Learning foreign and native accents: the role of production and listening

D i s s e r t a t i o n

zur Erlangung des akademischen Grades

Doktor der Philosophie

in der Philosophischen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Ann-Kathrin Grohe, M.A.

aus Heilbronn

2017

Gedruckt mit Genehmigung der Philosophischen Fakultät

der Eberhard Karls Universität Tübingen

Dekan: Prof. Dr. Jürgen Leonhardt

Hauptberichterstatterin: Prof. Dr. Andrea Weber

Mitberichterstatter: Prof. Dr. Gregory Poarch, Prof. Dr. Harald Baayen

Tag der mündlichen Prüfung: 27. Oktober 2017

Universitätsbibliothek Tübingen: TOBIAS-lib

Learning foreign and native accents: the role of production and listening

Acknowledgements

At the beginning of my post-graduate studies, I had the privilege to myself define the topic I wanted to investigate for the following three and a half years. I decided to look at accents. Accented speech is a topic that has always been fascinating to me. I have been raised with a southern German dialect that includes specific accent markers that I myself became aware of only after several semesters of linguistic studies and extensive contact with speakers from other dialect areas in Germany.

I experienced similar observations during my academic stays abroad, for instance in Sydney, College Park (Maryland), Montréal, Toulouse, and Montpellier. As a German High School student, I have never been aware of what exactly constitutes my German accent in English and French. I learned this explicitly in university courses, but mostly from native speakers of the respective language. With the awareness of the specific German accent markers in English or French, I realized that I as a native German speaker understand German-accented English and French with no further effort. The same applies for my own southern German native accent.

Whereas other people who did not grow up in the same southern German area frequently have problems understanding that specific accent, I myself as a dialect speaker do not experience any difficulty. I was wondering, whether this is due to the fact that I myself speak the same accent, or because I hear it frequently from other accent speakers. This is how I finally formