second language accented if they are recorded in a naturalistic service encounter and are rated by a listener with L2 knowledge in an audio only (or audiovisual for Asian speakers) condition with the range of speakers including NESs. The highest accentedness rating score is expected to be received by a female NNES speaking on a technical topic rated by a listener with no L2 experience (in the audiovisual condition – for Caucasian speakers) in a line up with other NNES voices only.

This chapter focused in the main on variation in perceived accentedness; the following chapter discusses variation in passing for a native speaker and the cues listeners use when making judgments as to the speaker's nativeness and origin.

Chapter 5 : Passing for a Native Speaker

This chapter revisits the results of Experiment 2, which studied the speakers' passing performance quantitatively, and explores the cues that listeners reported they used for identification of speakers' origin. It addresses the specific questions about passing:

- What is the variation in NNESs' passing for a NS of different English dialects?
- What are some factors that contribute to a successful passing performance?
- What are some of the elements that listeners notice in the input when a speaker succeeds or fails at passing?

Additionally, speaker self-reports and beliefs about passing are compared to the experimental results. Three case studies of German L1 speakers Kahui and Jack and Korean L1 speaker Emily attempt to link production and perception in passing for a native speaker: their production of vowels, exhibited passing performance, and listener cues reported by the participants are discussed.

5.1 Experiment 2 Revisited: Passing for a native speaker of different varieties

This section discusses the results of Experiment 2, described in detail in Section 4.2, pertaining to the speakers' passing performance. In Experiment 2 the listeners (1) rated clips produced by native and non-native speakers of English on an accentedness scale, (2) attempted to identify the geographical origin of the speaker, and (3) commented on the cues that affected their decisions. This section focuses on (2): listener identifications of the speakers' origin.

Table 5.1 and Table 5.2 represent the percentage of times the speakers were assigned to a particular regional category. The shaded cells mark the highest percentage assignment for each speaker. All NESs but one, Amy, received majority identifications by region correctly with the percentage ranging from 83.9 to 94.1 (Table 5.1, Figure 5.1). The female L1 speaker of SBE Amy, however, was most often judged to be a New Zealander (35.6%) with the correct answer following closely behind (33.9%). She moved to New Zealand at the youngest age of all NESs of other varieties (18) and had lived there the longest (4 years), so it is possible that she had picked

up some local features and abandoned some SBE ones. As she is a NES, though, her production is beyond the scope of this thesis. This high level of recognition was expected for in-group and Standard varieties as listeners would be quite familiar with them through everyday communication and mass media (Kerswill & Williams, 2002; see listener comments in Section 5.2). If not recognized correctly in terms of their region of origin, the NESs were most often assigned to other English-speaking regions (range: 5-12.7%, excluding Amy), but NNES guesses, such as Europe or Asia, were also made but were quite rare (range: 0-8.4%).

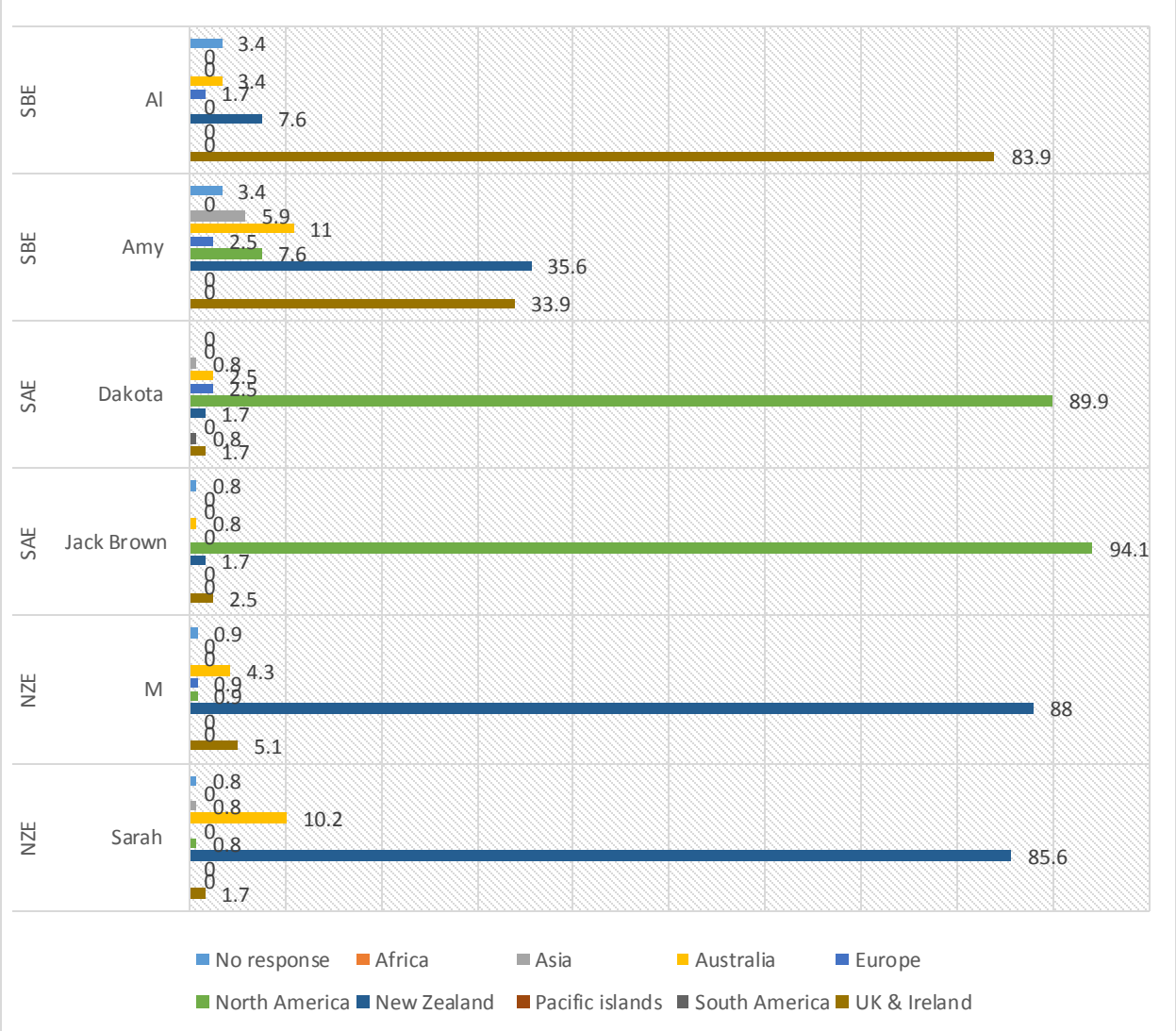

Figure 5.1: Percentage of regional assignments for first language English speakers

Table 5.1: Percentage of regional assignments for first language English speakers, most popular choice for speaker shaded

Dialect	Response / Speaker	No response	Africa	Asia	Australia	Europe	North America	NZ	Pacific islands	South America	UK & Ireland
SBE	Al	3.4	0	0	3.4	1.7	0	7.6	0	0	83.9
SBE	Amy	3.4	0	5.9	11	2.5	7.6	35.6	0	0	33.9
SAE	Dakota	0	0	0.8	2.5	2.5	89.9	1.7	0	0.8	1.7
SAE	Jack Brown	0.8	0	0	0.8	0	94.1	1.7	0	0	2.5
NZE	M	0.9	0	0	4.3	0.9	0.9	88	0	0	5.1
NZE	Sarah	0.8	0	0.8	10.2	0	0.8	85.6	0	0	1.7

In contrast, and as expected, in general there was a lot of variation in judgments for the NNESs (Table 5.2, Figure 5.2 and Figure 5.3). First of all, only one NNES Vincent did not pass for a NES of any variety at least once and the remaining 17 did which suggests that passing can be quite common.[6] In fact, for higher proficiency speakers passing was very frequent. For example, three higher proficiency speakers Emily, Kahui, and Sam were judged to be New Zealanders most of the time with German L1 speaker Kahui and Korean L1 speaker Sam receiving well over 50% of such judgments: 71.2% and 87.3% respectively, that of Sam being in the NES range. The three next most popular choices for Kahui and Sam are other English-speaking regions (Kahui: UK & Ireland 11.9%, Australia 8.5%, North America 3.4%; Sam: Australia 4.2%, UK & Ireland 4.2%, North America 1.7%) with only 2.5% and 1.6% respectively classifying them as NNESs. This suggests that studies of second language acquisition may benefit from considering the cases of passing for a NS of different varieties. Disregarding passing for a native speaker of other varieties underestimates the NNSs' ability to pass for a native speaker. In fact, the two speakers of interest, who were not classified as native

[6] Of course, this experiment is only an approximation to natural communication. Also, the clips were quite short and an attempt was made to avoid passages with grammatical errors, which might have constituted favorable conditions for passing. On the other hand, clips were not altered, so the results at least speak to the potential for passing behavior.

speakers of Egyptian Arabic by 38% of listeners in Ioup et al. (1992), may have still passed for a native speaker of another variety of Arabic to some or all of them.

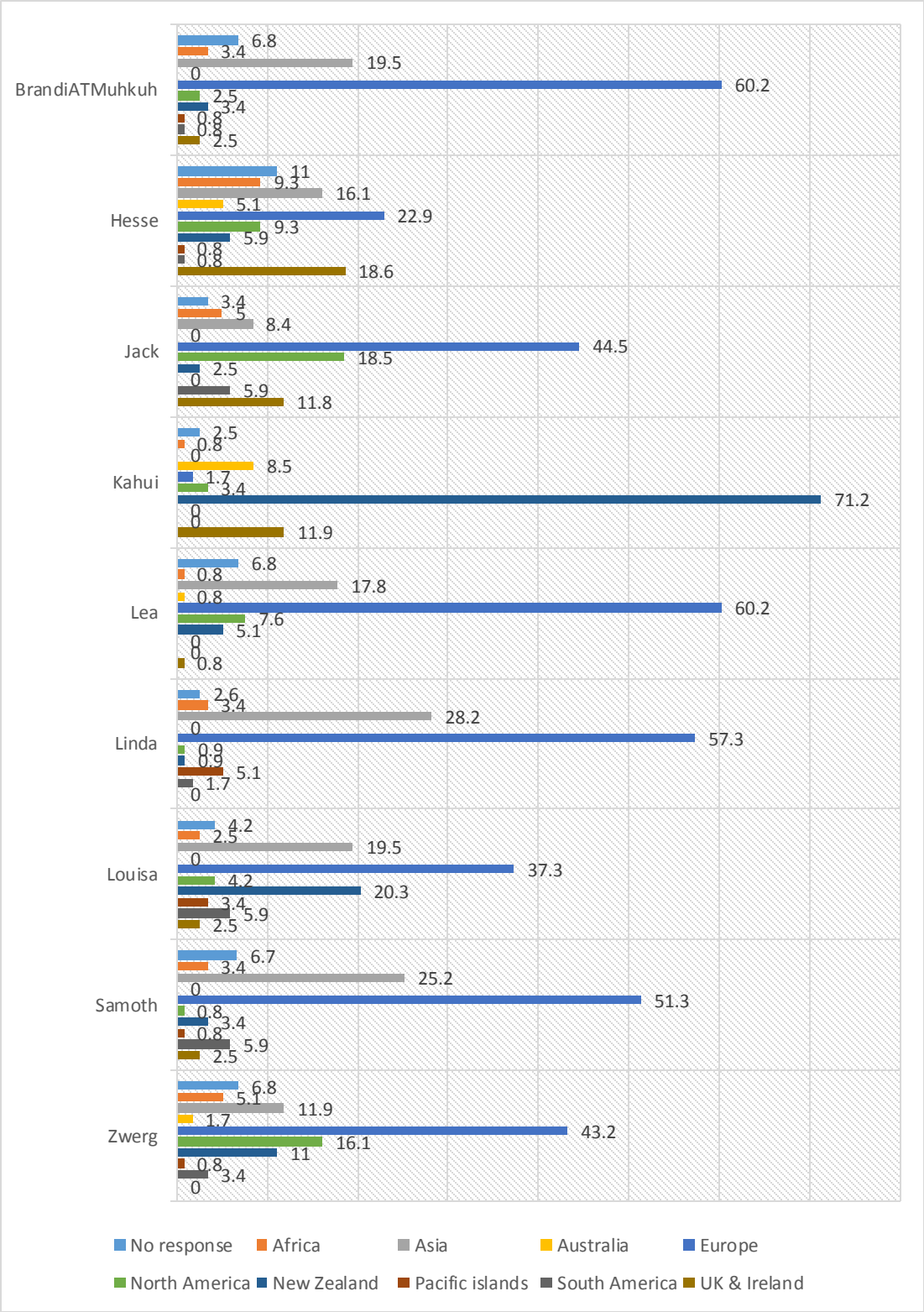

Figure 5.2: Percentage of regional assignments for first language German speakers

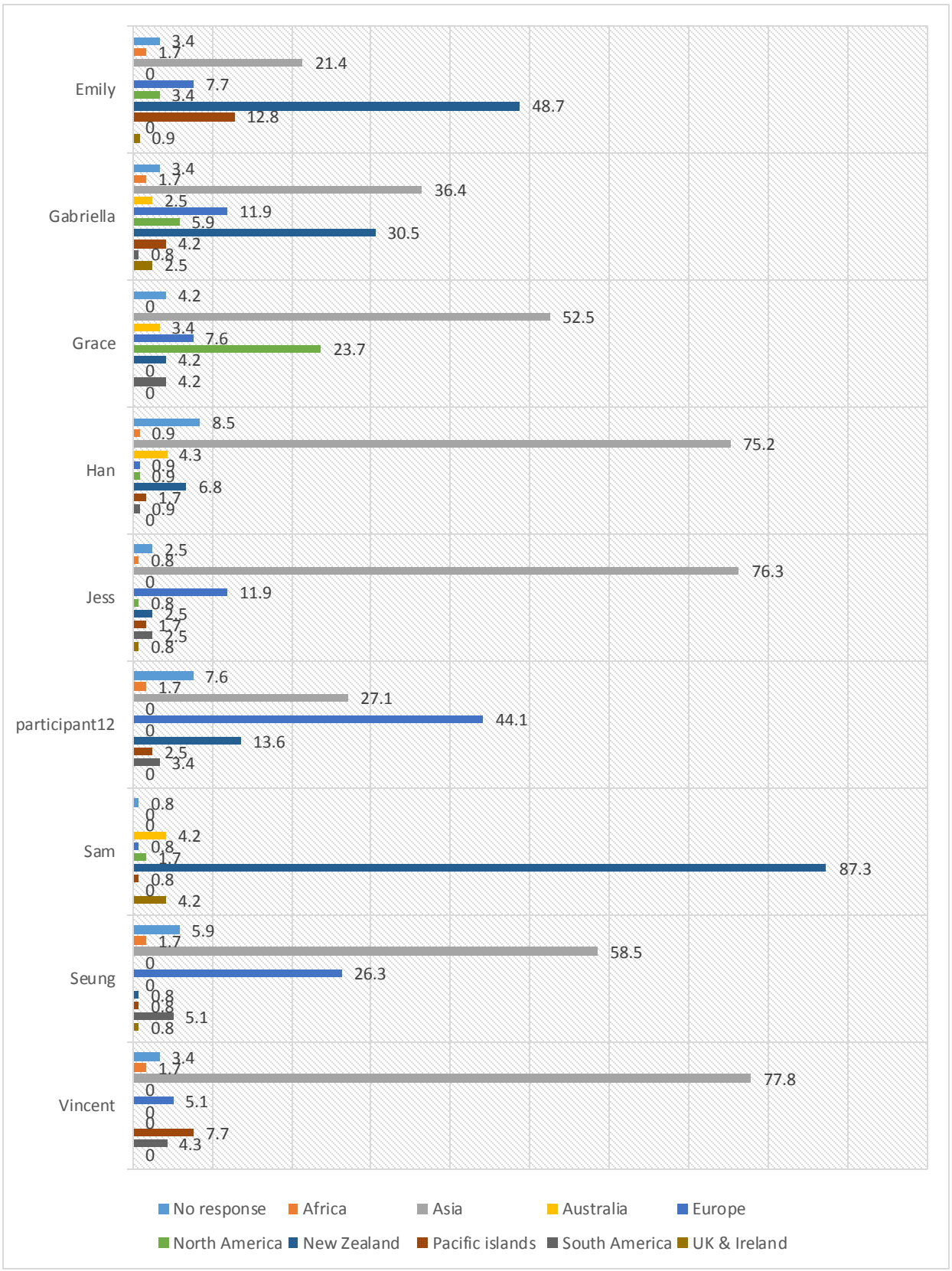

Figure 5.3: Percentage of regional assignments for first language Korean speakers

Table 5.2: Percentage of regional assignments for second language English speakers, most popular choice for speaker shaded

L1	Proficiency (NNESs)	Response / Speaker	No response	Africa	Asia	Australia	Europe	North America	NZ	Pacific islands	South America	UK & Ireland
German	Higher	Hesse	11	9.3	16.1	5.1	22.9	9.3	5.9	0.8	0.8	18.6
	Higher	Jack	3.4	5	8.4	0	44.5	18.5	2.5	0	5.9	11.8
	Higher	Kahui	2.5	0.8	0	8.5	1.7	3.4	71.2	0	0	11.9
	Higher	Louisa	4.2	2.5	19.5	0	37.3	4.2	20.3	3.4	5.9	2.5
	Higher	Samoth	6.7	3.4	25.2	0	51.3	0.8	3.4	0.8	5.9	2.5
	Higher	Zwerg	6.8	5.1	11.9	1.7	43.2	16.1	11	0.8	3.4	0
	Lower	BrandiATMuhkuh	6.8	3.4	19.5	0	60.2	2.5	3.4	0.8	0.8	2.5
	Lower	Lea	6.8	0.8	17.8	0.8	60.2	7.6	5.1	0	0	0.8
	Lower	Linda	2.6	3.4	28.2	0	57.3	0.9	0.9	5.1	1.7	0
Korean	Higher	Emily	3.4	1.7	21.4	0	7.7	3.4	48.7	12.8	0	0.9
	Higher	Gabriella	3.4	1.7	36.4	2.5	11.9	5.9	30.5	4.2	0.8	2.5
	Higher	Grace	4.2	0	52.5	3.4	7.6	23.7	4.2	0	4.2	0
	Higher	Han	8.5	0.9	75.2	4.3	0.9	0.9	6.8	1.7	0.9	0
	Higher	participant12	7.6	1.7	27.1	0	44.1	0	13.6	2.5	3.4	0
	Higher	Sam	0.8	0	0	4.2	0.8	1.7	87.3	0.8	0	4.2
	Lower	Jess	2.5	0.8	76.3	0	11.9	0.8	2.5	1.7	2.5	0.8
	Lower	Seung	5.9	1.7	58.5	0	26.3	0	0.8	0.8	5.1	0.8
	Lower	Vincent	3.4	1.7	77.8	0	5.1	0	0	7.7	4.3	0

The remaining NNESs received a wide range of guesses, which suggests that, generally, the listeners could not uniformly identify foreign accents as well as native English accents. None of them was assigned to a region with the same percentage of agreement as for NESs, the highest being 'Asia' for L1 Korean lower proficiency speaker Vincent at 77.8%. This number includes a wide range of countries that were mentioned, e.g., China, Japan, and Korea, which means that the listeners were even less accurate at identifying the exact country. Other NNESs who received listener agreement on a NNES region at over 50% were L1 German speakers of lower

proficiency BrandiATMuhkuh, Lea, Linda, and higher proficiency speaker Samoth and L1 Korean speakers of higher proficiency Grace and Han and those of lower proficiency Jess and Seung. In all these cases the majority of listeners identified the German L1 speakers as European and the Korean L1 speakers as Asian. All six speakers of lower proficiency in the two L2 groups were correctly identified as being Asian or European by origin by more than 50% of the listeners while only three of the twelve higher proficiency speakers received a majority rating in their correct region of origin. This suggests that lower proficiency speakers had more stereotypical accents which were easier for listeners to recognize.

The high variation in the listeners' judgments can be seen when the percentages of the most popular guesses are compared for the NES and NNES of higher and lower proficiency groups. For the NES group the range for the percentage of the most popular choice was 35.6-94.1 (mean=79.5); for the NNES of lower proficiency group the range was 57.3-77.8 (mean=65.1), and for the NNES of higher proficiency group the range was 22.9-87.3 (mean=51.2). The mean percentages of guesses indicate that there was more listener agreement about the origin of the NESs than the NNESs (more listeners identified the same region for NESs than NNESs) and more listener agreement about the NNESs of lower proficiency than the NNESs of higher proficiency as a group. This suggests that it was more difficult for listeners to identify the origin of the NNESs than NESs and NNESs of higher proficiency than those of lower. For example, L1 German speaker of higher proficiency Hesse was classified as mainland European by only 22.9% of respondents, the lowest of the most popular choices for all speakers. The next most popular choice for Hesse, UK & Ireland, follows only 4.3% behind at 18.6%; the third, fourth, and fifth most popular are Asia 16.1%, no response 11.0%, and Africa and North America with a tie at 9.3%. The wide range of variation in regional guesses, with the lowest standard deviation of all speakers at 7.0, suggests that the listeners found it quite difficult to place her accent. This interpretation is supported by Hesse's high 'no response' score. Hesse had the highest 'no response' score of all the speakers with 11% of listeners failing to provide a response. When asked about where people usually think that she is from in the post-recordings questionnaire, Hesse herself showed awareness of the mixed nature of her accent and its perception:

> … and I don't know if they're just being nice when they say, "Oh, I thought you were British". Some people say they can pick up a British accent; others say they can hear kind

> of like an American accent in there. I don't think I have either to be honest. Both are really strong, and I don't see myself there. I must be somewhere in between. And there must be some German accent as well because I have problems with sound/soundless voices. I notice that myself. There must be something. (Hesse.AA)

She does not comment on whether the mixed accent is her conscious choice as is suggested might be the case in Piller (2002) and Rindal & Piercy (2013), and it could also be explained by her personal linguistic history. In her post-recordings questionnaire she indicated that she had lived in the USA, Australia, and India for 10, 9, and 2 months respectively, which could have resulted in many different influences on her accent and identity.

Such a 'neutral' or 'mixed' accent, presumably influenced by several dialects, would most likely exhibit features pertaining to these varieties, which the listeners notice and classify accordingly. For example, Grace, Jack, and Zwerg are the only NNESs of all the higher proficiency speakers who consistently pronounced /r/ in non-prevocalic position. The same speakers were the ones who received over 15% of responses in the North America category. Such regularity suggests that there is some level of reliance on a cue (see Section 5.4 for more detail on feature salience). This suggests that the production of non pre-vocalic /r/, even when coupled with other non-target performance, may be a strong and sufficient marker of North America for some speakers of other English varieties.[7] Miller (2010), for example, argues that listeners may think they recognize an accent when they hear one highly salient feature despite other less salient features pointing in the other direction. From the point of view of the speaker, Piller (2002, p. 193) argues that 'L2 users may strategically employ stereotypical features characteristic of a particular variety in order to pass' and interprets one of her participants' heavy use of local feature word-initial /sp/ and /st/ instead of /ʃp/ and /ʃt/ in Standard German as a way to 'flag' her nativeness. Therefore, for a successful case of passing, NNESs do not have to project an ideal image but rather a plausible one (Giles & Williams, 1992 as cited in Giles, 2001) as languages have tolerance 'for the amount of difference that can be allowed within the normal range' (Davies, 2002, p. 144).

[7] This can only be a suggestion at this stage. Further work along the lines of Campbell-Kibler (2007) and Watson & Clark (2013) would be needed to tease out the perceptual effects of the presence vs absence of non-prevocalic /r/.

Passing for a native speaker of a different dialect may be an intermediate step between passing for a native speaker of the same dialect and non-passing from the production and perception angles, as discussed in Section 2.1.

(Figure 5.4) is a more compact version of Table 5.2 displaying the NNESs split in three nativeness categories and ordered by sum of percentages of the NS of the same and different dialect (passing for a native speaker of any English variety) values with 'no response' excluded. The speakers fall into several groups, which behaved differently. First, Sam and Kahui, who I have described above, behave similarly to NESs. They passed for a NS of the same dialect most of the time, followed by NS of a different dialect, and only rarely did they not pass for a NS of any English variety. The second group, consisting of Emily, Gabriella, and Louisa passed for a NS of the same dialect at least 20% of the time but did not pass for a NES at all over 40% of the time. The third group, consisting of Hesse, Jack, Grace, and Zwerg, did not pass for a NES the majority of the time and passed for a NS of a different dialect more often than for a NS of the same dialect. The speakers in these three groups were often regarded as native-like, and inconsistencies in their production were attributed to dialectal variation. The last group failed to pass for a NES more than 70% of the time (Lea, participant12, Han, BrandiATMuhkuh, Samoth, Jess, Linda, Seung, and Vincent) and passed for a NS of the same dialect or a NS of a different dialect a similar number of times suggesting that their overall non-nativeness was quite clear for listeners and only certain favorable conditions resulted in passing.

Table 5.3: Percentage of speaker assignments to NNES and NS of same and different dialect categories

Speaker	NNES	NS of a different dialect	NS of the same dialect
Sam	1.6	10.1	87.3
Kahui	2.5	23.8	71.2
Emily	43.6	4.3	48.7
Gabriella	55	10.9	30.5
Hesse	49.9	33	5.9
Jack	63.8	30.3	2.5
Grace	64.3	27.1	4.2
Zwerg	64.4	17.8	11

Louisa	68.6	6.7	20.3
Lea	78.8	9.2	5.1
participant12	78.8	0	13.6
Han	79.6	5.2	6.8
BrandiATMuhkuh	84.7	5	3.4
Samoth	86.6	3.3	3.4
Jess	93.2	1.6	2.5
Linda	95.7	0.9	0.9
Seung	92.4	0.8	0.8
Vincent	96.6	0	0

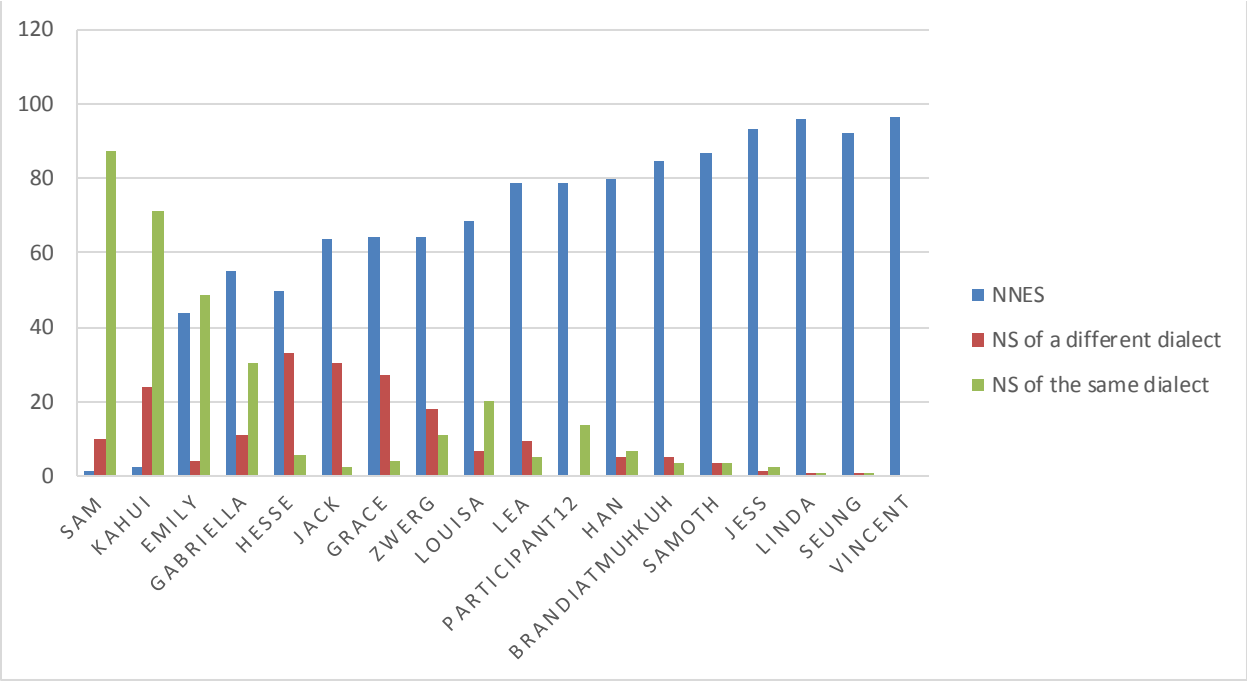

Figure 5.4: Percentage of speaker assignments to NNES and NS of same and different dialect categories

5.2 Speakers' beliefs about passing

The post-recordings questionnaire revealed the speakers' awareness of their ability to sometimes pass for a native speaker. All of the higher proficiency Korean L1 speakers except for Han, 4 out

of 6 higher proficiency German L1 speakers, with the exception of Jack and Samoth, and a lower proficiency German L1 speaker Lea could remember a successful passing experience. The quantitative results in Section 5.1 suggest that passing is more common than the speakers believe as even the speakers who did not report passing in their interviews passed for a native speaker at least once with the exception of a lower proficiency speaker Vincent.[8] This suggests that the speakers may not be aware of some successful cases of passing and, therefore, underestimate their frequency of passing. This finding supports previous claims that the language need not be perfectly native-like for an act of passing to occur because certain favorable conditions can affect it (Pattinson, 2010). The speakers also seemed aware of some of the conditions that may have such a positive effect (e.g., amount of talk):

> Some people at least say they thought I was a native speaker, but again I don't know if they're just being nice. Just a couple of weeks ago I went tramping, and we were sitting around… There was a group; we would talk to each other, but you wouldn't talk to just one person. You talk to the entire group. And one guy thought I was British. That's at least what he said. But maybe it's because I didn't really talk that much. He would only pick up once in a while that I would comment on something or ask him if he wanted a drink or something. That was very little conversation. (Hesse.AA)

With much variation in the stimulus, one can expect a range of variables to have an effect on passing. Several of the studies based on self-reports suggest that there may be contextual variation in passing and certain settings are more conducive to passing than others. For example, as suggested by Piller (2002), short service encounters may be one of such situations. The participants in this study corroborate this claim. Grace gives an example of her talking to a shopkeeper, and Kahui makes a generalization to all service encounters.

> When I went to States, for my first time, I landed at the airport in California, LA, and I was just talking with a shopkeeper. It was a young guy. And I just said this is very interesting country. It's really interesting to be here. And he said, "Where have you

[8] Of course, the clips were quite short and an attempt was made to avoid passages with grammatical errors, which might have constituted favorable conditions for passing. But clips were not altered, so the results at least speak to the potential for passing behavior.

been?" Like … he thought that I was. He didn't pick up that I was from another country. He thought I was a bit weird. And then I told him that I was from New Zealand or something. (Grace.AA)

Well, it just depends on the situation. When I meet people in service encounters, I don't talk about where I come from, but if you talk to friends or you talk to people, this topic comes up in a conversation, where are you from. (Kahui.AA)

Here Kahui supports a point brought up by Piller (2002) that passing is most common in service encounters with strangers where the identity of the speaker is not important and passing for a NS will not be regarded as deceit. Once the topic of origin is brought up in conversation, passing without deceit becomes impossible. Therefore, some speakers note that any first encounter with a stranger has potential for passing.

Some people do… mmm … I think it was when I was talking to the lady at the airport. And she's like, "oh are you from New Zealand or other". I'm like, "Oh, I'm from Korea." "You have a very good accent" so I thought she thought I was from New Zealand who was born in… yeah … (Emily.AA)

I had a few in England. So they thought I was American. But probably… I don't know. For instance, in the common room. So I was the new person. So they were like, "hey, how are you? What's your name?" and I was like, "blah-blah", five sentences. And they were like, "so you are from the States". (Lea.AA)

As one's origin is often a topic that comes up early in an informal conversation with strangers, sustaining the passing performance for a long time without deceit may be impossible. However, if the origin is not brought up, Zwerg believed that certain words, expressions, and topics, namely, more familiar and frequent ones may be helpful:

I wasn't talking a lot of sentences to them, just a few. "Hi, how are you? Da-da-da. What are you doing? What are you studying?" Maybe. And that's sentences you always say the

> same because you meet a lot of people. So it's more or less always the same sentence. And you're improving it. (Zwerg.AA)

The speakers were also aware of factors hindering the passing performance. The Asian participants in my study were acutely aware of societal stereotypes and the effect of ethnicity on passing. While the Caucasian speakers who possess a 'default' ethnicity, did not mention it at all in their debriefing interview, three higher proficiency Asian speakers claimed that people expected them to be foreign because of their looks:

> Non-native. I might be biased, but I am very convinced … I don't wanna sound racist, but kiwis expect Asians to be foreigners. Even my friends who are born here and raised here fully, and they don't speak their parents' mother tongue well, they are still asked where you're from. No, they don't ask that question thinking that you're a native speaker. It doesn't matter whether you have an accent or not. I think it's determined mostly by what the eyes see. (Gabriella.A.A.)

5.3 Discussion

The thesis aims to fill the gap in our understanding of passing for a native speaker not covered by qualitative studies based on self-reports (e.g., Piller, 2002) and quantitative studies of ultimate attainment (Abrahamsson & Hyltenstam, 2009). Reliance on speakers' self-reports foregrounds the speaker at the expense of the listener. Self-reports can also be unreliable as passing cannot be considered to have occurred unless confirmed by the listener and normally NNSs do not elicit and systematically record their interlocutors' judgments as to their assumed origin after every single encounter. This means that on the one hand, speakers may believe that passing occurred when it actually did not; on the other, there may be cases when passing occurred and the speakers were not aware of it. My results suggest that passing, as confirmed by the listeners, can be quite common. In fact, most of the speakers pass with varying regularity. As expected, speakers of high proficiency pass quite often, but even lower proficiency NNESs and / or those who do not self-report passing pass in a small proportion of cases. This suggests that NNESs' self-reports often underestimate the amount of passing that they experience on a daily basis.

Both qualitative and quantitative studies often discuss passing as a binary phenomenon (successful vs unsuccessful) and do not distinguish between passing for a native speaker of the same or different variety of the language as the listener. However, certain patterns emerge when this facet is introduced. For most speakers, this study found variation in passing for a native speaker of different dialects. This may be due to a number of speaker- or listener-related reasons. Firstly, listeners may judge certain variation as dialectal (Major, 2001). Secondly, some NNSs may choose to preserve certain features distinguishing them from the target community (Piller, 2002). Lastly, L2 speakers may not always simply move diachronically from foreign accented to native-like production in the target variety. If this is true, one may expect that most NNESs would diachronically exhibit an incremental increase in ratings from a majority NNES to NS of a different dialect to NS of the same dialect. The passing trends in the majority of speakers were consistent with this explanation, but several speakers passed for a speaker of a different dialect less often than for a native speaker of the same dialect or not pass at all. This contradicts the explanation that NNESs are first regarded to pass for a NS of a different dialect before they can pass for a NS of the same dialect. Of course, a number of factors influence a speaker's production, but it seems that at least some NNESs do not use (salient) influences from other varieties, which otherwise could be an attempt to pass for a NS of a different dialect. The listeners, in turn, likely vary in terms of the extent to which non-native and native-like features are noticed and so also the degree to which they are able to influence their final judgments.

5.4 Experiment 2 Revisited: Listener cues

The open-ended responses given by the listeners in Experiment 2 to question 3 ('Please comment on what made you think that the speaker is from that particular place: for example, is it what they said, how they said it, or something else?') were categorized into the following classes: accent (56.0%)[9], example (26.5%; the way he/she said 'X'), intonation (14.4%), zero response (6.0%), segments (5.3%), vocabulary (4.1%), content (1.5%), and grammar (1.3%). The most common response of all was 'accent', which subsumed a wide range of responses that did not give a hint towards the specific cue which the listeners used. A large number of the responses in this

[9] The percentage indicates the percent of listeners that commented on that particular category. Several participants made comments which related to more than one category, so the total sum exceeds 100%.

category were 'accent' or 'sounds familiar'. [10] The responses in the class 'example' were of the type 'the way he/she said X' (see Section 5.5), but without elaboration it was unclear what triggered the noticing of accent: vowels, consonants, suprasegmental features, or something else. The impossibility to follow up with such responses for clarification is a methodological disadvantage compared to an interview where a researcher directly interacts with the rater as in Hayes-Harb & Hacking (2015).

Some listeners commented on the overall proficiency of the speaker, and native-likeness was equated with 'good English': when judged a NES, Al was said to exhibit a 'good spoken English', and when judged a NNES, Grace was said to have 'incorrect English'. Similar evaluations of non-native English as 'bad' English have been found in both NES (McKenzie, 2015) and NNES (Zając, 2015) listeners. The equation of 'correct' English with nativeness and 'incorrect' English with non-nativeness stems from the prescriptive approach of the Standard Language Ideology (Lippi-Green, 1997) and Native Speaker Ideology (McKenzie, 2015) and gives rise to negative attitudes and stereotypes towards non-native speakers.

Naturally, non-target-like phonology can be a feature that listeners notice. In terms of intonation, raters commented on pauses, speed, and the like for those judged native and non-native English speakers. When commenting on speakers' pronunciation of segments, listeners often made quite general remarks, such as 'English twang to the words and vowels' for NES assignments and 'strange soft consonants' for NNES ones. Sometimes, however, the raters attempted to be more precise and named specific segments that stood out to them: ' "sister" and "hearted" had the r's pronounced like an American' or 'E/EY vowels accentuated' for a NES and 'aei vowels a little too similar sounding' for a NNES identification. Listeners commented more on the segments of speakers who they assigned to groups other than New Zealand (only 11.4% of the responses in the 'segments' category were given to clips with New Zealand guesses), probably because of a relative ease of describing something that is different rather than 'standard' or default, which is similar to the sentiment of one of the participants in Hayes-Harb & Hacking (2015).

Listeners used what they thought of as grammar mistakes to justify their classifications of speakers as NNESs: 'Grammar (missing an indefinite article)' or 'text instead of texted'.

[10] Because of a large number of spelling mistakes and typos in the listeners' comments, which the reader may find distracting, the examples used in the thesis have been edited to enhance readability unless it was assumed that the purposeful misspelling on the part of the listener indicated the speaker's mispronunciation of a word.

Grammatical mistakes were not a perfect predictor of being regarded a NNES as L1 Korean speaker Gabriella's clip 1 in the family setting received NS guesses at least one third of the time despite missing an indefinite article in an obligatory context. And even when noticed, the native-like accent was a stronger predictor of nativeness as Sam was judged to be a New Zealander with a foreign family history despite a grammar mistake: 'kiwi accent and inflection, however perhaps European parent as he said how much hours instead of many'.

The speaker's choice of words was sometimes used by listeners as a linguistic or an extra-linguistic cue. Failure to use an appropriate word ('oil from lamb' for lanolin) or using a dialectal variant ('they say the word math rather than maths') signalled foreignness. Use of technical jargon and terminology, on the other hand, suggested nativeness: Jack's university recording speaking of 'hydroxide complex solutions' was rated as 'most likely English first language due to vocab and clarity of speech'. At the same time, the use of slang and casual words was commented on when identifying speakers as in-group: 'like', 'yup', 'cos', 'kinda', 'that kind of stuff', 'yeah', 'cheers man', 'bro', and 'sweet' were perceived as markers of native-like identity. This suggests that appropriate use of colloquial expressions may raise one's chances of being perceived as a native speaker, at least among young adults as listeners. Use of slang may also work to create an impression of the speaker being at ease with the language. Using more formal language, on the other hand, was noticed as a mark of non-nativeness: 'Using the phrase "for example" instead of something more colloquial'.

Some speakers believed that the content of the clip provided them with a clue as to the speaker's accentedness and origin. For example, in one services clip Grace was guessed to be from India, and a listener justified it by linking topics and origins: 'she was asking for spices', and Amy was thought to be English because she was 'speaking about tea'. At the same time, reference to the same object could receive a different interpretation from different listeners: Zwerg, in the same clip which mentioned pizza and pasta, was once rated to be Italian ('food talked about') and once American ('talking about girls' night and pizza and pasta'). Such a connection is based on a listener's stereotyping of certain groups of people. Spices are used by native and non-native English speakers alike, and pizza and pasta are popular foods in many countries, so such a connection is reflective of listener stereotypes. Such stereotyping process was noticed and self-reflected upon by another listener as she guessed Han's recording in a

university setting to be from a Chinese person because of his talking about finance and commerce: 'Talking about business (ah, I'm such a stereotyper!)'.

Sometimes, the listeners indicated that they compared the clips to some sort of an ideal representation of the accent in their mind, such as their own accent (e.g., 'similar accent to me', 'similar to what I hear every day in NZ', or 'sounds like NZ TV presenter'). Kerswill & Williams (2002, p. 200) defined three reasons for accent familiarity: '(1) the degree of contact between one's own community and the community represented by the voice; (2) whether a voice sounds like someone the judge happens to know, (3) the influence of broadcast media'. These are all reflected in the listeners' justifications. The listeners often referred to examples of imported media in order to justify their guesses of speakers coming from other English-speaking countries: 'Sounds very English. I watch the TV show *The Only Way is Essex* and accent sounds kind of familiar', 'Sounds like BBC English everything pronounced clearly', 'sounds like *Coronation street*', and 'typical from what you hear in Hollywood movies'. Reference to foreign-accented media was very rare, reflecting the low listener exposure to second language-accented popular culture; however, one speaker was identified as a NNES because 'his voice sounded similar to my favourite Norwegian musician'. One can notice, however, that the speakers seem more familiar with different English varieties than foreign accents through mass media, which is not surprising because of the prevalence of English-medium programs on New Zealand TV. Reference to the accents of people the listeners had previously met was a popular justification for both native and foreign accents: 'accent was familiar to Americans I have met from overseas', 'sounded like the guy from America who used to help us with any computer troubles', 'accent sounds like Americans I have met', 'sounds like a Nepal girl I know', 'sounds exactly like my old German flatmate', 'sounds like the accent of my friend from Malaysia', and 'some Czech friends of mine sounded like this before they were exposed to more English speakers on their travels here in NZ'.

On the other hand, when the variety sounded unfamiliar, but the speaker's linguistic cues suggested native-likeness, the raters used the method of exclusion. For example, several listeners explained their reasoning behind the assignment of the speaker to Canada because of his/her perceived native-likeness but not to one of more familiar varieties: 'clearly English speaking but not NZ, USA, Australia, SA or British', 'didn't sound like any kind of recognizable accent

(England, SA, NZ, USA, Australia) yet definitely English-speaking person so went for Canada', 'not quite American', and 'vowels were long but did not sound American'.

Sometimes listeners did not limit themselves to the speakers' origin and made fine-tuned distinctions between accents in order to paint a more detailed picture of the speaker. Based on the linguistic input, the listeners made assumptions about the NNESs' language education: for instance, Zwerg was identified as a 'German taught American' and Emily 'sounds like a slight American accent on top of a Chinese accent, and lots of people in Asia seem to learn from American teachers'. The listeners took into consideration potential influences of mass media (Amy 'sounded like a New Zealander but one who consumes a lot of American media') and permanent or temporary speech disturbances, such as speech impediments or colds (e.g., Louisa 'sounded a lot like a kiwi girl who was suffering from a cold and had a blocked nose').

Some non-target-like production in speakers guessed to be NSs prompted the listeners to explain it through hypothetical speaker histories. Speakers' deviations were explained through possible non-native parent influences, knowledge of another language, or extensive travel experience. Sam, for example, was once characterized as 'born or raised in NZ; parents perhaps American', Gabriella - 'a person of Asian descent born in New Zealand and probably bilingual from a young age', and Zwerg – 'ambiguous accent, perhaps has lived in more than one country as her accent sounded American but also like something else; probably English as a first language though'. Sometimes deviation from 'standard' was explained through ethnic accent differences in both NZE and other varieties. Sam, for example, 'sounds like possibly a Māori boy, south Auckland accent, definitely New Zealand English though', and Emily 'sounds like a Māori girl' while Grace was thought to be Hawaiian American and Louisa – African American. These examples are reminiscent of 'speaker models' identified by Hayes-Harb and Hacking (2015). It can be seen that the listeners did not limit themselves to one category, native accentedness, but created a rich picture of the speaker, which included their socio-economic background, family history, and even personality.

To explore the manifestations of listener employment of their socio-cultural knowledge in the perception tasks, the descriptive and evaluative vocabulary used by the listeners to describe the speakers was analyzed. First, individual words in listener comments were automatically tagged for part of speech by the CLAWS part-of-speech tagger (Garside & Smith, 1997). Then, words tagged as adjectives were manually corrected for spelling, and clearly

misidentified non-adjectives were excluded from analysis as well as words pertaining to geographical origin (e.g., American) as they had been covered earlier when 'speaker models' were discussed. Word clouds were created for remaining adjectives in WordleTM (Feinberg, 2014). Figure 5.5, Figure 5.6, Figure 5.7, and Figure 5.8 show the adjectives that listeners used when justifying their judgment for different groups of speakers. For the purpose of comparison of reactions to native and non-native speech, word clouds were created for native speakers of two English varieties and non-native speakers of L1 Korean and L1 German when identified as a non-native speaker. The size of the word corresponds with its frequency in the listener judgments, and the adjectives of interest which are discussed below are circled.

Certain social stereotypes surfaced when connections were made between paralinguistic behavior and speaker origin. Sounding comfortable and confident was often associated with being a native speaker. It can be seen in Figure 5.5, Figure 5.6, Figure 5.7, and Figure 5.8 that the words *comfortable* and *confident* were more frequent in descriptions on SAE and SBE L1 speakers than Korean or German L1 speakers when judged a NNES. For example, Hesse was judged to be from the UK and was assessed to be 'fluid and confident'. Similarly, Watanabe (2008) found a link between confidence and L1 native-likeness in a language attitudes study. Hesitation, on the other hand, was often connected with non-nativeness: 'the hesitation before some words suggests it's a second language' and 'the hesitation, like she was trying to find the words, definitely suggested that English wasn't her first language'. The descriptions *hesitant*, *broken*, and *disjointed* are present in Figure 5.7 and Figure 5.8 but absent in descriptions of NESs (Figure 5.5 and Figure 5.6). When thought to be American, both native and non-native English speakers were commented on as 'confident and bubbly sounding', 'strong and dominating', and having 'very animated and excited speech' (see Figure 5.5 for SAE L1 speakers). The speakers' assumed socio-economic class also came into consideration: Kahui comes from a 'wealthy family', Samoth is 'well educated', Al 'sounds kind of posh', and Amy is a 'posh kiwi' (see Figure 5.6 for SBE L1 speakers).

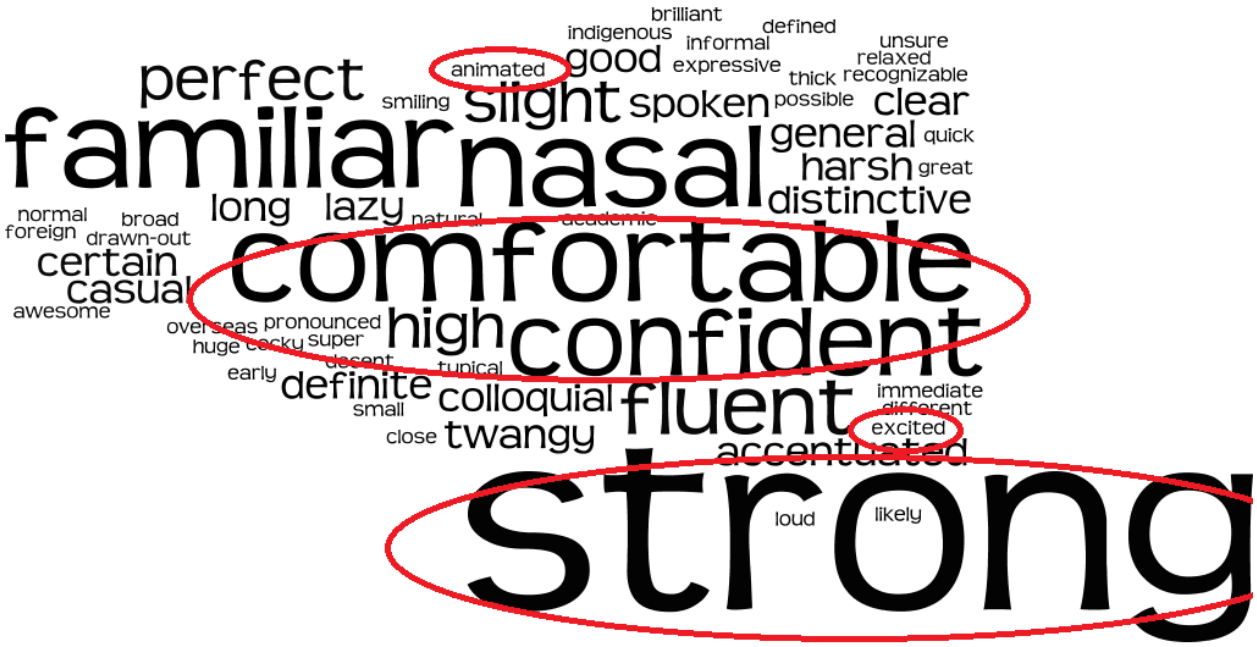

Figure 5.5: The word cloud for adjectives used in descriptions of SAE L1 speakers

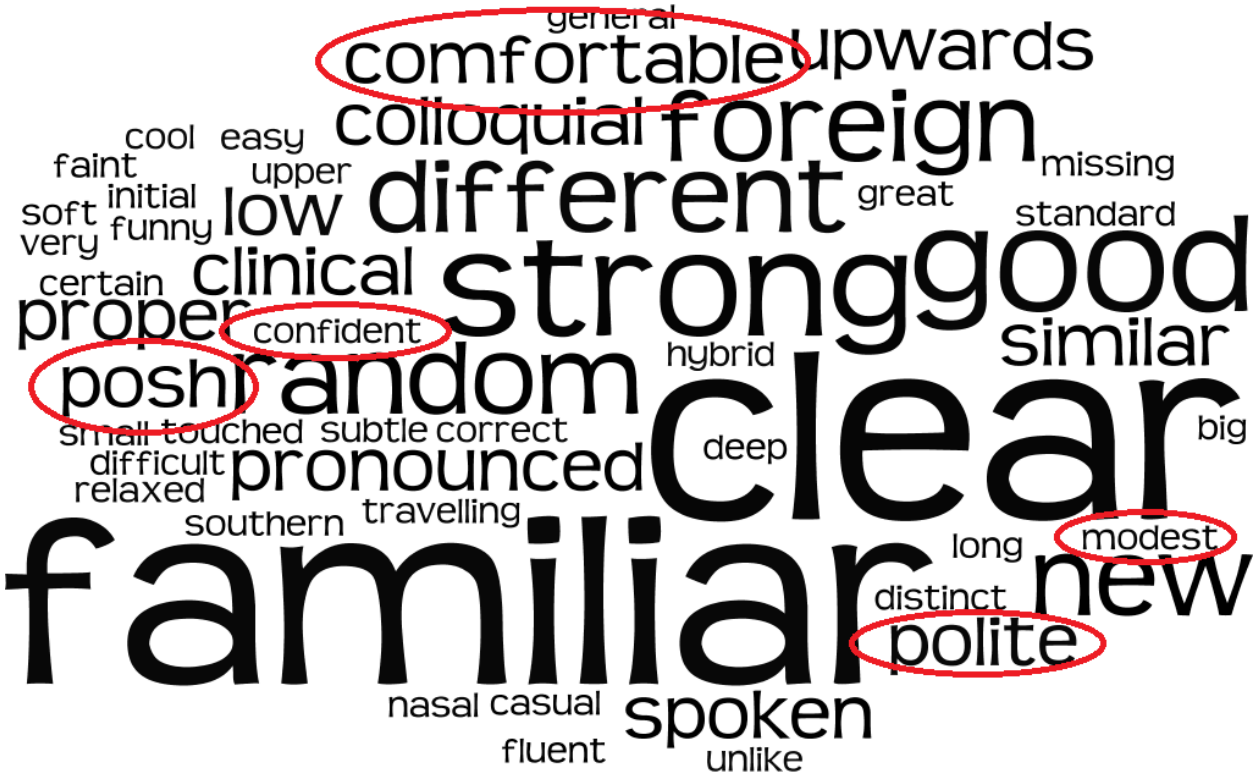

Figure 5.6: The word cloud for adjectives used in descriptions of SBE L1 speakers

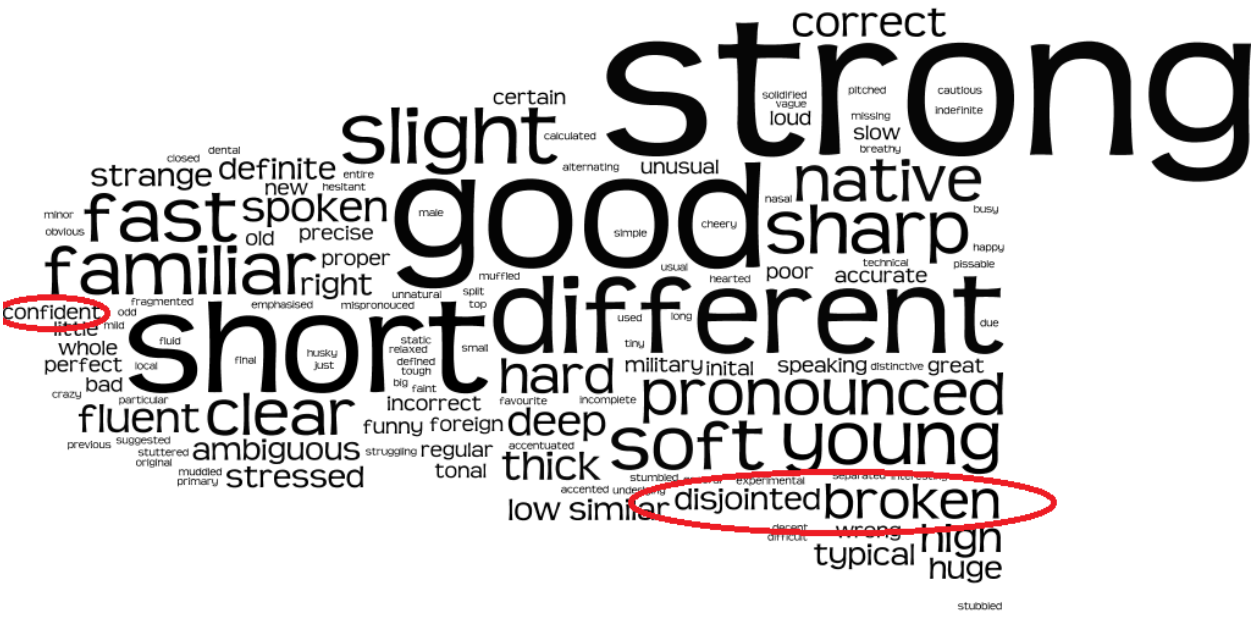

Figure 5.7: The word cloud for adjectives used in descriptions of Korean L1 speakers when judged a NNES

Figure 5.8: The word cloud for adjectives used in descriptions of German L1 speakers when judged a NNES

The examples above illustrate the listeners' use of 'speaker models' and descriptive vocabulary referring to extra- and paralinguistic features when participating in a perception task. Experimental studies have often found an effect of social information on speech perception (Drager, 2010; Rubin, 1992; Section 4.3). In such experiments the raters' stereotypes and attitudes are inferred from their behavior in different conditions (e.g., an Asian speaker is perceived to be more accented than a Caucasian speaker; Rubin, 1992). The listeners' explicit comments on the social features pertaining to the speakers lend extra support to listener reliance on social cues.

5.5 Case studies: Passing for a native speaker in production and perception

In this section I compare several NNESs' passing in production and perception and once again consider the importance of salient features for a passing performance. I focus on Kahui, who passed for a NS of NZE most of the time, Emily, who passed for a NS of NZE approximately as often as she failed to pass of a NES, and Jack, who rarely passed for a NS of NZE but passed for a NES over 20% of the time. The stressed vowels produced by Kahui, Emily, and Jack in the four settings (Chapter 3) were plotted for visual comparison with an ideal NZE speaker's vowel space. Their passing for a NS is revisited and discussed in light of the listeners' comments on what they believed made them guess where the speaker was from.

5.5.1 Kahui

Kahui is a 23-year-old male near-native speaker of English, a German L1 speaker. He began his formal study of English at the age of about nine, in his 'English as a foreign language' classes at school. Later, as an adult, he visited England for one month with the purpose of preparing for a standardized test of English proficiency before he moved to New Zealand 18 months before this study took place. He reported using English almost 100% of the time with the exception of weekly Skype sessions with his family in Germany.

It is clear that Kahui's vowel space is very similar to the prototypical NZE vowel space (Figure 5.9). Many vowels are quite NZE-like: for example, TRAP is somewhat raised, DRESS is very high and fronted, KIT is centralized, and GOOSE and NURSE are front; however, the overlap is not perfect with Kahui's LOT and STRUT vowels being higher compared to the NZE

ideal. Despite the visually quite native-like vowel-space, a speaker may not pass for a native speaker for other segmental (consonants), suprasegmental (intonation), grammatical (syntactic deviations), and other reasons (see Section 5.4).

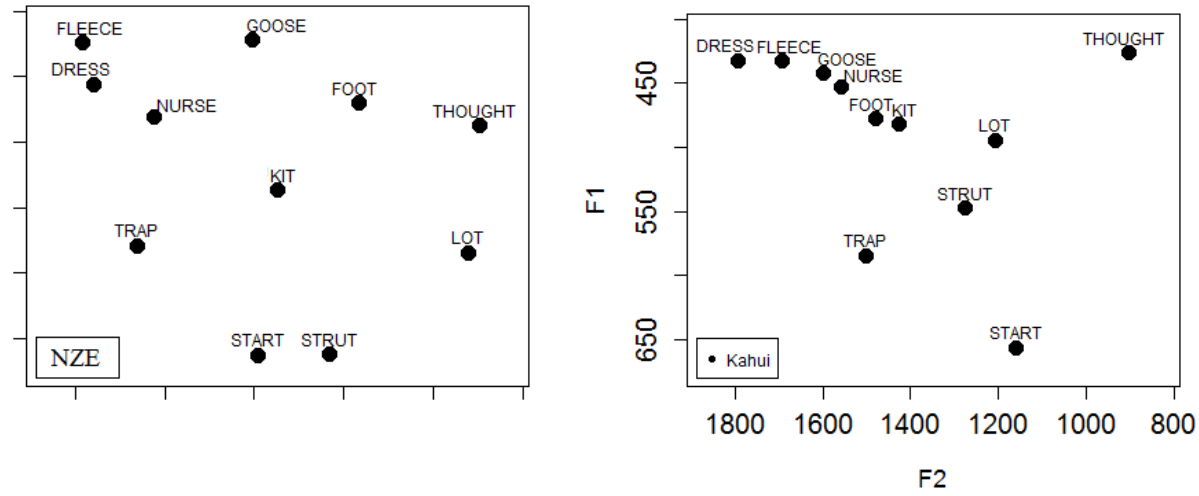

Figure 5.9: Vowel spaces of NZE (left panel) and L1 German speaker Kahui (right panel)

In native listeners' perception, Kahui passed for a native speaker of NZE in the majority of cases (71.2%), a few times for a native speaker of other varieties of English (23.8%), and for just 2.5% of the listeners he did not pass for a native speaker while 2.5% gave no response (Section 5.1). Content analysis of the listeners' comments revealed many general phrases, where listeners gave holistic judgments (e.g., 'kiwi accent'). However, some listeners identified particular segments as a trigger (e.g., 'maybe Australian with the vowel sounds'). Sometimes the listeners were more detailed and provided lexical examples (e.g., 'frish not fresh', which probably illustrates the raised quality of the DRESS vowel, typical of NZE). Such imitation of speaker features may be reflective of the non-linguist listeners' lack of terms for description but, at the same time, a certain degree of awareness of noticeable differences (Preston, 1996).

These examples were categorized by lexically stressed vowel (see Table 5.4 for monophthongs). The listeners used five lexically stressed examples of the DRESS vowel, four of KIT, three each of GOOSE and TRAP, two of NURSE, and one each of FLEECE, START, and STRUT. Most of these vowels are quite distinctive in NZE, and the listeners may be using them as a shibboleth. DRESS, KIT, and TRAP, which are involved in a chain shift in NZE, were commented on the largest number of times.

Table 5.4: Listeners' lexical examples when identifying Kahui as a NS of NZE

DRESS	FLEECE	GOOSE	KIT	NURSE	START	STRUT	TRAP
eleven x 2 fresh x 2 ten	unbelievable	school you youtube	chilli x 2 think x 2	working x 2	mark	mum	thank x 3

Because of the small number of NNES guesses, there were no examples provided by listeners to support their identification, but almost a quarter of listeners thought Kahui to be a NS of another English variety, and some of them used illustrations in their comments (see Table 5.5). The listeners used three lexical examples containing the stressed FLEECE vowel and one each of DRESS, GOOSE, STRUT, and THOUGHT. One can see that there is some difference between the vowels involved when Kahui was judged to be a NS of NZE and when he was judged to be a NS of another English variety. The short front vowels DRESS, KIT, and TRAP, as well as GOOSE and NURSE, which are distinctive of NZE, are prevalent in Table 5.4 but only emerge twice in Table 5.5. Additionally, in Table 5.5, there are more illustrations of the FLEECE vowel which was only mentioned once in Table 5.4.

Table 5.5: Listeners' lexical examples when identifying Kahui as a NS of a different English variety.

DRESS	FLEECE	GOOSE	STRUT	THOUGHT
ten	believe x 2 unbelievable	computes	suddenly	always

For comparison, a NS of NZE M also received many illustrations which contained characteristic NZE vowels: four of the DRESS vowel, two each of FLEECE, GOOSE, STRUT, and THOUGHT, and one each of KIT, LOT, NURSE, and START (Table 5.6), which suggests that the listeners were noticing the native-like production of salient vowels.

Table 5.6: Listeners' lexical examples when identifying M as a NS of NZE.

DRESS	FLEECE	GOOSE	KIT	LOT	NURSE	START	STRUT	THOUGHT
benefit	cheese	movies	will	what	perfectly	partner	lovely	awesome

guess then vet	freeze	view					lunch	pause

To explore how the number of the characteristic vowels in a given clip influences its passing for a NS of NZE, along the lines of Watson, Leach, and Gnevsheva (submitted), I calculated the number of words (Nwords), the number of lexically stressed vowels DRESS, KIT, TRAP, NURSE, and GOOSE, and the percentage of speakers that it successfully passed for a NS of NZE (pass) for each of the twelve Kahui's clips. I conducted Principle Components Analysis (PCA) on these data. The loading plot in Figure 5.10 represents the relationships between the variables in the space of the first two components. We can see that the percentage of passing and the number of lexically stressed KIT vowels in a clip have similar heavy loadings which suggests that they're positively correlated while the number of lexically stressed TRAP vowels seems to be negatively correlated with the percentage of passing.

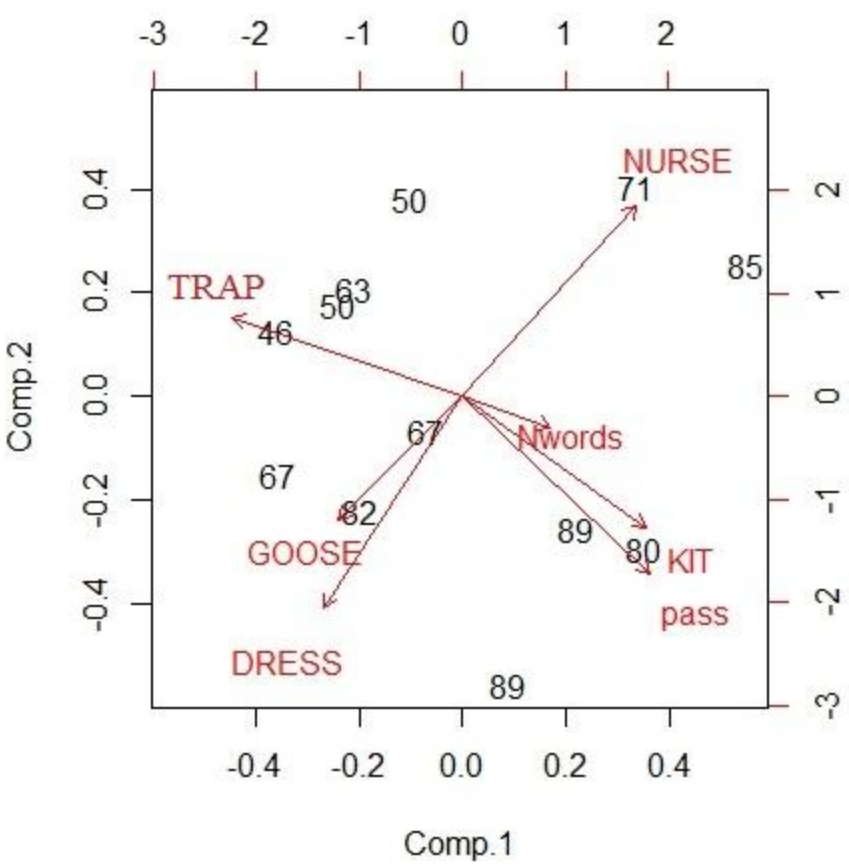

Figure 5.10: The loading plot of PCA for Kahui.

To sum up, Kahui's overall monophthongal vowel production was quite native-like, with NZE-like production of many of the characteristic vowels, despite no perfect overlap between his vowel space and the prototypical NZE vowel space. Kahui self-reported having experienced passing for a NS (Section 5.2), and his claim was supported by the perception experiment in which he passed for a NS of NZE over 70% of the time. When these listeners had to justify their decision, the words mentioned as examples contained vowels which are salient markers of New Zealand identity (Hay & Drager, 2010) and which Kahui produced in a NZE-like fashion, according to his vowel space. This finding is based on only few observations, but it nevertheless highlights the importance of salient features.

5.5.2 Emily

Emily is a 21-year-old female near-native speaker of English, a Korean L1 speaker. She started learning English with a tutor in her home country at the age of 10. She spent 1 month in Australia at the age of 11 before moving to New Zealand permanently at the age of 12 in order to continue her education there. She reported speaking both languages an approximately similar amount at the time of the study: Korean with her family and friends in Korea and Korean friends in New Zealand and English with her flat-mates, at the university, and with some Korean friends.

The monophthongs produced by Emily were also quite native-like (Figure 5.11): for instance, TRAP and DRESS are raised, GOOSE is fronted; however, there are some differences as Emily's KIT is high and front and very close to DRESS, NURSE is mid-central, and LOT and STRUT are higher compared to the NZE ideal.

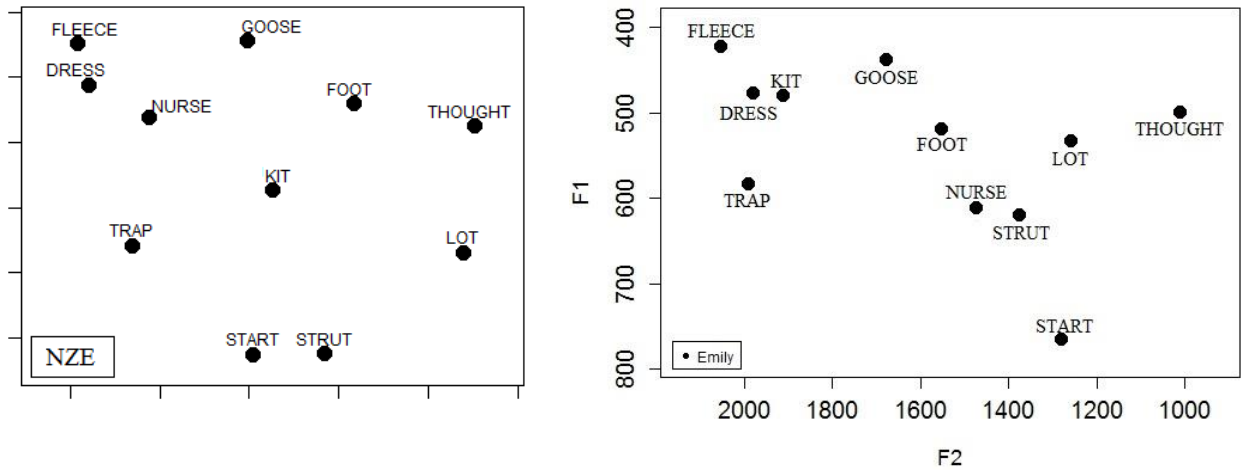

Figure 5.11: Vowel spaces of NZE (left panel) and L1 Korean speaker Emily (right panel)

In native listeners' perception, Emily passed for a native speaker of NZE (48.7%) much less frequently than Kahui despite her quite NZE-like vowel space, which illustrates listeners' reliance on a wide range of cues probably including both segmentals and suprasegmentals. Interestingly, several of the listeners who thought Emily to be from New Zealand and gave her a low accentedness score (more native-like) clarified that they believed her to be Māori ('sounds like a Māori girl'). In their justifications for this identification, the listeners commented on her pronunciation of consonants (e.g., 'fank you'), which is a clear example of how foreign

influences can be heard as dialectal (Markham, 1997 as cited in Major, 2001) in the case of the dental fricative /θ/ which is absent in Korean.

When looking at the lexical examples that listeners identified, we can notice that, if classified by stressed monophthongs, five categories are represented (Table 5.7). The listeners used four lexically stressed examples of the GOOSE and the TRAP vowels, two of the DRESS vowel, and one each of START and STRUT. Once again, we see the vowels that are representative of NZE. It is also a subset of Kahui's lexical examples (Table 5.4). Intriguingly, two of the categories that are missing for Emily in comparison to Kahui are KIT and NURSE, the two vowels which are less NZE-like in her vowel space (Figure 5.11). I take this as evidence that the pronunciation of other vowels which the listeners noticed was more target-like and the sum of more and less native-like productions was enough in order to justify a NS of NZE identification.

Table 5.7: Listeners' lexical examples when identifying Emily as a NS of NZE

DRESS	GOOSE	START	STRUT	TRAP
yeah yep	huge you x 3	market	just	dad grandmother thank x 2

Emily received few NS of a different variety identifications, and no illustrations were used then. However, she was believed to be a NNES almost as often as a NS of NZE, and several lexical examples were used for justifications (Table 5.8). Some differences between the examples in Table 5.7 and Table 5.8 can be observed: there are fewer examples of distinguishing NZE vowels ('thank you' and 'tend') and examples of vowels which were not found in Table 5.7 (e.g., 'positions' and 'product').

Table 5.8: Listeners' lexical examples when identifying Emily as a NNES

DRESS	FLEECE	GOOSE	KIT	LOT	START	TRAP
tend	people	you	positions	product	market	thank

To explore how the number of the characteristic vowels in a given clip influences its passing for a NS of NZE for Emily, I calculated the number of words, the number of lexically stressed vowels DRESS, KIT, TRAP, NURSE, and GOOSE, and the percentage of speakers that it successfully passed for a NS of NZE for each of the twelve clips in the same way that I did for Kahui. In Figure 5.12 we can see that Emily's percentage of passing is positively correlated with the number of lexically stressed TRAP vowels in the clip. This suggests that the more of the TRAP vowels were present in the clips, the more likely she was to pass. Emily's native-like production of the vowel seen in Figure 5.11 and the listeners' frequent comments on this vowel when identifying her as a NS of NZE (Table 5.7) suggest that the listeners were relying on the TRAP vowel as a marker of NZE-likeness and a larger number of it present in a given clip supported their NS of NZE categorization.

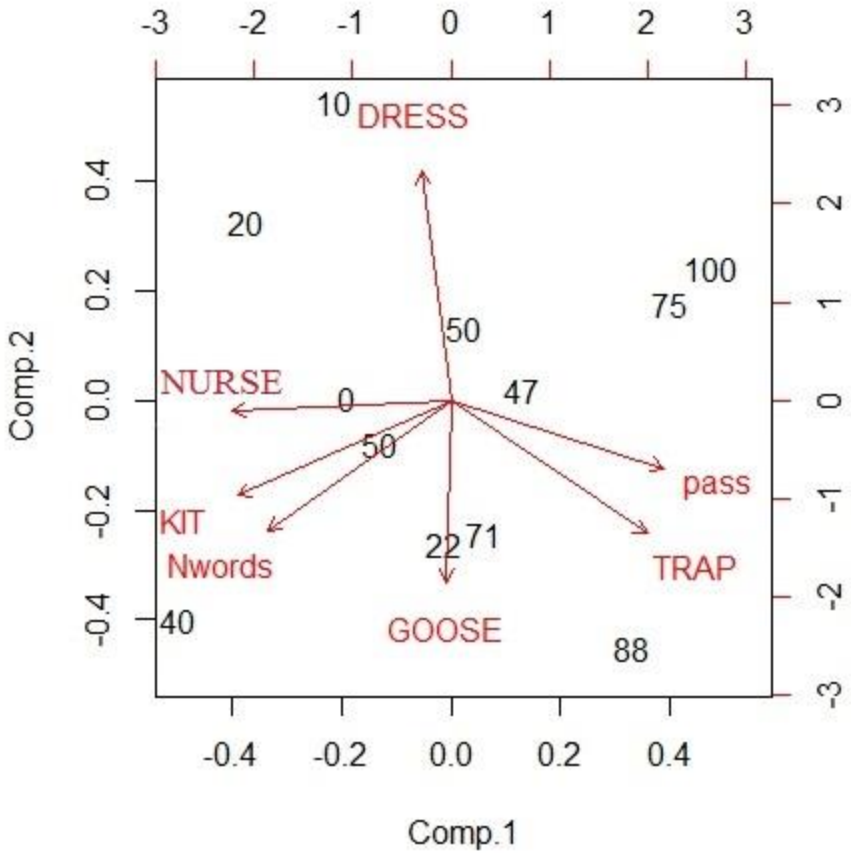

Figure 5.12: The loading plot of PCA for Emily.

Emily self-reported passing cases in the past (Section 5.2); however, her near-native vowel space did not result in frequent passing in the experimental setting comparable to that of Kahui. This may be reflective of her non-native-likeness in other linguistic and extra-linguistic domains. Additionally, unlike Kahui's, her production of one of the salient markers of NZE, the KIT vowel, was less NZE-like, which may have made her non-native-likeness more noticeable through more deviations from the ideal NZE speaker in the listeners' expectation.

5.5.3 Jack

Jack is a 26-year-old male near-native speaker of English, a German L1 speaker. He began his formal study of English at the age of 12, in his 'English as a foreign language' classes at school. He visited England for 2-3 weeks as an adult, and he moved to New Zealand 6 months before the study. He reported using English only 40-50% of the time as he spoke German at home with his German wife and at the university with German office-mates.

The monophthongs produced by Jack are less NZE-like compared to Kahui and Emily (Figure 5.13): although GOOSE is very front and TRAP is raised, TRAP and DRESS lack distinction, KIT is high and front, NURSE is mid central, LOT is fronter, THOUGHT is lower, and STRUT is higher compared to the NZE ideal. Reflective of this, Jack passed for a NS of NZE only 2.5% of the time (much less than Kahui and Emily) and for a native speaker of a different variety of English 30.3%. Together with other potential deviations from the listeners' NZE ideal, the quite non-target-like vowel space resulted in rare cases of passing for a NZE speaker for Jack. However, despite his not remembering ever being taken for a NES, over 30% of listeners believed he was a NS of another variety of English in this experimental setting. Because of the small number of NS of NZE guesses, there were no examples provided by listeners to support their identification, and the analysis of lexical examples is not conducted for Jack.

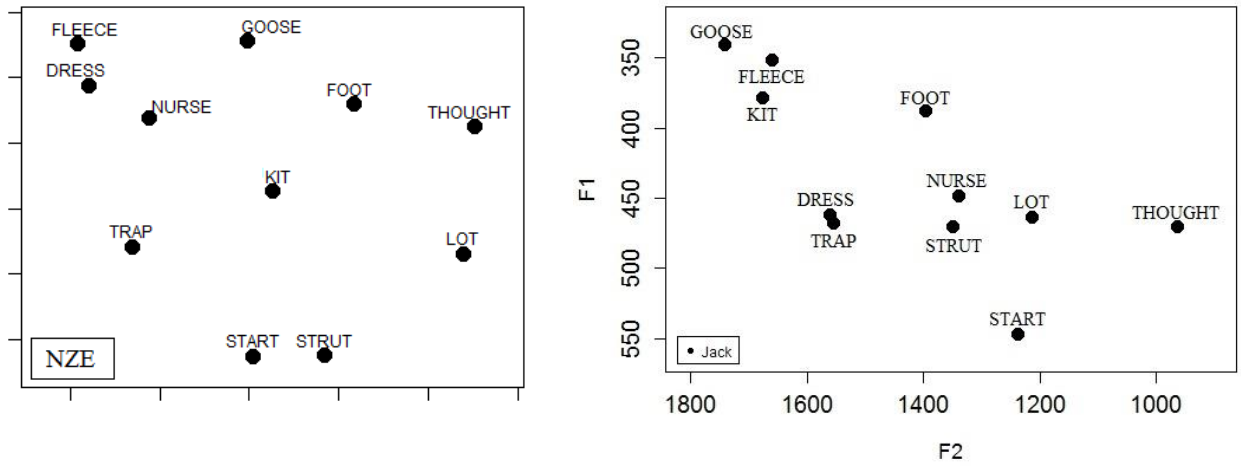

Figure 5.13: Vowel spaces of NZE (left panel) and L1 German speaker Jack (right panel)

Overall, these case studies illustrate the relationship between passing for a native speaker in production and perception. The speakers whose vowel spaces were more similar to the ideal NZE vowel space passed for a NS of NZE more often than the speaker with more differences did. The importance of target-likeness in salient features is highlighted in the listeners' lexical examples of native-likeness or non-native-likeness. The next chapter will compare and contrast the variation by setting found in production and perception models and discuss the implications for passing for a native speaker.

Chapter 6 : Discussion and Conclusion

6.1 Summary of results

In this section I return to the eleven specific research questions set out in the introduction of this thesis by theme.

6.1.1 Patterns in production

1. Do L2 speakers use differences between L1 and L2 vowel systems for situational style-shifting?
2. Does L2 speakers' style-shifting use the differences between L1 and L2 systems as a continuum as opposed to a binary choice?
3. Do speakers of different language backgrounds style-shift differently?
4. Do male and female L2 speakers style-shift differently?

The production study investigated style-shifting across the family, services, and university settings in Korean and German L1 speakers. Both L1 groups were found to vary in their production of several vowels, which I explain by applying the audience design (Bell, 1984) and identity construction accounts (Eckert, 2000). I had predicted the family setting to be most L1-like with an L1-related referee and topic and non-L2-related audience, followed by the university setting with an L2-related referee and topic and non-L2-related audience, and the services setting to be most L2-like of the three with L2-related audience, referee, and topic. The attested variation supported this hypothesis with the speakers style-shifting in production of the vowels of interest on an accentedness continuum between more L1-like and more L2-like in the three settings rather than two extremes.

The variation across the two L1 groups was not identical. The Korean group showed the most variation with significant differences found in three of the explored vowels across two or three settings. The German group, on the other hand, exhibited less variation with significant differences in only one of the vowels between two settings. The available data do not allow me to explain this finding, but I consider some of the reasons for such a result. For example, the

Korean L1 speakers have a longer length of residence in New Zealand as a group compared to the German L1 speakers. This means that Korean speakers had had a longer time to get familiar with NZE and develop their sociolinguistic styles. The variation was uniform across males and females.

6.1.2 Patterns in perception

The perception experiments revealed several significant effects.

5. Is there an effect of recording setting on perceived accentedness of a NNES?

First, recording setting was found to have a significant effect on perceived accentedness such that in Experiment 1 the clips in the services setting (family setting was significant at p=0.1) and in Experiment 2 the clips in the services and family settings (for female speakers) were found to be significantly less accented compared to the clips in the university setting which partially supported Piller (2002). In the former experiment, setting participated in an interaction with proficiency with no effect of setting on accentedness for lower proficiency speakers. I argued that this may be reflective of variation present in higher proficiency speakers only. In the latter, an interaction with speaker sex predicted no significant effect of setting for males and an interaction with listener L2 knowledge showed a significantly lower accentedness score in the friends and services settings for listeners with L2 knowledge. I argued that the interaction with speaker sex may be due to listener expectation in regards to the association between speaker sex and certain topics and the interaction with listener L2 knowledge suggests that listeners with and without L2 knowledge may be relying on different cues in their perception and accentedness rating.

6. What is the effect of availability of visual information for Asian NNESs in an accentedness perception task?
7. What is the effect of availability of visual information for Caucasian NNESs in the same accentedness perception task?
8. Will these effects for Asian and Caucasian NNESs be better predicted by reverse linguistic stereotyping or an audiovisual mismatch?

Moreover, Experiment 3 investigated the effect of ethnicity on accentedness perception and revealed that Caucasian NNESs were rated significantly more accented in the audiovisual condition and less accented in the video condition compared to the audio only one while Asian NNESs' accentedness was not judged significantly different among the three conditions. I explained this finding by an effect of an audiovisual mismatch (McGowan, 2011) between the speakers' actual production and listener expectation in regards to the speakers' accentedness based on their ethnicity.

6.1.3 Patterns in passing for a native speaker in experimental conditions

9. What is the variation in NNESs' passing for a NS of different English dialects?
10. What are some factors that contribute to a successful passing performance?
11. What are some of the elements that listeners notice in the input when a speaker succeeds or fails at passing?

The results of Experiment 2 suggest that there is a lot of variation in passing for a native speaker. First of all, NNESs can pass for both NSs of the same dialect and NSs of a different dialect. The patterns in passing for a native speaker of the same or different dialect suggest that, while some speakers may be using a 'mixed' accent as their target or an intermediate step in accent acquisition, some do not incorporate features of other dialects and are more commonly judged to be a NNES or a NS of the same dialect. It was tentatively suggested that the incorporation of certain stereotypical features of other dialects can benefit passing for a NS of a different dialect.

This study also compared the speakers' self-reports of passing to those in Piller (2002) and found certain regularities: the speakers in both studies believed that service encounters and communication with strangers was conducive to passing. The statistical analysis of passing in the four different settings supported Piller's (2002) claims that certain environments can be conducive to passing. Unlike qualitative predictions, however, service encounters were not judged to be significantly more native-like compared to the university setting.

The listeners' comments revealed rater reliance on a number of linguistic and extra-linguistic factors. The linguistic ones included mention of segments, suprasegmentals, grammar, and vocabulary. The analysis of listeners' use of examples (imitation) highlighted the importance of salient elements. The extralinguistic comments revealed rater consideration of socio-cultural information in the task. The results of this study suggest that passing for a native speaker is a highly variable phenomenon with many speaker-, listener-, and situation-dependent factors affecting it. The next section compares and contrasts within-speaker variation in production and perception of NNESs.

6.2 General discussion

6.2.1 Variation in production

This study is the most thorough investigation of situational style-shifting in L2 speakers to date, comparing shifts in a number of monophthongs in the speech of 12 NNESs of two different language backgrounds. The found incremental variation on the accentedness continuum from more L1-like to more L2-like production suggests that L2 speakers can use this accentedness continuum for sociolinguistic positioning on top of the sociolinguistic variation found in the L1 community, a Type 3 variation. This finding highlights the role of L2 speakers as independent and creative users of a language who can employ resources unavailable to its L1 speakers and underlines the importance of regarding L2 speakers as such. Moreover, accounts of L1 variation, such as audience design (Bell, 1984) and identity construction (Eckert, 2000) were fruitfully applied to L2 sociolinguistic variation in this thesis, which once more puts L1 and L2 speakers on a similar level. This study exemplifies how sociolinguistic tools, most often used to study of L1 variation, can be successfully applied to L2 variation.

6.2.2 Variation in perception

The role of listener expectation in perception surfaces in this thesis and confirms earlier observations of its profound effect on foreign-accented speech perception (Lindemann & Subtirelu, 2013). Reverse-linguistic stereotyping (Rubin, 1992) is intricately connected with

listener expectation and predicts that assumed speaker-related social information (e.g., ethnicity) will influence perceived phonetic information (e.g., foreign accentedness); that is listener expectation to hear accented speech when they see an Asian speaker will effect this auditory illusion even when the speech sample is standard-accented. Audiovisual mismatch (McGowan, 2011), on the other hand, predicts that a mismatch between auditory and visual information can affect perceived accentedness. The results of my study supported the audiovisual mismatch effect; however, these two accounts need not contradict each other as the mismatch between auditory and visual information can only be defined as a mismatch due to expectation of certain perceptual conditions. That is an Asian face and standard-accented speech can only be considered a mismatch if there is an expectation of an Asian face appearing with foreign-accented speech. One may attempt to extrapolate the mismatch / incongruence effect between expected and perceived information beyond audiovisual data.

The experiments investigating the effect of setting on accentedness perception have also highlighted the role of listener expectation and experience supporting previous literature on speaker- and stimulus- independent factors (Lindemann & Subtirelu, 2013; Levi et al., 2007). My experiments found that the services and the family settings were more conducive to a less accented rating compared to the university setting. From the point of view of the listener, there may be an expectation to hear standard-accented speech in the university setting with its scientific topics and technical vocabulary. Hearing foreign-accented speech might have constituted a mismatch or incongruity effect in that setting then. I have also attempted to use a mismatch effect and expectations argument to explain speaker sex by setting and listener L2 knowledge by setting interactions (see Section 4.2.2).

Literature on L1 linguistic behavior has often used expectations which are formed by previous experience to explain variation in multiple domains (e.g., Hay, Warren et al., 2006; Niedzielski, 1999 discussed above). Weatherholtz, Walker, Melvin, Royer, & Clopper (2014) argued that recent experience with and dialect priming influenced intelligibility of that dialect in noise. Fine, Jaeger, Farmer, & Qian (2013) found that syntactic comprehension is affected by expectations based on, for example, language statistics. Nass & Brave (2005) reported on a number of different studies investigating machine voice perception and concluded that incongruous voice characteristics and various types of information such as personality, and consequently, an inconsistency between people's expectations and perceived speech, affect

people's behavior. The results presented in this thesis suggest that similar theoretical and experimental tools can be used to discuss perception of both L1 and L2 speech.

6.2.3 Relationship between production and perception

Communication is a joint performance between the speaker and the listener. If the speaker style-shifts in a certain manner, it is important to know whether it is salient for the listener and if variation in production reflects the variation in perception. In this series of accentedness studies, I have found variation in both production and perception as summarized in Section 6.1. I have excluded the friends setting from discussion in the production study completely, and although I retained it in the discussion of the perception study, I want to be cautious in interpreting the results pertaining to it as there was much variation in topic, audience, and flow of individual friends encounters.

In the remaining settings, the variation by setting in production partially matched the variation by setting in perception. My findings support Piller's (2002) claims about the services setting. The services setting was most native-like in the speakers' production of the vowels and the listeners' assessment of accentedness. Speakers in this setting often employed formulaic expressions and high frequency words. It is also a very common, highly practiced situation with clear, defined roles for the parties involved. This relative lighter accentedness in production and perception would, presumably, make it easier to pass for a native speaker. This is supported by the speakers' examples of their passing performances in service situations and other first encounters with strangers (Section 5.2); however, the difference in passing between the university and services settings in the experiment did not reach significance (Section 4.2.3).

The inter-relationship between the family and university settings in production and perception is not exact. Korean L1 participants were more native-like in their production in the university setting compared to the family setting while German L1 speakers' production was not different between the two settings. In perception, on the other hand, the university setting was rated as more accented for females of both language groups (Experiment 2) or no significant difference at $p<0.05$ was found (Experiment 1). Purnell (2010) argues that the mapping of acoustic and perceptual cues is not exact, and Munro & Derwing (2015) also note that acoustic measures do not always correlate with perception ratings, so while the speakers may be

signalling nativeness in some elements, if the listeners are focusing on a different set of elements, the signal may not be noticed. Additionally, I argued that topic and listener expectations about the speaker on a given topic may be an extralinguistic factor influencing accentedness perception.

Speaker sex was found to participate in a significant interaction with setting in Experiment 2, but neither its main effect nor their interaction reached significance in the production study. This suggests that the significant interaction may be the result of listener factors, that males and females may style-shift differently on features that I did not analyse (e.g., consonants), or the production study did not have enough participants and statistical power to detect a significant effect. A production experiment analysing more features with a larger number of speakers may help to clarify this.

6.2.4 Passing for a native speaker

Passing is an aspect of social behavior influenced by many factors. In this thesis I link it with accentedness in production and perception and assume that lighter accentedness in production and perception is correlated with successful passing for a native speaker. Based on the results of the production and perception studies and the speakers' self-reports, I argue that the same speaker is more likely to have a successful passing performance in a short encounter with native-speaking strangers. This finding supports Piller's (2002) claims that short service encounters are conducive to passing.

Piller's (2002) claims about a facilitative effect of communication with friends may only be partially supported through re-interpreting the family setting as communication with friends. The friends setting was excluded from the production analysis because of the variation in topic and audience in the self-recordings, and it was not judged less accented compared to the university setting in my perception experiments. When the participants in Piller (2002) were mentioning communication with friends, they probably had a general effect of audience in mind rather than a more literal passing for a native speaker because passing can only refer to communication with strangers when the real identity of the speaker is not known. By definition, one cannot pass for a NS to a friend who is familiar with the speaker's background. I find some support for Piller's claims if I re-interpret the family setting in my experiments as a 'friends'

setting because of a familiar environment, topic, and interviewer to my speakers. The production in the family setting was only significantly less native-like for Korean L1 speakers, and in perception, the results of my Experiment 2 suggest that it was only the female speakers who were judged significantly more accented in the university setting compared to the family setting. I argued that listener expectations play an important role in passing.

Listener expectation may also have an effect on passing in relation to visual cues. Experiment 3 investigated the effect of a non-linguistic variable, namely ethnicity, on perceived accentedness. Ethnicity and other visual factors, such as clothing, have often been linked to the phenomenon of passing. Pattinson (2010), as briefly discussed in Chapter 2, described what visual factors made it easier for British nationals to pass easier for a French person during WWII: stereotypical 'French looks' (eye and hair color, height, etc.) and clothing (no tweed jacket or plus-fours). Marx (2002) started to avoid running shoes, men's jeans, and T-shirts in order to pass. Some of my Korean participants believed that their ethnicity prevented them from passing for a native speaker. In the accentedness experiment, however, the Korean speakers were rated similarly in audio only and audiovisual conditions. It is the German speakers that were rated significantly more accented in the audiovisual condition compared to audio only. However, this study explored the effect of ethnicity on accentedness, and its effect on passing for a native speaker may well be different.

6.3 Implications

This thesis aimed to explore some aspects of the social meaning of accent and the phenomenon of passing for a native speaker. It questioned the assumption of the accent being a learner's 'curse' and highlighted the social potential thereof. The production study was an exploration into within-speaker variation in the production of target vowels by two L1 groups of L2 speakers of English. The participants style-shifted on an accentedness continuum between more and less L2-like production of vowels in three settings which differed in topic and/or audience. This finding indicates that NNSs can use L2 variants on an accentedness continuum from more to less native-like for sociolinguistic positioning (something I call Type 3 variation) and are not limited to (arguably) uncontrolled interspeaker accentedness variation or intraspeaker variation mirroring NS patterns of sociolinguistic variation. This highlights the NNS relative agency and creativity in

their engagement with an L2 and allows them to be seen as independent users who appropriate the language rather than 'language borrowers' who have no influence over it.

The observed within-speaker variation in production among settings may have a number of practical implications. First, future production studies may benefit from a tighter control of topic and audience. For L2 pedagogy in general and pronunciation teaching in particular, the results suggest that learner exposure to more L2-related topics and audiences may be facilitative in making more L2-like forms available to the learner. The use of a variety of topics and audiences may also be beneficial in helping the learners explore different identities and extend their production repertoire. At the same time, regarding native-like forms as the only valid target may be questioned by both teachers and learners as the continuum between more and less native-like forms may be employed by L2 speakers for sociolinguistic and identity construction purposes.

The perception experiments highlighted the importance of listener expectation and listener-dependent factors. The effect of listener expectation is broadly in line with usage-based models (Pierrehumbert, 2003) as our current behavior is influenced by expectations based on previous experiences. The results of Experiment 3 suggest that an ethnicity effect may be found for Caucasian speakers as well as Asian speakers, as has been highlighted in previous research. Whereas we may be better aware of stereotyping of minority ethnicities and its effects on people's judgments, we might not be aware of an adverse effect of a majority ethnicity on perceived accentedness to the same extent. Based on the findings of Experiment 2, I tentatively suggest that listener expectation as to the speaker sex and topic congruence may also have an effect on speech perception. Such listener-related effects may have implications for L2 assessment. High-stakes decisions are often based on subjective impressions of a speaker's language ability. At the very least, professionals working in the field of L2 assessment must be aware of listener factors potentially having an effect on listener judgments.

The results suggest that focusing on passing as a purely inter-speaker phenomenon is too simplistic. The speakers' self-reports and the results of the quantitative analysis suggest that certain settings may be more conducive to passing: first encounters with strangers and communication with a friend in a comfortable environment. The discussion of passing for a native speaker of different varieties of English further suggests that speakers may try to avoid passing for a native speaker of the same dialect and instead aim for a 'mixed' accent allowing

them to pass for a native speaker of different dialects. The discussion of salient features in the speakers' production to listeners and their effect on passing suggests that native-like production of particular elements (e.g., vowels, slang) may be more important for a passing performance and suggests more attention paid to such elements in the L2 classroom.

6.4 Future directions

The results of this thesis suggest many intriguing questions still left to tackle. For example, the production data set did not allow for exploration of such factors as proficiency and length of residency, and although differences between the L1 groups were found, it is not certain that they should be attributed to the L1 and not other factors. The speakers' variation was examined from the point of view of their production and other people's perception of their production; variation in NNESs' perception of other people's speech remains an open avenue of research.

I have found variation in the production of vowels in the speakers' L2 and linked it with changes in audience and topic. Whereas I found a continuum in the production of L1- and L2-like elements, the presence of a phonetic continuum as opposed to a continuum of probabilities has not been confirmed and should be investigated further. Audience design and identity construction will also predict variation in the speakers' production of their L1. One might expect to find variation in the production of L1 vowels on the continuum from more to less L2-like depending on the topic and / or audience.

The naturalistic spontaneous speech of NNESs used in this thesis is definitely an advantage that allowed me to explore more realistic production and perception; however, it comes with certain challenges like little control over the production content. On the other hand, reading passages allow researchers to control for the exact words produced by the speaker at the expense of naturalness. However, for future research attempting to make connections between production and perception, I intend to use reading passages making direct comparisons easier. For a perception study, the matched-guise technique (Lambert et al., 1960) can help to exercise even more control. Acoustic manipulations of segments in a clip will help to study salience and noticing with a more robust method.

Passing for a native speaker is one extreme of accent production; having a noticeable stereotypical accent which is easily recognized – the other. A complex study of stereotypical

accents can shed light on how good listeners are at recognizing them, how they group them, and how they evaluate them. Perceptual dialectology has provided us with numerous interesting findings (e.g., see Preston, 1999, and Bradlow, Clopper, Smiljanic & Walter, 2010); perceptual accentology can prove prolific as well.

The ethnicity experiment can be extended with more speaker and listener groups and conditions. As listeners' expectations are representative of their past experiences and of societal stereotypes, the current findings may only be applicable to societies with a similar demographic distribution; therefore, it would be interesting to replicate this study in a different setting (e.g., Hong Kong). In the same setting, quantification of listener experience with Asian NESs and Caucasian NNESs may allow to explore the effect of such experience on accentedness perception ratings in more detail. Future research using Asian NES and NNES listeners may help to clarify whether there is an effect of listener ethnicity on perception of foreign accented speech produced by Asian and Caucasian speakers.

6.5 Conclusion

This thesis was an exploration into within-speaker variation in NNESs' production, perception, and, ultimately, passing for a native speaker of English. In Chapter 3 the analysis concerned the L2 speakers' production of several vowels in several recording settings differing in topic and audience. The results display that NNESs can use the distance between L1 and L2 systems for sociolinguistic positioning. Chapter 4 discussed the factors that were found to influence the speakers' perceived accentedness, with a particular emphasis on recording setting and ethnicity. Chapter 5 focused on passing for a native speaker; it also differentiated and quantified NNESs' passing for a native speaker of the same and different dialect as the listener and discussed the listeners' noticing and commenting on linguistic features in the input.

This thesis demonstrates the benefits of using naturalistic data in speech production research as well as combining production and perception analysis of variation and quantitative and qualitative data. I also advocate for the focus on NNESs as independent L2 users with agency and creativity in sociolinguistic variation, not prisoners of their non-native-likeness and accent. This suggests that sociolinguistic methods can be successfully applied to the study of within-speaker variation in NNESs. This study, then, is an example of a fruitful combination of

the quantitative and qualitative methods of second language acquisition and variationist sociolinguistic literature.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition,* 30(4), 481-509.

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and native-likeness in a second language: listener perception versus linguistic scrutiny. *Language Learning, 59*(2), 249-306.

Adamson, H. D., & Regan, V. M. (1991). The acquisition of community speech norms by Asian immigrants learning English as a second language. A preliminary study. *Studies in Second Language Acquisition, 13*(1), 1-22.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The Relationship Between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentais, Prosody, and Syllable Structure. *Language Learning, 42*(4), 529-555. doi: 10.1111/j.1467-1770.1992.tb01043.x

Baayen, R. H., Davidson, D. J., & Bates, D. M. (1998). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Baayen, R. H., Piepenbrock, R. & Gulikers, L. (1995). *The CELEX lexical database (CD-ROM).* Philadelphia: PA: Linguistic Data Consortium, University of Pennsylvania.

Baker, W., & Trofimovich, P. (2005). Interaction of Native- and Second-Language Vowel System(s) in Early and Late Bilinguals. *Language and Speech, 48*(1), 1-27.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278.

Barras, C., Geoffrois, E., Wu, Z., & Liberman, M. (2001). Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication, 33*(1-2), 5-22.

Beebe, L. M. (1980). Sociolinguistic Variation and Style Shifting in Second Language Acquisition. *Language Learning, 30*(2), 433-445.

Beebe, L. M., & Zuengler, J. (1983). Accomodation theory: An explanation for style shifting in second language dialects. In N. Wolfson, E. Judd & E. Hatch (Eds.), *Sociolinguistics and Language Acquisition*. Rowley: Newbury House.

Beinhoff, B. (2008). Looking for the "real" native speaker: The perception of native and non-native English accents by non-native speakers of English. In E. Waniek-Klimczak (Ed.), *Issues in Accents of English* (pp. 120-139). Newcastle upon Tyne: Cambridge Scholars Publishers.

Bell, A. (1984). Language Style as Audience Design. *Language in society, 13*(2), 145-204.

Block, D. (2006). Identity in applied linguistics. In T. Omoniyi & G. White (Eds.), *The sociolinguistics of identity* (pp. 34-49). London, New York: Continuum.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version Version 5.1.22). Retrieved from http://www.praat.org/

Bongaerts, T. (1999). Ultimate Attainment in L2 Pronunciation: The Case of Very Advanced Late L2 Learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133-159). Mahwah: Lawrence Erlbaum Associates, Publishers.

Bradlow, A., Clopper, C., Smiljanic, R., & Walter, M. A. (2010). A perceptual phonetic similarity space for languages: Evidence from five native language listener groups. *Speech Communication*, 52(11), 930-942. doi: 10.1016/j.specom.2010.06.003

Bucholtz, M., & Hall, K. (2005). Identity and interaction: a sociocultural linguistic approach. *Discourse Studies, 7*(4-5), 585-614.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences, 11*(12), 535-543.

Campbell-Kibler, K. (2007). Accent, (ING), and the social logic of listener perceptions. *American Speech, 82*(1), 32-64.

Childs, J. A. (2012). *Suprasegmental features and their classroom application in pronunciation instruction*. (Master of Arts), The University of Texas at Austin.

Coupland, N. (1984). Accommodation at work: Some phonological data and their implications. *International Journal of the Sociology of Language, 46*, 49-70.

Davies, A. (2002). *The Native Speaker: Myth and Reality*. Clevedon: Multilingual Matters LTD.

Derwing, T.M., & Munro, M.J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition, 19*(1), 1-16.

Dolgova Jacobsen, N. (2008). "Identity [iz] a [djifikəlt] question": A Variationist Analysis of the Relationship Between L1 Features and Ethnic Identity in the Speech of Russian Learners of English. *Georgetown Working Papers in Language, Discourse, & Society, 2*, 1-28.

Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass, 4*(7), 473-480.

Drummond, R. (2011). Glottal variation in /t/ in non-native English speech. Patterns of acquisition. *English World-Wide, 32*(3), 280-308.

Eckert, P. (2000). *Linguistic variation as social practice: the linguistic construction of identity in Belten High* (Vol. 27.). Malden, Mass: Blackwell Publishers.

Feinberg, J. (2014). WordleTM. Available from http://www.wordle.net. Accessed August 2015.

Fine, A. B., Jaeger, T. F., Farmer, T., and Qian, T. 2013. Rapid expectation adaptation during syntactic comprehension. *PLoS ONE 8*(10), e77661. doi:10.1371/journal.pone.0077661

Fischer, J. L. (1958). Social influences on the choice of a linguistic variant. *Word, 14*, 47-56.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America, 91*(1), 370-389.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78-104.

Fromont, R., & Hay, J. (2012). *LaBB-CAT: an Annotation Store*. Paper presented at the Australasian Language Technology Association Workshop.

Garrett, P. (2010). *Attitudes to language*. Cambridge: Cambridge University Press.

Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech, & A. McEnery (Eds.), *Corpus Annotation: Linguistic Information from Computer Text Corpora* (pp. 102-121). London: Longman.

Gatbonton, E., Trofimovich, P., & Magid, M. (2005). Learners' ethnic group affiliation and L2 pronunciation accuracy: A sociolinguistic investigation. *TESOL Quarterly, 39*(3), 489-511.

Giles, H. (2001). Couplandia and beyond. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 211-219). Cambridge, UK: Cambridge University Press.

Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation* (Vol. 7). London: Academic Press.

Gluszek, A., & Dovidio, J. F. (2010). The Way They Speak: A Social Psychological Perspective on the Stigma of Nonnative Accents in Communication. *Personality and Social Psychology Review, 14*(2), 214-237.

Gnevsheva, K. (2013). Problematising the age of acquisition: effect of current language use and language dominance on L2 sound production. *New Zealand English Journal, 27*, 7-19.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics, 48*(4), 865-892.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review, 23*(3). doi: 10.1515/tlr.2006.014

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*(4), 458-484. doi: 10.1016/j.wocn.2005.10.001

Hayes-Harb, R., & Hacking, J. (2015). Beyond rating data: What do listeners believe underlies their accentedness judgments? *Journal of Second Language Pronunciation, 1*(1), 43-64.

Hayes-Harb, R., & Watzinger-Tharp, J. (2012). Accent, Intelligibility, and the Role of the Listener: Perceptions of English-Accented German by Native German Speakers. *Foreign Language Annals, 45(2),* 260-282.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis. A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition, 16*, 73-98.

Isaacs, T., & Trofimovich, P. (2011). Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of second language speech. *Applied Psycholinguistics, 32*(1), 113-140.

Johnson, J. S., & Newport, E. L. (1989). Critical Period Effects in Secon Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language *Cognitive Psychology, 21*(1), 60-99.

de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of Speaking Proficiency. *Studies in Second Language Acquisition, 34*(01), 5-34. doi: 10.1017/s0272263111000489

Kachru, B. (1992). *The Other Tongue: English across cultures*. University of Illinois Press.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental Measures of Accentedness and Judgments of Language Learner Proficiency in Oral English. *The Modern Language Journal, 94*(4), 554-566.

Kerswill, P., & Williams, A. (2002). Dialect recoginition and speech community focusing in new and old towns in England: The effects of dialect levelling, demography and social

networks. In D. Long & D. R. Preston (Eds.), *Handbook of Perceptual Dialectology* (Vol. 2, pp. 173-204). Amsterdam: John Benjamins Publishing Company.

Kraut, R., & Wulff, S. (2013). Foreign-accented speech perception ratings: a multifactorial case study. *Journal of Multilingual and Multicultural Development, 34*(3), 249-263. doi: 10.1080/01434632.2013.767340

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Lambert, W.E., Hodgson, R.C., Gardner, R.C., & Fillenbaum, S. (1960). Evaluational reactions to spoken language. *Journal of Abnormal and Social Psychology, 60:* 44-51.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of Second Language development*. Oxford: Oxford University Press.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America, 121*(4), 2327-2338.

Li, X. (2010). Sociolinguistic Variation in the Speech of Learners of Chinese as a Second Language. *Language Learning, 60*(2), 366-408.

Lindemann, S., & Subtirelu, N. (2013). Reliably biased: The role of listener expectation in the perception of second language speech. *Language Learning, 63*(3), 567-594.

Lippi-Green, R. (1997). *English with an Accent : Language, Ideology and Discrimination in the United States*. Hoboken: Routledge.

Long, M. H. (1990). Maturational Constraints on Language Development. *Studies in Second Language Acquisition, 12*(3), 251-285.

MacKenzie, L., & Turton, D. (2013). *Crossing the Pond: Extending automatic alignment techniques to British English dialect data*. Paper presented at the New Ways of Analyzing Variation.

Magen, H.S. (1998). The perception of foreign-accented speech. *Journal of Phonetics, 26*(4), 381-400.

Magnusson, J. E., & Stroud, C. (2012). High proficiency markets of performance: A sociocultural approach to native-likeness. *Studies in Second Language Acquisition, 34*(2), 321-345. doi: 10.1017/s0272263112000071

Major, R. C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah: Lawrence Erlbaum Associates, Publishers.

Major, R. C. (2004). Gender and stylistic variation in second language phonology. *Language Variation and Change, 16*, 169-188.

Marx, N. (2002). Never quite a "Native Speaker": accent and identity in the L2 - and the L1. *The Canadian Modern Language Review, 59*(2), 264-281.

McGowan, K. B. (2011). *The Role of Socioindexical Expectation in Speech Perception.* (Dissertation/Thesis), ProQuest, UMI Dissertations Publishing.

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

McKenzie, R. M. (2015). The sociolinguistics of variety identification and categorisation: Free classification of varieties of spoken English amongst non-linguist listeners. *Language Awareness, 24*(2), 150-168.

Miller, N. (2010). Foreign accent syndrome. In C. Llamas & D. Watt (Eds.), *Language and identities* (pp. 67-75). Edinburgh: Edinburgh University Press.

Milroy, L. (1987). *Language and Social Networks*. New York: Blackwell.

Moyer, A. (2004). *Age, Accent, and Experience in Second Language Acquisition: An Integrated Approach to Critical Period Inquiry*. Clevedon: Multilingual Matters Limited.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: acoustic measurements and accentedness ratings. *Language and Speech*, 36(1), 39-66.

Munro, M. J. (2008). Foreign Accent and Speech Intelligibility. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 193-218). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Munro, M. J. & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73-97.

Munro, M. J. & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century: A point of view. *Journal of Second Language Pronunciation, 1* (1), 11-42.

Munro, M.J., Derwing, T.M., & Burgess, C.S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication, 52*, 626-637.

Munro, M. J., Derwing, T. M., & Holtby, A. K. (2011). *Evaluating individual variability in foreign accent comprehension.* Paper presented at the The 3rd annual pronunciation in second language learning and teaching conference, Iowa State University.

Nass, C. & Brave, S. (2005). *Wired for Speech: How voice activates and advances the human-computer relationship*. Cambridge: The MIT Press.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*(1), 62-85.

Pattinson, J. (2010). 'Passing unnoticed in a French crowd': The passing performances of British SOE agents in Occupied France. *National Identities, 12*(3), 291-308.

Pierrehumbert, J. (2003). Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. *Language and Speech, 46*(2-3), 115-154.

Piller, I. (2002). Passing for a native speaker: identity and success in second language learning. *Journal of Sociolinguistics, 6*(2), 179-206.

Preston, D.R. (1996). Whaddayaknow?: The Modes of Folk Linguistic Awareness. *Language Awareness, 5*(1), 40-74.

Preston, D.R. (Ed.) (1999). *Handbook of Perceptual Dialectology*. Amsterdam: John Benjamins Publishing Company.

Prikhodkine, A. (2012). *"Tell me your name and I'll say if you speak with an accent". Effects of proper names' ethnicity on listener expectations.* Paper presented at the 'Folk linguistics': Language from speakers' perspective, St. Petersburg.

Purnell, T. (2010). Phonetic detail in the perception of ethnic varieties of US English. In D. R. Preston & N. Niedzielski (Eds.), *A Reader in Sociophonetics* (pp. 289-326). New York: Walter de Gruyter.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rampton, B. (2011). Style in a Second Language. *Working Papers in Urban Language & Literacies*, 65, www.kcl.ac.uk/ldc.

Regan, V. (1996). Variation in French interlanguage. In R. Bayley & D. R. Preston (Eds.), *Second Language Acquisition and Linguistic Variation* (pp. 177-201). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Rehner, K., Mougeon, R., & Nadasdi, T. (2003). The learning of sociolinguistic variation by advanced FSL learners: The case of Nous versus On in immersion French. *Studies in Second Language Acquisition, 25*, 127-156.

Rindal, U., & Piercy, C. (2013). Being 'neutral'? English pronunciation among Norwegian learners. *World Englishes, 32*(2), 211-229. doi: 10.1111/weng.12020

Rosenfelder, Ingrid; Fruehwald, Joe; Evanini, Keelan and Jiahong Yuan. 2011. FAVE (Forced Alignment and Vowel Extraction) Program Suite. http://fave.ling.upenn.edu.

Rubin, D. L. (1992). Nonlanguage Factors Affecting Undergraduates' Judgments of Nonnative English-Speaking Teaching Assistants. *Research in Higher Education, 33*(4), 511-531.

Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations, 14*, 337-353.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*, 421-436.

Schleef, E., Meyerhoff, M., & Clark, L. (2011). Teenagers' acquisition of variation: A comparison of locally-born and migrant teens' realisation of English (ing) in Edinburgh and London. *English World-Wide, 32*(2), 206-236.

Schmid, M. S., Gilbers, S. & Nota, A. (2014). Ultimate attainment in late second language acquisition: Phonetic and grammatical challenges in advanced Dutch–English bilingualism. *Second Language Research, 30*(2), 129-157.

Schneider, W., Eschman, A., and Zuccolotto, A. (2012). E-Prime User's Guide. Pittsburgh: Psychology Software Tools, Inc.

Sharma, D. (2011). Style repertoire and social change in British Asian English. *Journal of Sociolinguistics, 15*(4), 464-492.

Shuck, G. (2004). Conversational performance and the poetic construction of an ideology. *Language in Society, 33*, 195-222.

Statistics New Zealand. (2013). 2013 Census totals by topic. Retrieved December, 12, 2014 from http://www.stats.govt.nz/Census/2013-census/data-tables/total-by-topic.aspx

Statistics South Africa. (2012). Census 2011: Census in brief. Retrieved May, 11, 2015 from http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf

Strange, W., Bohn, O.-S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America, 115*(4), 1791-1807.

Subtirelu, N. C. (2015). "She does have an accent but ...": Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com. *Language in Society, 44*(1), 35-62. doi: 10.1017/s0047404514000736

Tarone, E. (1979). Interlanguage as chameleon. *Language Learning, 29*, 181-191.

Thomas, E. R., & Kendall, T. (2007). NORM: The vowel normalization and plotting suite Retrieved 06.11, 2013, from http://ncslaap.lib.ncsu.edu/tools/norm/

Thomas-Hunt, M. C. & Phillips, K. W. (2004). When what you know is not enough: Expertise and gender dynamics in task groups. *Personalilty and Social Psychology Bulletin, 30*(12), 1585-1598. doi: 10.1177/0146167204271186

University of Camridge. 2014. *HTK*. Retrieved from http://htk.eng.cam.ac.uk/

Vishnevskaya, G. (2008). Foreign accent: Phonetic and communicative hazards. In E. Waniek-Klimczak (Ed.), *Issues in accents of English* (pp. 235-251). Newcastle upon Tyne: Cambridge Scholars Publishers.

Walker, A. (2014). *Crossing oceans with voices and ears: Second dialect acquisition and topic-based shifting in production and perception.* (PhD), The Ohio State University.

Watanabe, Y. (2008). New Zealand attitudes towards foreign-accented English. *Te Reo, 51*, 99-127.

Watson, K. & Clark, L. (2013). How salient is the NURSE~SQUARE merger? *English Language and Linguistics, 17(2)*, 297-323.

Watson, K., Leach, H. & Gnevsheva, K. (submitted). Perceptual dialectology in northern England: exploring the effects of proximity and salience on accent identification.

Waylard, R.P. (2007). The relationship between identification and discrimination in cross-language perception: The case of Korean and Thai. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 201-220). Amsterdam: John Benjamins Publishing Company

Weatherholtz, K., Walker, A., Melvin, S., Royer, A., & Clopper, C. G. (2014). Effects of experience and expectations on adaptation to dialect variation in noise. Poster presented at *The 27th Annual CUNY Conference on Human Sentence Processing*, Columbus, OH, March 14-15th.

Wells, J. C. (1982). *Accents of English*. Vol. 3. Cambridge, Cambridge University Press.

Williams, A., Garrett, P., & Coupland, N. (1999). Dialect recognition. In D. R. Preston (Ed.), *Handbook of perceptual dialectology*. Amsterdam: John Benjamins Publishing Company.

Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics, 24*, 245-261.

Yi, H.-G., Phelps, J., E.B., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America, 134*(5), 387-393.

Yi, H.-G., Smiljanic, R., & Chandrasekaran, B. (2014). The neural processing of foreign-accented speech and its relationship to listener bias. *Frontiers in Human Neuroscience, 8,* 1-12. doi: 10.3389/fnhum.2014.00768

Young, R. (1988). Variation in the interlanguage hypothesis. *Studies in Second Language Acquisition*, 10, 281-302.

Zając, M. (2015). *Phonetic convergence in the speech of Polish learners of English*. (PhD), The University of Lodz.

Appendix A: Human Ethics Applications

Production study

Ksenia Gnevsheva

Department of Linguistics, Locke 210a

15 Jan.2013

CONSENT FORM

Word Choice in a Variety of Communicative Situations

I have read and understood the description of the above-named project. On this basis I agree to participate as a subject in the project, and I consent to publication of the results of the project with the understanding that anonymity will be preserved.

I understand also that I may at any time withdraw from the project, including withdrawal of any information I have provided.

I note that the project has been reviewed ***and approved*** by the University of Canterbury Human Ethics Committee.

NAME (please print): …….

Signature:

Date:

Te Kāhui Roro Reo • New Zealand Institute of
Language Brain & Behaviour

UC
UNIVERSITY OF
CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Information

You are invited to participate as a subject in the research project **Word Choice in a Variety of Communicative Situations**. The aim of this project is to see how the word choice is different when people speak on different topics.

Your involvement in this project will include a pre-recordings questionnaire, a post-recordings questionnaire or short audio-recorded interview, two audio-recorded interviews (at home and at work/University (the work one will be additionally video-recorded)) and two self- audio-recordings of a meeting with friends and a service-encounter resulting in a total of 4 recordings of about 10-15 minutes each.

The tasks are not difficult, and you might even find them fun, but it is true that some people might feel uneasy when being interviewed and audio- or video-recorded. This feeling usually goes away after a few minutes of being recorded. If you change your mind about participating in the project, you have the right to withdraw at any time, including withdrawal of any information provided without penalty.

The results of the project may be used for future research projects or published, but you are assured of the complete confidentiality of data gathered in this investigation: the identity of participants will not be made public without their prior consent. To ensure anonymity and confidentiality, the researcher will not use real names of the participants and will keep all identifiable information in a secure location.

The project is being carried out as a requirement for the degree of Doctor of Philosophy by Ksenia Gnevsheva under the supervision of Dr Kevin Watson, who can be contacted at [+64 3 364 2987] or by email: ksenia.gnevsheva@pg.canterbury.ac.nz; kevin.watson@canterbury.ac.nz. Ksenia and/or Kevin will be pleased to discuss any concerns you may have about participation in the project.

The project has been reviewed ***and approved*** by the University of Canterbury Human Ethics Committee.

Te Kāhui Roro Reo • New Zealand Institute of
Language Brain & Behaviour

UC
UNIVERSITY OF
CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Pre-Recordings Questionnaire

Word Choice in a Variety of Communicative Situations

Please, choose a pseudonym for yourself:

1. How old are you?
2. What is your gender?
3. What's the highest academic degree you have achieved or are studying towards?
4. What is(are) your native language(s)? If it's English, which dialect (New Zealand, American, etc.)?
5. Where were you born?
6. (If born outside of New Zealand) When did you move to New Zealand?

Ksenia Gnevsheva

Department of Linguistics, Locke 210a

15 Jan.2013

DEBRIEFING AND CONSENT FORM

Variation in Degree of Accentedness in Second Language English Speakers

The aim of this project is to see whether second language speakers of English exhibit a different degree of accentedness in different naturalistic environments and if different environments facilitate or hinder their ability to pass for a native speaker in communicative situations. You were not told of the true nature of the project because it could have influenced your linguistic behavior. If you are interested to learn about the findings of the study, please, let Ksenia know, and she'll provide you with a summary after the project is completed.

I have read and understood the description of the above-named project. On this basis I agree to participate as a subject in the project, and I consent to the following RESEARCH USE of the results of the project with the understanding that confidentiality will be preserved (Please, tick the boxes if you are comfortable with them. Strike out any that don't apply.):

□ I agree to audio/video recordings being played to research participants in future research studies.

□ I agree to transcript/audio/video recordings being used in teaching, public lectures, and presentations.

I understand also that I may withdraw from the project at this moment, including withdrawal of any information I have provided. I note that the project has been reviewed ***and approved*** by the University of Canterbury Human Ethics Committee.

NAME (please print): …….

Signature:

Date:

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Post-Recordings Questionnaire

Variation in Degree of Accentedness in Second Language English Speakers

Your pseudonym:

Please, answer the following questions:

1. List all the languages you speak and how well you speak them.
2. Which dialect of English do you consider your native one (New Zealand English, American English, etc.)?
3. What English-speaking countries have you travelled to/lived in and for how long?

Perception study

ETHICAL APPROVAL OF LOW RISK RESEARCH I INVOLVING HUMAN PARTICIPANTS REVIEWED BY DEPARTMENTS

Please read the important notes appended to this form before completing the sections below

1 RESEARCHER'S NAME: Ksenia Gnevsheva

2 NAME OF DEPARTMENT OR SCHOOL: Department of Linguistics

3 EMAIL ADDRESS: ksenia.gnevsheva@pg.canterbury.ac.nz

4 TITLE OF PROJECT: Accents of English

5 PROJECTED START DATE OF PROJECT: March 2014

6 STAFF MEMBER/SUPERVISOR RESPONSIBLE FOR PROJECT: Dr. Kevin Watson, Professor Jen Hay

7 NAMES OF OTHER PARTICIPATING STAFF AND STUDENTS:

8 STATUS OF RESEARCH: (pilot study, thesis, staff research – please include status of student researchers involved if this is a staff-led project)

Thesis

9 BRIEF DESCRIPTION OF THE PROJECT:
Please give a brief summary (approx. 500 words) of the nature of the proposal in lay language, including the aims/objectives/hypotheses of the project, rationale, participant description, and procedures/methods of the project:-

The aim of this research project is to investigate accent perception. The hypothesis is that the topic of the recording, speaker's language background, and listener's language background will have an influence on accent perception. In the experiment the participants will hear/watch a number of short clips collected under HEC approval Ref: HEC 2012/176. The speakers in the clips are from different regional and language backgrounds speaking on a variety of different topics.

The experiment will have several instantiations differing in the type of input participants will receive (audio only or audio + video) and how detailed the provided response should be. There are 4 different conditions: short Audio 1, long Audio 1, short Audio2, and Video. In short Audio 1 and long Audio 1 conditions participants will listen to 96 15-20 second clips. In the short condition the participants will be asked to rate the speaker in each clip on a scale from "First language speaker of English" to "Second language speaker of English". The task will take approximately 40 min. In the long condition, the participants will have an extra open-response task which will ask them to guess where the speaker is from and comment on their decision. The whole task will take approximately 1 hour. In both short Audio 2 and Video conditions participants will hear the same 24 15-20 second clips, but in the Video condition, participants will additionally see the video input associated with the audio. The task is to rate the speaker in each clip on a scale from "First language speaker of English" to "Second language speaker of English" and will take approximately 20 min.

At last, the listeners will be asked to complete a short anonymous questionnaire eliciting the information about their sex, age, education, language background, and experience with varieties of English. See attachment.

Participants will be 100 native New Zealand English speaking adults who may be recruited through ads and the researcher's social networks. Participants will receive between $10 as compensation for their time.

10 WHY IS THIS A LOW RISK APPLICATION?

Description should include issues raised in the Low Risk Checklist
Please give details of any ethical issues which were identified during the consideration of the proposal and the way in which these issues were dealt with or resolved.

This is a low risk application because participation is anonymous and no sensitive information about or from participants is collected. Participation in this project is designed to avoid causing any sort of physical, mental, or emotional stress or other risks. This project does not raise any issue of deception, threat, invasion of privacy, mental, physical or cultural risk or stress.

11 PROVIDE COPIES OF INFORMATION & CONSENT FORMS FOR PARTICIPANTS

These forms should be on University of Canterbury departmental letterhead. The name of the project, name(s) of researcher(s), contact details of researchers (and for PhD students, the supervisor), names of who has access to the data, the length of time the data is to be stored, that participants have the right to withdraw participation and data provided, and what the data will be used for should all be clearly stated. A statement that the project has been reviewed approved by the appropriate department and the UCHEC Low Risk Approval process should also be included.

Please see Attachments: Information, Ad, and Questionnaire.
In "Information", the instructions for the 4 different conditions are given in parentheses.

Please ensure that Section A (where appropriate), B and C below are all completed

Applicant's Signature: ..... Date .....

A SUPERVISOR'S DECLARATION FOR PhD RESEARCH:

1 I have made the applicant fully aware of the need for and the requirement of seeking HEC approval for research involving human participants.
2 I have ensured the applicant is conversant with the procedures involved in making such an application.
3 In addition to this form the applicant has individually filled in the full application form which has been reviewed by me.

Signed (Supervisor): ..... Date .....

B SUPPORTED BY THE DEPARTMENTAL/SCHOOL RESEARCH COMMITTEE:

Name .....

Signature: ..... Date .....

C APPROVED BY HEAD OF DEPARTMENT/SCHOOL:

Name .....

Signature: ..... Date .....

SUBMISSION OF APPLICATION:

- Please attach copies of any Information Sheet and Consent Form
- Forward two hard copies to: The Secretary, Human Ethics Committee, Okeover House
- Forward an electronic copy to: human-ethics@canterbury.ac.nz

NOTES ON PROCEDURE:

The Chair of the University of Canterbury Human Ethics Committee and two other Human Ethics Committee members will review this application.

In normal circumstances queries will be forwarded via email to the applicant within 7 days

If you are a PhD student, please include a copy of this form as an appendix in your thesis

ACTION TAKEN BY HUMAN ETHICS COMMITTEE:

☐ Added to PhD & Staff Low Risk Reporting Database

☐ Referred to University of Canterbury HEC

☐ Referred to another Ethics Committee – please specify:

……

REVIEWED BY: …….. (HEC Chair)

…….. (HEC Member)

…….. (HEC Member)

Date ……

NOTES CONCERNING LOW RISK REPORTING SHEETS

1. This form should only be used for proposals which are Low Risk as defined in the University of Canterbury Human Ethics Committee Principles and Guidelines policy document and which may therefore be properly considered and approved at departmental level and by the Chair and two members of the University of Canterbury Human Ethics Committee under Section 5 of that document.

2. Low Risk applications are:

   PhD thesis, pilot studies and staff research where the projects do not raise any issue of deception, threat, invasion of privacy, mental, physical or cultural risk or stress, and do not involve gathering personal information of a sensitive nature about or from individuals.

3. No research can be counted as low risk if it involves:

   (i) invasive physical procedures or potential for physical harm
   (ii) procedures which might cause mental/emotional stress or distress, moral or cultural offence
   (iii) personal or sensitive issues
   (iv) vulnerable groups
   (v) Tangata Whenua
   (vi) cross cultural research
   (vii) investigation of illegal behaviour(s)
   (viii) invasion of privacy
   (ix) collection of information that might be disadvantageous to the participant
   (x) use of information already collected that is not in the public arena which might be disadvantageous to the participant
   (xi) use of information already collected which was collected under agreement of confidentiality
   (xii) participants who are unable to give informed consent
   (xiii) conflict of interest e.g. the researcher is also the lecturer, teacher, treatment-provider, colleague or employer of the research participants, or there is any other power relationship between the researcher and the research participants.
   (xiv) deception
   (xv) audio or visual recording without consent
   (xvi) withholding benefits from "control" groups
   (xvii) inducements
   (xviii) risks to the researcher

   This list is not definitive but is intended to sensitise the researcher to the types of issues to be considered. Low risk research would involve the same risk as might be encountered in normal daily life.

4. Responsibility

   Supervisors are responsible for:
   Theses where the projects do not raise any issues listed below.

   Heads of Department are responsible for:
   (i) Giving final approval for the low risk application.
   (ii) Ensuring a copy of all applications are kept on file in the Department/School.

   NOTE: If the HOD is the applicant, then a senior member of staff and preferably also the department and/or school research committee should give final approval. The HOD is still responsible for (ii) above.

4. A separate low risk form should be completed for each research proposal involving human participants and for which ethical approval has been considered or given at Departmental level.

5. Two completed and signed Application forms, together with a copies of Information Sheets and/ or Consent Forms, should be submitted to the Secretary, Human Ethics Committee, Okeover House, as soon as the proposal has been considered at departmental level. Please also submit an electronic version to human-ethics@canterbury.ac.nz. .

6. The Information Sheet and Consent Form include the statement "This proposal has been reviewed and approved by the University of Canterbury Human Ethics Committee low risk process".

7. Please ensure the Consent Form and the Information Sheet are on University of Canterbury letterhead and have been carefully proof-read; the institution as a whole is likely to be judged by them.

9. The research must be consistent with the University of Canterbury Human Ethics Committee Principles and Guidelines. Refer to the appendices of the University of Canterbury Human Ethics Committee Principles and Guidelines for guidance on information sheets and consent forms.

10. Please note that if the nature, procedures, location or personnel of the research project changes after departmental approval has been given in such a way that the research no longer meets the conditions laid out in Section 5 of the Principles and Guidelines, a full application to the Human Ethics Committee must be submitted.

11. This form is available electronically at: http://www.canterbury.ac.nz/humanethics

CHECKLIST

Please check that your application/summary has discussed:

- Procedures for voluntary, informed consent
- Privacy & confidentiality
- Risk to participants
- Obligations under the Treaty of Waitangi
- Needs of dependent persons
- Conflict of interest
- Permission for access to participants from other individuals or bodies
- Inducements

In some circumstances research which appears to meet low risk criteria may need to be reviewed by the University of Canterbury Human Ethics Committee. This might be because of requirements of:

- The publisher of the research
- An organisation which is providing funding resources, existing data, access to participants etc.
- Research which meets the criteria for review by a Health and Disability Ethics Committee – see HRC web site.

If you require advice on the appropriateness of research for low risk review, please contact the Chair of the University of Canterbury Human Ethics Committee

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

UC UNIVERSITY OF CANTERBURY Te Whare Wānanga o Waitaha CHRISTCHURCH NEW ZEALAND

Information

You are invited to participate in an experiment that studies how people identify different regional and international accents of English. Your involvement in this project will include a questionnaire and the main part in which you will listen to (watch) short recordings and will be asked to rate the speakers on an accentedness scale, (guess where you think that person is from and comment on why you think so). The whole experiment will take about 40 (20) minutes (1 hour).

Participation is voluntary and you have the right to withdraw at any stage until you've submitted the completed questionnaire at the end of the experiment without penalty. The results of the project may be used for future research projects or published, but you may be assured of the complete confidentiality of data gathered in this investigation. To ensure anonymity and confidentiality, we do not ask for your contact details and, though we record your consent to participate, the signed forms are stored separately from the anonymised data.

The responses gathered will be kept on password-protected computers and servers for which only the primary investigator, supervisors and bona fide researchers have access. Since there is no information, which identifies participants, this data will be kept indefinitely to allow for future re-analysis. You may receive a copy of the project results by contacting the researcher at the conclusion of the project.

The project is being carried out as a requirement for the degree of Doctor of Philosophy by Ksenia Gnevsheva under the supervision of Dr Kevin Watson, who can be contacted by email: ksenia.gnevsheva@pg.canterbury.ac.nz; kevin.watson@canterbury.ac.nz. Ksenia and/or Kevin will be pleased to discuss any concerns you may have about participation in the project.

The project has been reviewed and **approved** by the Department of Linguistics and the University of Canterbury Human Ethics Committee Low Risk Approval process.

By signing below, you indicate that you have read and understood the description of the above-named project. On this basis, you agree to participate as a subject in the project and consent to publication of the results of the project. You understand that you may withdraw at any time.

NAME (please print): …….

Signature: Date:

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Help us with our research!

Is New Zealand English the first language you learnt to speak?

Are you over the age of 18?

Then you are invited to participate in an experiment that studies how people identify different regional and international accents of English.

In this experiment you will listen to short recordings and will be asked to rate the speaker on an accentedness scale and guess where they are from. The whole experiment will take up to 1 hour and you will be compensated to the value of **NZ$10** for your time.

If you are interested, please e-mail Ksenia to set up an appointment:

Ksenia.gnevsheva@pg.canterbury.ac.nz

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

UC UNIVERSITY OF CANTERBURY Te Whare Wānanga o Waitaha CHRISTCHURCH NEW ZEALAND

Questionnaire

Participant number:

1. How old are you?

2. Are you male or female?

3. What's the highest academic degree you have achieved or are studying towards?

4. List all the languages you speak and how well you speak them.

5. What English-speaking countries have you travelled to/lived in and for how long?

6. Where do you think the research assistant that you met earlier is from?

7. Any comments you would like to make about the experiment:

Appendix B: Mixed-effects model output for the production study models

Emily

Table B.1: Summary for model of KIT F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-1.378	0.692	181.960	-1.991	0.048	*
fol.env_ð	-0.239	0.731	181.960	-0.327	0.744	
fol.env_f	0.152	0.231	181.960	0.656	0.512	
fol.env_g	-0.490	0.683	181.960	-0.717	0.474	
fol.env_ʧ	-0.983	0.790	181.960	-1.244	0.215	
fol.env_k	0.659	0.827	181.960	0.797	0.426	
fol.env_l	2.111	1.094	181.960	1.930	0.055	
fol.env_m	0.908	0.708	181.960	1.283	0.201	
fol.env_n	0.243	0.529	181.960	0.459	0.647	
fol.env_ŋ	0.604	0.536	181.960	1.126	0.262	
fol.env_p	1.417	0.858	181.960	1.653	0.100	
fol.env_s	0.015	0.699	181.960	0.022	0.983	
fol.env_ʃ	-0.314	0.720	181.960	-0.436	0.664	
fol.env_t	0.561	0.530	181.960	1.059	0.291	
fol.env_v	2.006	1.024	181.960	1.959	0.052	
fol.env_z	-0.250	0.534	181.960	-0.468	0.640	
pre.env_b	0.204	0.335	181.960	0.609	0.544	
pre.env_d	0.292	0.497	181.960	0.587	0.558	
pre.env_ð	-0.230	0.535	181.960	-0.430	0.668	
pre.env_f	0.501	0.328	181.960	1.528	0.128	
pre.env_g	-2.358	0.996	181.960	-2.369	0.019	*
pre.env_h	-0.118	0.426	181.960	-0.277	0.782	
pre.env_ʧ	-1.670	1.074	181.960	-1.555	0.122	
pre.env_l	-1.930	0.911	181.960	-2.119	0.035	*
pre.env_r	-0.423	0.497	181.960	-0.852	0.395	
pre.env_s	-0.107	0.436	181.960	-0.246	0.806	
pre.env_t	-1.364	0.842	181.960	-1.620	0.107	
pre.env_w	0.484	0.508	181.960	0.953	0.342	

celex.frequency	0.082	0.036	181.960	2.263	0.025	*
setting_family	-0.207	0.084	181.960	-2.465	0.015	*
setting_services	0.539	0.246	181.960	2.192	0.030	*

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.2: Summary for model of KIT F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	1.980	0.539	181.960	3.672	0.000	***
fol.env_ð	-1.350	0.661	181.960	-2.042	0.043	*
fol.env_f	-0.591	0.216	181.960	-2.736	0.007	**
fol.env_g	-1.947	0.671	181.960	-2.901	0.004	**
fol.env_ʧ	-0.975	0.718	181.960	-1.357	0.176	
fol.env_k	-0.465	0.796	181.960	-0.584	0.560	
fol.env_l	-2.209	1.002	181.960	-2.205	0.029	*
fol.env_m	-0.700	0.700	181.960	-1.001	0.318	
fol.env_n	-1.131	0.494	181.960	-2.288	0.023	*
fol.env_ŋ	-0.764	0.548	181.960	-1.393	0.165	
fol.env_p	-2.700	0.783	181.960	-3.447	0.001	***
fol.env_s	-1.042	0.696	181.960	-1.496	0.136	
fol.env_ʃ	-1.243	0.690	181.960	-1.802	0.073	
fol.env_t	-1.248	0.486	181.960	-2.571	0.011	*
fol.env_v	-1.525	0.938	181.960	-1.626	0.106	
fol.env_z	-1.487	0.551	181.960	-2.700	0.008	**
pre.env_b	0.213	0.241	181.960	0.884	0.378	
pre.env_d	-1.182	0.482	181.960	-2.454	0.015	*
pre.env_ð	-0.226	0.529	181.960	-0.428	0.669	
pre.env_f	-0.860	0.301	181.960	-2.856	0.005	**
pre.env_g	1.038	0.922	181.960	1.125	0.262	
pre.env_h	-0.690	0.427	181.960	-1.615	0.108	
pre.env_ʧ	0.278	0.987	181.960	0.282	0.779	
pre.env_l	0.254	0.844	181.960	0.301	0.764	
pre.env_r	0.352	0.472	181.960	0.746	0.457	
pre.env_s	-0.970	0.380	181.960	-2.551	0.012	*

pre.env_t	0.510	0.777	181.960	0.656	0.512	
pre.env_w	-0.397	0.464	181.960	-0.855	0.394	
category_function	-0.413	0.180	181.960	-2.299	0.023	*
setting_family	-0.036	0.076	181.960	-0.474	0.636	
setting_services	0.252	0.224	181.960	1.122	0.263	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.3: Summary for model of DRESS F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-1.326	0.319	76.980	-4.160	0.000	***
fol.env_ð	-0.010	0.651	76.980	-0.015	0.988	
fol.env_k	0.443	0.486	76.980	0.911	0.365	
fol.env_l	1.967	0.547	76.980	3.598	0.001	***
fol.env_m	0.286	0.646	76.980	0.443	0.659	
fol.env_n	0.285	0.487	76.980	0.586	0.560	
fol.env_r	0.737	0.401	76.980	1.839	0.070	
fol.env_s	-0.633	0.610	76.980	-1.037	0.303	
fol.env_ʃ	-0.282	0.710	76.980	-0.397	0.692	
fol.env_t	0.253	0.715	76.980	0.354	0.724	
fol.env_v	0.497	0.667	76.980	0.746	0.458	
fol.env_ʒ	1.696	0.771	76.980	2.201	0.031	*
pre.env_b	2.692	0.694	76.980	3.879	0.000	***
pre.env_d	-0.308	0.659	76.980	-0.468	0.641	
pre.env_ð	0.217	0.516	76.980	0.421	0.675	
pre.env_f	0.390	0.703	76.980	0.554	0.581	
pre.env_g	0.042	0.663	76.980	0.064	0.949	
pre.env_j	0.530	0.590	76.980	0.898	0.372	
pre.env_k	-0.195	0.612	76.980	-0.319	0.750	
pre.env_l	0.366	0.527	76.980	0.694	0.490	
pre.env_m	0.120	0.538	76.980	0.223	0.824	
pre.env_n	-0.511	0.612	76.980	-0.836	0.406	
pre.env_p	-0.180	0.489	76.980	-0.368	0.714	
pre.env_r	0.465	0.458	76.980	1.015	0.313	

pre.env_s	-0.493	0.451	76.980	-1.093	0.278	
pre.env_t	-0.753	0.612	76.980	-1.231	0.222	
pre.env_w	0.044	0.383	76.980	0.115	0.909	
pre.env_z	0.937	0.582	76.980	1.609	0.112	
category_function	0.602	0.184	76.980	3.267	0.002	**
setting_family	0.463	0.182	76.980	2.540	0.013	*

Note. * p<0.05; ** p<0.01, *** p<0.001; the services setting is missing from the model because Emily did not produce any DRESS vowels in her services recordings

Table B.4: Summary for model of DRESS F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	1.954	0.255	76.980	7.662	0.000	***
fol.env_ð	-0.300	0.506	76.980	-0.593	0.555	
fol.env_k	-1.422	0.381	76.980	-3.733	0.000	***
fol.env_l	-2.119	0.423	76.980	-5.013	0.000	***
fol.env_m	-2.156	0.455	76.980	-4.739	0.000	***
fol.env_n	-0.828	0.355	76.980	-2.334	0.022	*
fol.env_r	-1.901	0.314	76.980	-6.060	0.000	***
fol.env_s	-0.708	0.461	76.980	-1.535	0.129	
fol.env_ʃ	-0.682	0.539	76.980	-1.266	0.209	
fol.env_t	-1.339	0.512	76.980	-2.618	0.011	*
fol.env_v	-1.041	0.521	76.980	-2.000	0.049	*
fol.env_ʒ	-1.549	0.567	76.980	-2.733	0.008	**
pre.env_b	-1.053	0.528	76.980	-1.995	0.050	*
pre.env_d	0.916	0.462	76.980	1.982	0.051	
pre.env_ð	0.289	0.396	76.980	0.729	0.468	
pre.env_f	0.205	0.548	76.980	0.375	0.709	
pre.env_g	0.450	0.465	76.980	0.968	0.336	
pre.env_j	-0.144	0.454	76.980	-0.317	0.752	
pre.env_k	0.671	0.457	76.980	1.468	0.146	
pre.env_l	-0.233	0.412	76.980	-0.567	0.572	
pre.env_m	0.150	0.366	76.980	0.410	0.683	
pre.env_n	0.546	0.454	76.980	1.201	0.234	

pre.env_p	-0.076	0.354	76.980	-0.216	0.830	
pre.env_r	-0.501	0.336	76.980	-1.490	0.140	
pre.env_s	0.291	0.322	76.980	0.902	0.370	
pre.env_t	0.322	0.457	76.980	0.704	0.484	
pre.env_w	-0.954	0.297	76.980	-3.210	0.002	**
pre.env_z	-0.623	0.434	76.980	-1.435	0.155	
duration	1.232	0.603	76.980	2.044	0.044	*
setting_family	-0.092	0.143	76.980	-0.646	0.520	

Note. * p<0.05; ** p<0.01, *** p<0.001; the services setting is missing from the model because Emily did not produce any DRESS vowels in her services recordings

Table B.5: Summary for model of TRAP F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	0.444	0.738	165.970	0.602	0.548	
fol.env_ʧ	0.101	1.039	165.970	0.097	0.923	
fol.env_k	-0.173	0.765	165.970	-0.226	0.821	
fol.env_l	-0.651	1.039	165.970	-0.626	0.532	
fol.env_m	-0.111	0.815	165.970	-0.137	0.892	
fol.env_n	-0.192	0.742	165.970	-0.259	0.796	
fol.env_ŋ	-0.425	0.853	165.970	-0.498	0.619	
fol.env_ʃ	-0.092	0.951	165.970	-0.096	0.923	
fol.env_t	0.064	0.770	165.970	0.083	0.934	
fol.env_v	-0.079	0.330	165.970	-0.238	0.812	
fol.env_z	-0.263	0.668	165.970	-0.394	0.694	
pre.env_b	0.169	0.413	165.970	0.409	0.683	
pre.env_d	-0.156	0.782	165.970	-0.200	0.842	
pre.env_ð	-0.304	0.247	165.970	-1.233	0.219	
pre.env_f	-0.381	0.491	165.970	-0.776	0.439	
pre.env_h	0.241	0.673	165.970	0.359	0.720	
pre.env_k	-0.901	0.456	165.970	-1.978	0.050	*
pre.env_l	0.872	0.325	165.970	2.688	0.008	**
pre.env_m	-0.236	0.317	165.970	-0.744	0.458	
pre.env_n	0.140	0.435	165.970	0.322	0.748	

pre.env_p	0.944	0.369	165.970	2.561	0.011	*
pre.env_r	-0.035	0.370	165.970	-0.095	0.924	
pre.env_t	0.981	0.604	165.970	1.624	0.106	
pre.env_ϴ	0.004	0.543	165.970	0.007	0.994	
duration	0.763	0.435	165.970	1.755	0.081	
setting_family	0.201	0.111	165.970	1.822	0.070	
setting_services	0.500	0.273	165.970	1.832	0.069	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.6: Summary for model of STRUT F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-1.420	0.136	81.970	-10.414	0.000	***
fol.env_d	0.924	0.155	81.970	5.963	0.000	***
fol.env_ð	0.696	0.192	81.970	3.616	0.001	***
fol.env_f	0.345	0.398	81.970	0.865	0.389	
fol.env_ʧ	0.752	0.318	81.970	2.361	0.021	*
fol.env_m	1.125	0.316	81.970	3.564	0.001	***
fol.env_n	0.741	0.171	81.970	4.339	0.000	***
fol.env_p	0.671	0.188	81.970	3.576	0.001	***
fol.env_s	0.951	0.357	81.970	2.660	0.009	**
fol.env_t	0.758	0.132	81.970	5.733	0.000	***
fol.env_z	1.078	0.186	81.970	5.802	0.000	***
pre.env_ʤ	0.361	0.355	81.970	1.017	0.312	
pre.env_k	-0.455	0.311	81.970	-1.462	0.148	
pre.env_l	-0.346	0.285	81.970	-1.216	0.227	
pre.env_m	-0.434	0.224	81.970	-1.941	0.056	
pre.env_n	0.768	0.486	81.970	1.579	0.118	
pre.env_r	-0.295	0.285	81.970	-1.037	0.303	
pre.env_s	-0.607	0.275	81.970	-2.205	0.030	*
setting_family	-0.042	0.075	81.970	-0.558	0.578	
setting_services	-0.013	0.174	81.970	-0.073	0.942	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.7: Summary for model of GOOSE F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	1.637	0.633	127.960	2.584	0.011	*
fol.env_dʒ	-0.209	0.441	127.960	-0.473	0.637	
fol.env_d	-0.616	0.314	127.960	-1.959	0.052	
fol.env_l	-0.767	0.218	127.960	-3.520	0.001	***
fol.env_m	-1.084	0.468	127.960	-2.317	0.022	*
fol.env_n	-0.349	0.353	127.960	-0.989	0.324	
fol.env_p	-0.876	0.415	127.960	-2.109	0.037	*
fol.env_s	0.179	0.482	127.960	0.370	0.712	
fol.env_t	-1.327	0.381	127.960	-3.485	0.001	***
fol.env_v	-0.531	0.352	127.960	-1.508	0.134	
pre.env_h	-0.832	0.398	127.960	-2.091	0.038	*
pre.env_j	0.508	0.179	127.960	2.844	0.005	**
pre.env_t	0.200	0.180	127.960	1.111	0.268	
pre.env_z	-0.487	0.442	127.960	-1.100	0.273	
celex.frequency	-0.136	0.053	127.960	-2.548	0.012	*
setting_family	-0.395	0.076	127.960	-5.192	0.000	***
setting_services	-0.150	0.148	127.960	-1.010	0.314	

Note. * p<0.05; ** p<0.01, *** p<0.001

German L1 speakers

Table B.8: Summary for model of KIT F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-0.586	0.480	1601.0	-1.222	0.222	
fol.env_b	-0.079	0.669	2026.0	-0.118	0.906	
fol.env_d	-0.052	0.479	2020.0	-0.108	0.914	
fol.env_ð	-0.057	0.493	1993.0	-0.115	0.908	
fol.env_f	0.060	0.475	2019.0	0.127	0.899	
fol.env_g	-0.112	0.484	2012.0	-0.231	0.817	

fol.env_ʧ	-0.228	0.493	2002.0	-0.463	0.644	
fol.env_k	0.128	0.479	2026.0	0.266	0.790	
fol.env_l	0.117	0.480	2026.0	0.243	0.808	
fol.env_m	0.267	0.482	2029.0	0.553	0.580	
fol.env_n	0.014	0.473	2022.0	0.029	0.977	
fol.env_ŋ	0.239	0.481	2029.0	0.497	0.619	
fol.env_p	0.001	0.504	2024.0	0.001	0.999	
fol.env_r	0.082	0.580	2024.0	0.142	0.887	
fol.env_s	-0.162	0.482	2028.0	-0.337	0.736	
fol.env_ʃ	-0.144	0.484	2028.0	-0.297	0.766	
fol.env_t	0.017	0.473	2024.0	0.035	0.972	
fol.env_v	0.149	0.483	2020.0	0.309	0.757	
fol.env_z	-0.161	0.474	2011.0	-0.339	0.734	
fol.env_ʒ	-0.006	0.506	2026.0	-0.012	0.991	
pre.env_ʤ	-0.224	0.481	2027.0	-0.465	0.642	
pre.env_b	-0.101	0.089	548.8	-1.139	0.255	
pre.env_d	-0.258	0.074	346.2	-3.484	0.001	***
pre.env_ð	0.087	0.117	366.0	0.741	0.459	
pre.env_f	0.023	0.095	1413.0	0.246	0.806	
pre.env_g	-0.580	0.114	803.5	-5.113	0.000	***
pre.env_h	-0.151	0.096	591.3	-1.576	0.116	
pre.env_ʧ	-0.178	0.152	1021.0	-1.172	0.241	
pre.env_k	-0.094	0.113	1366.0	-0.829	0.407	
pre.env_l	-0.057	0.071	587.0	-0.794	0.428	
pre.env_m	-0.338	0.098	890.3	-3.456	0.001	***
pre.env_n	-0.960	0.480	2023.0	-2.000	0.046	*
pre.env_p	-0.115	0.110	1288.0	-1.048	0.295	
pre.env_r	0.002	0.103	1247.0	0.022	0.983	
pre.env_s	-0.210	0.065	833.0	-3.220	0.001	**
pre.env_ʃ	-0.086	0.297	1980.0	-0.291	0.771	
pre.env_t	-0.219	0.103	719.0	-2.129	0.034	*
pre.env_Ɵ	-0.027	0.102	864.3	-0.261	0.794	
pre.env_v	-0.198	0.108	1541.0	-1.835	0.067	

pre.env_w	-0.033	0.131	1605.0	-0.248	0.804	
duration	2.448	0.304	2031.0	8.051	0.000	***
setting_family	0.007	0.040	6.1	0.167	0.873	
setting_services	0.294	0.044	10.3	6.673	0.000	***

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.9: Summary for model of KIT F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-0.053	0.498	1258	-0.106	0.916	
fol.env_d	0.905	0.490	2025	1.847	0.065	
fol.env_ð	0.681	0.503	2025	1.352	0.177	
fol.env_f	0.647	0.487	2025	1.328	0.184	
fol.env_g	1.239	0.495	2025	2.505	0.012	*
fol.env_ʧ	0.874	0.503	2025	1.738	0.082	
fol.env_k	0.879	0.487	2025	1.805	0.071	
fol.env_l	0.038	0.488	2025	0.078	0.938	
fol.env_m	0.609	0.490	2025	1.242	0.214	
fol.env_n	0.838	0.486	2024	1.727	0.084	
fol.env_ŋ	1.128	0.490	2023	2.303	0.021	*
fol.env_p	0.357	0.513	2024	0.697	0.486	
fol.env_r	0.320	0.590	2022	0.542	0.588	
fol.env_s	0.634	0.490	2024	1.296	0.195	
fol.env_ʃ	0.767	0.495	2024	1.549	0.121	
fol.env_t	0.640	0.485	2024	1.318	0.188	
fol.env_v	0.743	0.495	2024	1.501	0.133	
fol.env_z	0.755	0.488	2024	1.549	0.122	
fol.env_ʒ	0.651	0.512	2024	1.271	0.204	
pre.env_ʤ	0.050	0.491	2026	0.102	0.918	
pre.env_b	-0.294	0.091	2029	-3.214	0.001	**
pre.env_d	-0.141	0.075	2033	-1.894	0.058	
pre.env_ð	-0.160	0.117	2025	-1.374	0.170	
pre.env_f	-0.273	0.100	2029	-2.741	0.006	**
pre.env_g	0.576	0.120	2027	4.788	0.000	***

pre.env_h	0.112	0.097	2026	1.165	0.244	
pre.env_ʧ	0.344	0.155	1895	2.216	0.027	*
pre.env_k	0.358	0.119	2027	3.005	0.003	**
pre.env_l	-0.377	0.078	2027	-4.817	0.000	***
pre.env_m	-0.695	0.103	2025	-6.748	0.000	***
pre.env_n	-0.394	0.491	2028	-0.803	0.422	
pre.env_p	-0.114	0.115	2022	-0.993	0.321	
pre.env_r	-0.104	0.108	2028	-0.972	0.331	
pre.env_s	-0.276	0.070	2030	-3.921	0.000	***
pre.env_ʃ	-0.045	0.303	2028	-0.149	0.882	
pre.env_t	0.087	0.105	2018	0.824	0.410	
pre.env_ϴ	-0.220	0.102	2016	-2.154	0.031	*
pre.env_v	-0.673	0.118	2026	-5.721	0.000	***
pre.env_w	-0.670	0.134	2032	-5.007	0.000	***
celex.frequency	-0.034	0.008	2032	-4.358	0.000	***
duration	3.329	0.309	2026	10.768	0.000	***
setting_family	0.039	0.039	7	1.009	0.345	
setting_services	0.054	0.075	4	0.710	0.520	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.10: Summary for model of DRESS F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	0.350	0.567	348.2	0.618	0.537	
fol.env_d	-0.217	0.537	1237.0	-0.405	0.686	
fol.env_ð	-0.239	0.552	1238.0	-0.433	0.665	
fol.env_f	0.165	0.547	1237.0	0.301	0.763	
fol.env_g	-0.799	0.613	1236.0	-1.304	0.193	
fol.env_ʧ	0.094	0.750	1237.0	0.126	0.900	
fol.env_k	-0.106	0.534	1236.0	-0.198	0.843	
fol.env_l	0.006	0.531	1236.0	0.011	0.991	
fol.env_m	-0.211	0.536	1236.0	-0.394	0.694	
fol.env_n	-0.217	0.530	1236.0	-0.410	0.682	
fol.env_p	0.044	0.568	1236.0	0.078	0.938	

fol.env_r	-0.371	0.541	1237.0	-0.685	0.494	
fol.env_s	-0.218	0.533	1237.0	-0.410	0.682	
fol.env_ʃ	-0.428	0.542	1236.0	-0.790	0.430	
fol.env_t	-0.185	0.535	1237.0	-0.346	0.730	
fol.env_ϴ	-0.084	0.598	1237.0	-0.140	0.889	
fol.env_v	-0.250	0.534	1236.0	-0.468	0.640	
fol.env_z	-1.847	0.750	1238.0	-2.461	0.014	*
fol.env_ʒ	-0.235	0.593	1236.0	-0.396	0.692	
pre.env_ʤ	-0.564	0.180	1237.0	-3.124	0.002	**
pre.env_aʊ	0.557	0.385	1234.0	1.447	0.148	
pre.env_b	-0.040	0.125	1238.0	-0.318	0.750	
pre.env_d	-0.465	0.135	1236.0	-3.446	0.001	***
pre.env_ð	0.096	0.074	1239.0	1.300	0.194	
pre.env_f	-0.322	0.147	1232.0	-2.199	0.028	*
pre.env_g	-0.301	0.098	1236.0	-3.055	0.002	**
pre.env_h	0.393	0.224	1236.0	1.752	0.080	
pre.env_ɪ	-0.299	0.533	1236.0	-0.561	0.575	
pre.env_j	-0.209	0.127	1224.0	-1.650	0.099	
pre.env_k	-0.080	0.133	1237.0	-0.604	0.546	
pre.env_l	-0.068	0.089	1238.0	-0.769	0.442	
pre.env_m	-0.305	0.101	1237.0	-3.010	0.003	**
pre.env_n	0.046	0.116	1238.0	0.395	0.693	
pre.env_p	-0.007	0.108	1239.0	-0.064	0.949	
pre.env_r	0.159	0.083	1239.0	1.918	0.055	
pre.env_s	-0.075	0.076	1237.0	-0.994	0.320	
pre.env_ʃ	-0.144	0.274	1217.0	-0.527	0.599	
pre.env_t	0.122	0.081	1239.0	1.501	0.134	
pre.env_v	-0.359	0.122	1237.0	-2.934	0.003	**
pre.env_w	-0.090	0.070	1233.0	-1.289	0.198	
pre.env_z	-0.618	0.378	1238.0	-1.633	0.103	
duration	3.169	0.494	1238.0	6.417	0.000	***
category_function	-0.098	0.046	1236.0	-2.148	0.032	*
setting_family	-0.018	0.060	6.8	-0.292	0.779	

setting_services	0.197	0.106	5.0	1.852	0.123	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.11: Summary for model of DRESS F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	0.768	0.503	198.900	1.5	0.129	
fol.env_d	-0.022	0.466	930.700	0.0	0.962	
fol.env_ð	-0.185	0.489	604.500	-0.4	0.705	
fol.env_f	-0.091	0.476	833.600	-0.2	0.849	
fol.env_g	-0.045	0.532	902.300	-0.1	0.933	
fol.env_ʧ	-0.071	0.646	1027.600	-0.1	0.913	
fol.env_k	0.035	0.462	975.300	0.1	0.940	
fol.env_l	-0.716	0.459	971.700	-1.6	0.119	
fol.env_m	-0.268	0.466	914.900	-0.6	0.566	
fol.env_n	0.084	0.457	986.900	0.2	0.854	
fol.env_p	-0.054	0.491	957.900	-0.1	0.912	
fol.env_r	-0.243	0.470	899.500	-0.5	0.606	
fol.env_s	-0.274	0.459	992.900	-0.6	0.551	
fol.env_ʃ	-0.208	0.470	935.300	-0.4	0.658	
fol.env_t	-0.011	0.462	951.400	0.0	0.981	
fol.env_Ɵ	0.233	0.522	816.200	0.4	0.656	
fol.env_v	-0.195	0.462	936.300	-0.4	0.673	
fol.env_z	0.714	0.646	1026.100	1.1	0.269	
fol.env_ʒ	-0.372	0.511	1000.400	-0.7	0.467	
pre.env_ʤ	0.005	0.160	584.200	0.0	0.978	
pre.env_aʊ	-2.125	0.332	961.700	-6.4	0.000	***
pre.env_b	-0.528	0.124	141.300	-4.3	0.000	***
pre.env_d	-0.337	0.126	287.000	-2.7	0.008	**
pre.env_ð	-0.614	0.101	40.200	-6.1	0.000	***
pre.env_f	-0.443	0.137	258.500	-3.2	0.001	**
pre.env_g	0.074	0.104	90.700	0.7	0.484	
pre.env_h	-0.137	0.200	569.000	-0.7	0.494	
pre.env_ɪ	0.326	0.459	1030.100	0.7	0.478	

pre.env_j	-0.053	0.140	62.300	-0.4	0.708	
pre.env_k	0.232	0.132	142.100	1.8	0.082	
pre.env_l	-0.556	0.087	167.800	-6.4	0.000	***
pre.env_m	-0.688	0.100	173.300	-6.9	0.000	***
pre.env_n	-0.156	0.117	137.100	-1.3	0.185	
pre.env_p	-0.606	0.103	222.800	-5.9	0.000	***
pre.env_r	-0.629	0.084	168.900	-7.5	0.000	***
pre.env_s	-0.309	0.077	138.900	-4.0	0.000	***
pre.env_ʃ	-0.029	0.243	530.700	-0.1	0.907	
pre.env_t	-0.230	0.083	123.100	-2.8	0.006	**
pre.env_v	-0.662	0.120	191.000	-5.5	0.000	***
pre.env_w	-0.950	0.084	53.600	-11.3	0.000	***
pre.env_z	-0.689	0.326	1011.200	-2.1	0.035	*
duration	2.618	0.420	1240.500	6.2	0.000	***
category_function	-0.162	0.052	78.600	-3.1	0.003	**
setting_family	-0.016	0.058	5.500	-0.3	0.786	
setting_services	-0.039	0.145	5.300	-0.3	0.799	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.12: Summary for model of TRAP F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	0.379	0.497	837.100	0.762	0.446	
fol.env_b	0.936	0.551	1183.000	1.700	0.089	
fol.env_d	0.464	0.487	844.200	0.953	0.341	
fol.env_f	-0.077	0.717	1114.000	-0.107	0.915	
fol.env_g	0.713	0.570	1277.000	1.250	0.212	
fol.env_ʧ	0.439	0.578	1349.000	0.759	0.448	
fol.env_k	0.553	0.468	1137.000	1.182	0.237	
fol.env_l	0.406	0.489	1049.000	0.831	0.406	
fol.env_m	0.711	0.476	1109.000	1.494	0.136	
fol.env_n	0.391	0.464	1173.000	0.843	0.399	
fol.env_ŋ	0.647	0.503	1176.000	1.286	0.199	
fol.env_p	0.948	0.501	1104.000	1.894	0.058	

fol.env_r	0.000	0.536	1176.000	0.000	1.000	
fol.env_s	0.710	0.510	1002.000	1.391	0.164	
fol.env_ʃ	0.293	0.573	1312.000	0.512	0.609	
fol.env_t	0.471	0.466	1104.000	1.011	0.312	
fol.env_ϴ	0.826	0.522	1095.000	1.581	0.114	
fol.env_v	0.508	0.487	865.200	1.042	0.298	
fol.env_z	0.072	0.484	682.300	0.148	0.882	
pre.env_ʤ	0.267	0.363	1161.000	0.736	0.462	
pre.env_b	0.167	0.160	136.300	1.045	0.298	
pre.env_d	-0.119	0.268	179.100	-0.444	0.658	
pre.env_ð	0.098	0.139	57.900	0.705	0.483	
pre.env_f	-0.024	0.202	348.900	-0.119	0.905	
pre.env_g	0.588	0.815	1636.000	0.722	0.470	
pre.env_h	0.589	0.146	162.200	4.036	0.000	***
pre.env_ɪ	-0.043	0.254	346.600	-0.168	0.867	
pre.env_ʧ	0.035	0.771	1651.000	0.045	0.964	
pre.env_k	-0.027	0.123	120.100	-0.218	0.828	
pre.env_l	0.447	0.189	514.300	2.373	0.018	*
pre.env_m	-0.120	0.190	516.600	-0.630	0.529	
pre.env_n	0.122	0.316	1124.000	0.386	0.700	
pre.env_p	0.116	0.331	1318.000	0.351	0.726	
pre.env_r	0.232	0.170	548.400	1.367	0.172	
pre.env_s	0.114	0.312	928.900	0.365	0.715	
pre.env_t	-0.268	0.234	773.300	-1.146	0.252	
pre.env_ϴ	-0.462	0.294	201.600	-1.572	0.118	
pre.env_v	-0.371	0.557	1521.000	-0.666	0.506	
pre.env_w	0.281	0.836	1622.000	0.336	0.737	
pre.env_z	-0.517	0.252	168.500	-2.054	0.042	*
celex.frequency	-0.043	0.015	235.400	-2.867	0.005	**
duration	3.859	0.246	2172.000	15.702	0.000	***
setting_family	-0.029	0.110	6.000	-0.262	0.802	
setting_services	0.089	0.065	12.400	1.367	0.196	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.13: Summary for model of FOOT F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-0.445	0.378	221.82	-1.18	0.239	
fol.env_k	-0.746	0.316	299.7	-2.362	0.019	*
fol.env_l	-0.605	0.851	299.21	-0.711	0.478	
fol.env_ʃ	-0.649	0.451	301.09	-1.439	0.151	
fol.env_t	-0.489	0.406	303.26	-1.206	0.229	
pre.env_f	-0.666	0.556	300.23	-1.197	0.232	
pre.env_g	0.437	0.371	302.93	1.178	0.240	
pre.env_k	0.504	0.356	300.23	1.421	0.156	
pre.env_l	0.643	0.205	304.08	3.135	0.002	**
pre.env_ʃ	0.416	0.394	303.04	1.054	0.293	
pre.env_t	1.090	0.269	303.52	4.056	0.000	***
pre.env_w	-0.334	0.368	301.98	-0.906	0.365	
setting_family	-0.119	0.139	6.02	-0.854	0.426	
setting_services	0.071	0.179	5.2	0.398	0.706	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.14: Summary for model of GOOSE F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-0.260	0.494	307.1	-0.526	0.599	
fol.env_ə	0.080	0.356	424.1	0.225	0.822	
fol.env_ʤ	-0.501	0.358	137	-1.4	0.164	
fol.env_d	0.156	0.192	135.8	0.81	0.420	
fol.env_ð	0.245	0.591	684.4	0.414	0.679	
fol.env_ɪ	-0.441	0.189	24.4	-2.326	0.029	*
fol.env_ʧ	-0.041	0.400	207.7	-0.101	0.919	
fol.env_l	-1.066	0.487	704.6	-2.191	0.029	*
fol.env_m	-0.114	0.263	216.6	-0.435	0.664	
fol.env_n	-0.482	0.300	431.4	-1.608	0.109	
fol.env_p	-0.494	0.323	383.6	-1.529	0.127	

fol.env_s	0.198	0.162	61.9	1.217	0.228	
fol.env_ʃ	0.231	0.404	122.8	0.572	0.568	
fol.env_t	0.055	0.158	71.1	0.35	0.728	
fol.env_ϴ	-0.840	0.336	347.8	-2.503	0.013	*
fol.env_v	-0.686	0.489	708.2	-1.401	0.162	
fol.env_z	0.320	0.140	59	2.281	0.026	*
fol.env_ʒ	0.187	0.308	76	0.608	0.545	
pre.env_d	0.867	0.483	308	1.797	0.073	
pre.env_f	0.198	0.618	624.4	0.319	0.750	
pre.env_h	-0.160	0.503	356.7	-0.319	0.75	
pre.env_j	1.067	0.459	593.8	2.323	0.020	*
pre.env_ʧ	0.799	0.615	363.2	1.298	0.195	
pre.env_k	0.028	0.653	620.2	0.044	0.965	
pre.env_l	0.057	0.506	401.1	0.112	0.911	
pre.env_m	0.009	0.668	660.7	0.014	0.989	
pre.env_n	0.988	0.533	483.2	1.852	0.065	
pre.env_p	-0.047	0.915	749.5	-0.051	0.959	
pre.env_r	-0.006	0.509	540.5	-0.011	0.991	
pre.env_s	0.986	0.506	429.4	1.947	0.052	
pre.env_ʃ	-0.129	0.648	688.1	-0.199	0.843	
pre.env_t	0.697	0.473	393.6	1.472	0.142	
duration	-1.518	0.241	982.1	-6.304	0.000	***
setting_family	0.098	0.049	9.8	2.005	0.073	
setting_services	0.097	0.097	6.9	0.997	0.352	

Note. * p<0.05; ** p<0.01, *** p<0.001

Korean L1 speakers

Table B.15: Summary for model of KIT F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-1.113	0.496	1437	-2.246	0.025	*
fol.env_ʤ	-0.391	0.695	1821	-0.563	0.573	
fol.env_b	0.622	0.514	1827	1.211	0.226	
fol.env_d	0.429	0.491	1826	0.874	0.382	

fol.env_ð	0.236	0.496	1827	0.476	0.634	
fol.env_f	0.422	0.488	1827	0.865	0.387	
fol.env_g	0.467	0.501	1827	0.933	0.351	
fol.env_ʧ	-0.268	0.501	1827	-0.536	0.592	
fol.env_k	0.708	0.498	1826	1.421	0.155	
fol.env_l	0.492	0.493	1826	0.997	0.319	
fol.env_m	0.686	0.496	1826	1.382	0.167	
fol.env_n	0.507	0.486	1826	1.043	0.297	
fol.env_ŋ	0.834	0.506	1827	1.649	0.099	
fol.env_p	0.526	0.524	1826	1.003	0.316	
fol.env_s	0.596	0.492	1826	1.211	0.226	
fol.env_ʃ	0.218	0.498	1826	0.437	0.662	
fol.env_t	0.557	0.486	1826	1.147	0.251	
fol.env_v	0.492	0.496	1827	0.993	0.321	
fol.env_z	0.298	0.486	1826	0.614	0.540	
fol.env_ʒ	-0.365	0.703	1824	-0.519	0.604	
pre.env_b	-0.037	0.075	1824	-0.494	0.622	
pre.env_d	-0.003	0.069	1830	-0.044	0.965	
pre.env_ð	-0.217	0.104	1828	-2.083	0.037	*
pre.env_f	0.232	0.090	1822	2.567	0.010	*
pre.env_g	-0.651	0.210	1827	-3.097	0.002	**
pre.env_h	-0.142	0.115	1827	-1.233	0.218	
pre.env_ʧ	-0.235	0.290	1825	-0.810	0.418	
pre.env_k	0.216	0.139	1825	1.555	0.120	
pre.env_l	0.194	0.079	1829	2.456	0.014	*
pre.env_m	0.051	0.106	1823	0.480	0.631	
pre.env_n	0.876	0.487	1827	1.801	0.072	
pre.env_p	-0.068	0.229	1825	-0.296	0.767	
pre.env_r	-0.208	0.086	1821	-2.415	0.016	*
pre.env_s	-0.208	0.071	1825	-2.925	0.003	**
pre.env_ʃ	-0.619	0.484	1825	-1.279	0.201	
pre.env_t	0.116	0.101	1829	1.150	0.250	
pre.env_ϴ	-0.203	0.145	1826	-1.395	0.163	

pre.env_v	-0.122	0.165	1827	-0.738	0.460	
pre.env_w	0.351	0.094	1830	3.714	0.000	***
pre.env_z	-0.412	0.498	1825	-0.828	0.408	
duration	1.722	0.312	1826	5.528	0.000	***
setting_family	-0.041	0.047	6	-0.870	0.416	
setting_services	0.414	0.108	4	3.848	0.018	***

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.16: Summary for model of KIT F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	0.428	0.494	1135	0.866	0.387	
fol.env_dʒ	0.200	0.690	1556	0.290	0.772	
fol.env_b	-0.077	0.514	1329	-0.150	0.881	
fol.env_d	0.175	0.490	1440	0.357	0.721	
fol.env_ð	0.211	0.501	1060	0.421	0.674	
fol.env_f	0.031	0.487	1456	0.063	0.950	
fol.env_g	0.462	0.504	1164	0.916	0.360	
fol.env_ʧ	0.260	0.504	1153	0.515	0.607	
fol.env_k	0.155	0.496	1507	0.313	0.754	
fol.env_l	-0.278	0.491	1548	-0.566	0.572	
fol.env_m	0.098	0.493	1561	0.199	0.842	
fol.env_n	0.105	0.484	1571	0.216	0.829	
fol.env_ŋ	0.450	0.503	1535	0.894	0.372	
fol.env_p	-0.155	0.523	1444	-0.296	0.768	
fol.env_s	-0.099	0.490	1530	-0.202	0.840	
fol.env_ʃ	0.457	0.495	1551	0.924	0.356	
fol.env_t	0.029	0.484	1516	0.060	0.952	
fol.env_v	0.333	0.496	1409	0.672	0.502	
fol.env_z	0.057	0.486	1445	0.117	0.907	
fol.env_ʒ	0.875	0.700	1491	1.251	0.211	
pre.env_b	-0.279	0.093	78	-2.990	0.004	**
pre.env_d	-0.158	0.085	82	-1.861	0.066	
pre.env_ð	-0.171	0.143	35	-1.200	0.238	

pre.env_f	-0.332	0.100	202	-3.307	0.001	**
pre.env_g	0.380	0.221	399	1.715	0.087	
pre.env_h	0.120	0.129	164	0.928	0.355	
pre.env_ʧ	-0.056	0.293	1026	-0.191	0.849	
pre.env_k	0.059	0.154	208	0.383	0.702	
pre.env_l	-0.602	0.100	76	-6.021	0.000	***
pre.env_m	-0.585	0.114	358	-5.149	0.000	***
pre.env_n	0.413	0.486	1457	0.850	0.395	
pre.env_p	0.104	0.231	1077	0.447	0.655	
pre.env_r	-0.044	0.103	110	-0.425	0.672	
pre.env_s	-0.321	0.081	145	-3.949	0.000	***
pre.env_ʃ	0.259	0.482	1563	0.537	0.592	
pre.env_t	0.001	0.111	236	0.005	0.996	
pre.env_Ɵ	0.040	0.158	256	0.252	0.801	
pre.env_v	-0.245	0.178	283	-1.371	0.172	
pre.env_w	-0.722	0.105	193	-6.889	0.000	***
pre.env_z	-0.122	0.495	1556	-0.246	0.806	
duration	1.098	0.304	1785	3.610	0.000	***
setting_family	-0.035	0.049	6	-0.707	0.506	
setting_services	-0.008	0.163	4	-0.050	0.962	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.17: Summary for model of DRESS F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-0.861	0.405	287.900	-2.1	0.034	*
fol.env_b	0.758	0.641	1060.000	1.2	0.238	
fol.env_d	0.442	0.384	1061.000	1.2	0.249	
fol.env_ð	0.621	0.417	1061.000	1.5	0.136	
fol.env_f	0.119	0.420	1061.000	0.3	0.776	
fol.env_g	0.028	0.643	1059.000	0.0	0.966	
fol.env_ʧ	0.930	0.637	1060.000	1.5	0.145	
fol.env_k	0.463	0.379	1061.000	1.2	0.222	
fol.env_l	0.937	0.379	1061.000	2.5	0.013	*

fol.env_m	0.618	0.383	1062.000	1.6	0.107	
fol.env_n	0.471	0.376	1061.000	1.3	0.211	
fol.env_p	0.461	0.418	1060.000	1.1	0.270	
fol.env_r	0.094	0.394	1060.000	0.2	0.811	
fol.env_s	0.498	0.380	1061.000	1.3	0.190	
fol.env_ʃ	-0.001	0.403	1061.000	0.0	0.997	
fol.env_t	0.682	0.382	1061.000	1.8	0.075	
fol.env_ϴ	1.124	0.425	1062.000	2.6	0.008	**
fol.env_v	0.383	0.382	1060.000	1.0	0.316	
fol.env_z	0.168	0.636	1060.000	0.3	0.792	
fol.env_ʒ	0.622	0.449	1062.000	1.4	0.166	
pre.env_ʤ	-0.082	0.182	1060.000	-0.5	0.652	
pre.env_b	0.243	0.194	1061.000	1.3	0.211	
pre.env_d	-0.085	0.143	1061.000	-0.6	0.553	
pre.env_ð	0.145	0.072	1058.000	2.0	0.043	*
pre.env_f	0.745	0.179	1061.000	4.2	0.000	***
pre.env_g	-0.506	0.095	1060.000	-5.3	0.000	***
pre.env_h	0.275	0.168	1061.000	1.6	0.101	
pre.env_j	-0.142	0.154	1057.000	-0.9	0.356	
pre.env_k	-0.327	0.207	1060.000	-1.6	0.113	
pre.env_l	0.510	0.099	1061.000	5.2	0.000	***
pre.env_m	-0.169	0.093	1054.000	-1.8	0.069	
pre.env_n	-0.036	0.100	1062.000	-0.4	0.722	
pre.env_p	0.252	0.110	1060.000	2.3	0.022	*
pre.env_r	0.246	0.084	1062.000	2.9	0.004	**
pre.env_s	0.119	0.081	1052.000	1.5	0.142	
pre.env_t	0.078	0.089	1060.000	0.9	0.384	
pre.env_ϴ	-0.091	0.527	1060.000	-0.2	0.863	
pre.env_v	-0.023	0.109	1062.000	-0.2	0.833	
pre.env_w	0.163	0.070	1065.000	2.3	0.021	*
pre.env_z	-0.564	0.185	1061.000	-3.1	0.002	**
duration	4.238	0.412	1063.000	10.3	0.000	***
setting_family	0.022	0.097	6.300	0.2	0.825	

setting_services	0.105	0.153	5.100	0.7	0.524	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.18: Summary for model of DRESS F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	0.386	0.342	202.7	1.129	0.260	
fol.env_b	0.341	0.528	877.3	0.646	0.518	
fol.env_d	0.347	0.319	684.7	1.088	0.277	
fol.env_ð	0.493	0.349	546.9	1.412	0.159	
fol.env_f	0.498	0.350	633.2	1.423	0.155	
fol.env_g	0.442	0.529	875.3	0.837	0.403	
fol.env_ʧ	0.280	0.522	935.1	0.536	0.592	
fol.env_k	0.437	0.315	662.4	1.387	0.166	
fol.env_l	-0.360	0.316	638.4	-1.139	0.255	
fol.env_m	0.105	0.321	573.6	0.327	0.744	
fol.env_n	0.451	0.312	661.5	1.442	0.150	
fol.env_p	0.238	0.346	717.8	0.688	0.491	
fol.env_r	0.465	0.333	451.7	1.399	0.163	
fol.env_s	0.371	0.315	677.2	1.175	0.240	
fol.env_ʃ	0.171	0.336	611.2	0.509	0.611	
fol.env_t	0.271	0.319	635.1	0.852	0.394	
fol.env_ϴ	0.486	0.357	524.0	1.364	0.173	
fol.env_v	0.244	0.318	634.0	0.768	0.443	
fol.env_z	0.370	0.522	933.2	0.708	0.479	
fol.env_ʒ	0.668	0.372	700.9	1.794	0.073	
pre.env_ʤ	-0.354	0.158	241.2	-2.247	0.026	*
pre.env_b	-0.225	0.163	538.3	-1.384	0.167	
pre.env_d	-0.166	0.120	586.0	-1.391	0.165	
pre.env_ð	-0.230	0.082	24.4	-2.799	0.010	**
pre.env_f	-0.460	0.150	597.1	-3.068	0.002	**
pre.env_g	0.289	0.091	66.4	3.180	0.002	**
pre.env_h	-0.066	0.140	626.8	-0.470	0.639	
pre.env_j	0.187	0.134	210.6	1.392	0.165	

pre.env_k	0.550	0.172	674.3	3.194	0.001	**
pre.env_l	-0.627	0.084	370.6	-7.458	0.000	***
pre.env_m	-0.491	0.083	152.0	-5.892	0.000	***
pre.env_n	-0.044	0.091	109.2	-0.479	0.633	
pre.env_p	-0.384	0.096	223.6	-4.007	0.000	***
pre.env_r	-0.296	0.075	175.4	-3.944	0.000	***
pre.env_s	-0.294	0.073	138.5	-4.027	0.000	***
pre.env_t	0.041	0.079	209.6	0.515	0.607	
pre.env_ɵ	-0.344	0.431	989.0	-0.799	0.424	
pre.env_v	-0.482	0.094	301.3	-5.117	0.000	***
pre.env_w	-0.908	0.071	41.7	-12.711	0.000	***
pre.env_z	-0.547	0.169	83.5	-3.227	0.002	**
duration	1.417	0.334	1059.7	4.250	0.000	***
setting_family	-0.120	0.073	6.3	-1.640	0.150	
setting_services	-0.168	0.112	7.3	-1.489	0.178	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.19: Summary for model of TRAP F1

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	0.287	0.378	345.2	0.759	0.449	
fol.env_b	0.199	0.368	476.1	0.540	0.589	
fol.env_d	0.533	0.361	326.7	1.474	0.141	
fol.env_ð	0.550	0.622	857.9	0.885	0.377	
fol.env_g	0.005	0.689	1165.0	0.008	0.994	
fol.env_ʧ	0.682	0.406	546.8	1.679	0.094	
fol.env_k	0.603	0.345	411.5	1.748	0.081	
fol.env_l	0.451	0.397	488.3	1.137	0.256	
fol.env_m	0.782	0.361	449.7	2.166	0.031	*
fol.env_n	0.431	0.340	422.3	1.267	0.206	
fol.env_ŋ	0.398	0.362	411.6	1.099	0.272	
fol.env_p	0.802	0.374	459.1	2.147	0.032	*
fol.env_r	0.158	0.394	513.1	0.399	0.690	
fol.env_s	0.531	0.424	623.3	1.252	0.211	

fol.env_ʃ	0.708	0.419	444.0	1.691	0.092	
fol.env_t	0.673	0.353	309.0	1.906	0.058	
fol.env_ɵ	0.679	0.400	402.9	1.695	0.091	
fol.env_v	0.397	0.357	339.9	1.112	0.267	
fol.env_z	0.168	0.350	312.6	0.480	0.632	
fol.env_ʒ	-0.088	0.695	1138.0	-0.127	0.899	
pre.env_ʤ	-0.165	0.365	941.5	-0.453	0.651	
pre.env_b	0.623	0.152	158.3	4.094	0.000	***
pre.env_d	-0.016	0.209	121.0	-0.078	0.938	
pre.env_ð	-0.073	0.113	31.8	-0.643	0.525	
pre.env_f	0.142	0.173	230.0	0.819	0.414	
pre.env_g	0.069	0.319	500.5	0.217	0.828	
pre.env_h	0.517	0.122	129.6	4.227	0.000	***
pre.env_ɪ	-0.200	0.291	782.4	-0.686	0.493	
pre.env_ʧ	0.151	0.475	1347.0	0.318	0.750	
pre.env_k	-0.147	0.114	81.2	-1.294	0.199	
pre.env_l	0.645	0.132	277.3	4.878	0.000	***
pre.env_m	-0.024	0.142	216.3	-0.166	0.868	
pre.env_n	-0.348	0.146	189.1	-2.384	0.018	*
pre.env_p	0.858	0.232	1077.0	3.704	0.000	***
pre.env_r	0.259	0.139	398.8	1.857	0.064	
pre.env_s	0.032	0.293	551.5	0.108	0.914	
pre.env_t	0.309	0.141	382.0	2.194	0.029	*
pre.env_ɵ	-0.307	0.225	142.5	-1.365	0.174	
pre.env_v	0.154	0.313	444.7	0.491	0.624	
pre.env_z	-0.364	0.295	418.7	-1.235	0.218	
celex.frequency	-0.063	0.014	165.5	-4.565	0.000	***
duration	2.921	0.208	1880.0	14.026	0.000	***
setting_family	-0.076	0.063	5.7	-1.197	0.279	
setting_services	0.420	0.105	5.7	3.889	0.009	**

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.20: Summary for model of STRUT F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	-0.573	0.395	1043.100	-1.449	0.148	
fol.env_ʤ	-0.190	0.543	1045.500	-0.349	0.727	
fol.env_b	-0.673	0.390	1045.000	-1.725	0.085	
fol.env_d	-0.228	0.403	1044.600	-0.566	0.571	
fol.env_ð	-0.101	0.385	1045.100	-0.262	0.793	
fol.env_f	-0.604	0.398	1044.700	-1.518	0.129	
fol.env_ʧ	-0.307	0.387	1045.300	-0.794	0.427	
fol.env_k	-0.241	0.389	1044.300	-0.621	0.535	
fol.env_l	-0.658	0.393	1045.700	-1.673	0.095	
fol.env_m	-0.560	0.384	1044.900	-1.459	0.145	
fol.env_n	-0.368	0.378	1045.000	-0.972	0.331	
fol.env_ŋ	-0.432	0.412	1044.900	-1.049	0.294	
fol.env_p	-0.577	0.388	1045.200	-1.489	0.137	
fol.env_r	0.285	0.414	1044.900	0.687	0.492	
fol.env_s	-0.336	0.408	1044.900	-0.823	0.411	
fol.env_ʃ	-0.881	0.521	1045.100	-1.691	0.091	
fol.env_t	-0.341	0.401	1045.500	-0.849	0.396	
fol.env_ϴ	-0.317	0.439	1045.200	-0.722	0.471	
fol.env_v	-0.599	0.388	1045.400	-1.541	0.124	
fol.env_z	-0.086	0.392	1045.500	-0.218	0.827	
pre.env_ʤ	0.706	0.150	1045.700	4.716	0.000	***
pre.env_b	-0.023	0.124	1046.800	-0.183	0.855	
pre.env_d	0.206	0.100	1045.200	2.056	0.040	*
pre.env_f	-0.197	0.115	1046.900	-1.707	0.088	
pre.env_g	0.314	0.187	1045.700	1.682	0.093	
pre.env_h	0.158	0.147	1045.900	1.077	0.282	
pre.env_j	0.408	0.241	1045.600	1.694	0.091	
pre.env_k	0.080	0.090	1046.300	0.887	0.375	
pre.env_l	-0.092	0.112	1047.100	-0.820	0.412	
pre.env_m	-0.240	0.090	1045.400	-2.670	0.008	**
pre.env_n	0.554	0.140	1046.200	3.961	0.000	***
pre.env_p	-0.071	0.136	1046.500	-0.524	0.601	

pre.env_r	-0.110	0.078	1043.000	-1.420	0.156	
pre.env_s	0.157	0.093	1045.700	1.684	0.092	
pre.env_t	0.410	0.145	1046.900	2.828	0.005	**
pre.env_w	-0.328	0.092	1047.500	-3.574	0.000	***
pre.env_z	0.357	0.241	1045.500	1.483	0.138	
celex.frequency	0.013	0.011	1045.900	1.174	0.241	
duration	-0.681	0.275	1050.500	-2.473	0.014	*
setting_family	-0.018	0.027	15.800	-0.646	0.528	
setting_services	0.092	0.062	367.500	1.497	0.135	

Note. * p<0.05; ** p<0.01, *** p<0.001

Table B.21: Summary for model of GOOSE F2

	Estimate	Standard error	df	t value	Pr(>|t|)	Significance
(Intercept)	0.605	0.182	19.8	3.321	0.003	**
fol.env_dʒ	-0.248	0.279	53.6	-0.886	0.379	
fol.env_b	-0.925	0.595	721.2	-1.555	0.120	
fol.env_d	-0.211	0.175	59.0	-1.207	0.232	
fol.env_ɪ	-0.360	0.193	30.9	-1.869	0.071	
fol.env_ʧ	0.075	0.374	171.6	0.200	0.842	
fol.env_k	-0.904	0.440	313.1	-2.054	0.041	*
fol.env_l	-1.139	0.317	226.9	-3.593	0.000	***
fol.env_m	-0.326	0.280	462.4	-1.161	0.246	
fol.env_n	-0.181	0.186	130.4	-0.973	0.333	
fol.env_p	0.222	0.380	330.2	0.584	0.560	
fol.env_s	0.006	0.201	57.4	0.031	0.976	
fol.env_ʃ	0.082	0.306	563.0	0.266	0.790	
fol.env_t	-0.037	0.161	60.5	-0.232	0.818	
fol.env_v	-0.630	0.314	375.5	-2.009	0.045	*
fol.env_z	0.172	0.153	54.8	1.128	0.264	
fol.env_ʒ	0.842	0.258	123.1	3.267	0.001	**
pre.env_f	-0.702	0.361	151.0	-1.942	0.054	
pre.env_h	-0.840	0.257	47.1	-3.271	0.002	**
pre.env_j	0.472	0.156	21.1	3.020	0.006	**

pre.env_ʧ	-0.217	0.297	31.3	-0.731	0.470	
pre.env_k	-0.291	0.363	101.3	-0.802	0.425	
pre.env_l	-1.034	0.212	79.7	-4.866	0.000	***
pre.env_m	0.113	0.376	162.3	0.302	0.763	
pre.env_p	-1.380	0.714	484.6	-1.934	0.054	
pre.env_r	-1.070	0.199	29.5	-5.368	0.000	***
pre.env_s	-0.323	0.426	227.8	-0.757	0.450	
pre.env_ʃ	0.686	0.625	480.3	1.098	0.273	
pre.env_t	-0.324	0.162	16.6	-2.002	0.062	
pre.env_z	0.209	0.454	222.8	0.460	0.646	
duration	-2.276	0.205	1203.0	-11.124	0.000	***
setting_family	-0.177	0.068	5.7	-2.582	0.044	*
setting_services	-0.074	0.102	1.5	-0.715	0.571	

Note. * $p<0.05$; ** $p<0.01$, *** $p<0.001$

Appendix C: Perception experiments instructions

Experiment 1.

Slide1

Hello!

In this experiment you will hear people speaking on a variety of different topics. You will hear 4 different recordings for each speaker. The recordings were made in conversations with other people, so if you sometimes hear silence, that means the speaker's interlocutor is speaking.

Your goal is to rank the recordings on the scale of accentedness from "I can hear a very strong foreign accent" to "I cannot hear a foreign accent at all".

Type a key to continue...

Slide 2

At the top of the screen you will see 4 different symbols associated with 4 different recordings for each speaker. You can play and re-play the recordings by clicking on the symbols.

At the bottom of the screen you will see a scale. If you click on the scale after you've clicked on one of the symbols, you can place that symbol on the scale, thus ranking the recording accordingly.

Type a key to continue...

Slide 3

If you decide to change the ranking of a recording, simply click on the symbol on the scale that you want to move and then click in the new position.

Once you are done ranking the 4 recordings for a speaker, click on "Done" on the bottom right on the screen, and that will take you to the next speaker.

Let's practice now!

Type a key when you're ready to begin...

Slide 4. After the practice trial.

If you have any remaining questions, please ask the research assistant now.

If you are ready to begin the experiment, type any key...

Experiment 2.

Slide 1

Hello!

In this experiment you will hear recordings of people speaking on a variety of different topics. The recordings were made in conversations with other people, so if you sometimes hear silence, that means the other person is speaking.

There are 3 questions that you will answer about every speaker.

Type a key to continue...

Slide 2

First, you will rate the speaker on the scale from 1 (Definitely a first language speaker of English) to 7 (Definitely a second language speaker of English). It can be any variety of English, not just New Zealand English. Please use keys 1 through 7 to indicate your answer.

Type a key to continue...

Slide 3

Next, you will be asked to name a country or a region where you think the speaker was born and raised. Try to be as precise as you can. The speakers may be from different English-speaking and non-English-speaking countries.

Lastly, please comment on what made you think that the speaker is from that particular place: for example, is it what they said, how they said it, or something else?

Type a key when you're ready to begin...

Experiment 3.

Audio condition.

Slide 1.

Hello!

In this experiment you will hear 24 different people speaking on a variety of topics.

Your goal is to rate the recordings on the scale of accentedness from "I can hear a very strong foreign accent" to "I cannot hear a foreign accent at all" (this person sounds like a native speaker of New Zealand English).

Let's practice now!

Type a key to continue...

Slide 2. After the practice trial.

If you have any remaining questions, please ask the research assistant now.

If you are ready to begin the experiment, press any key now.

Video condition.

Slide 1.

Hello!

In this experiment you will see 24 different people speaking on a variety of topics; however, you will not hear what they are saying.

Your goal is to rate the recordings on the scale of accentedness from "I think this person has a very strong foreign accent" to "I don't think this person has a foreign accent at all" (a native speaker of New Zealand English).

Let's practice now!

Type a key to continue...

Slide 2. After the practice trial.

If you have any remaining questions, please ask the research assistant now.

If you are ready to begin the experiment, press any key now.

Audiovisual condition.

Slide 1.

Hello!

In this experiment you will see 24 different people speaking on a variety of topics.

Your goal is to rate the recordings on the scale of accentedness from "I can hear a very strong foreign accent" to "I cannot hear a foreign accent at all" (this person sounds like a native speaker of New Zealand English).

Let's practice now!

Slide 2. After the practice trial.

If you have any remaining questions, please ask the research assistant now.

If you are ready to begin the experiment, press any key now.

Appendix D: Permissions from publishers

Dear Ksenia Gnevsheva,

Your email requesting permission as mentioned below was forwarded to me by my colleague Susan.

Herewith I have the pleasure to give you permission to use your own articles for a part of your PhD thesis.

Article: "Style-shifting and intra-speaker variation in the vowel production of nonnative speakers of New Zealand English"
Expected to be published in our journal: Journal of Second Language Pronunciation 1:2 (2015), pages: 135-156.

Article: "Acoustic analysis in the Accents of Non-Native English (ANNE) corpus"
Expected to be published in our journal: International Journal of Learner Corpus Research 1:2 (2015), pages: 256-267.

This permission is given on condition that full acknowledgement of the original source is given.

Best regards,

--

Ineke Elskamp
Rights & Royalties Department
John Benjamins Publishing Co.
Postal address: P.O. Box 36224, 1020 ME AMSTERDAM, The Netherlands
Visiting address: Klaprozenweg 75G, 1033 NN AMSTERDAM, The Netherlands
Tel.: +31 (0)20 6304747
Fax: +31 (0)20 6739773
Web: www.benjamins.com
E-mail: rights@benjamins.nl

FACULTY OF ARTS
CHARLES UNIVERSITY
IN PRAGUE

Institute of Phonetics

Herewith I have the pleasure to give you permission to use your own article for a part of your PhD thesis.

Article: Gnevsheva, K. (2015). Passing for a native speaker: Production and Perception. *Proceedings of the EPIP4 4th International Conference on English Pronunciation: Issues & Practices*, pp. 37–40. Prague: Institute of Phonetics.

This permission is given on condition that full acknowledgement of the original source is given.

doc. PhDr. Jan Volín, Ph.D.
EPIP4 – Head of Organizing Committee

nám. Jana Palacha 2, 116 38 Praha
Czech Republic

Tel: (+420) 221 619 100
Fax: (+420) 221 619 310

fu@ff.cuni.cz
http://fu.ff.cuni.cz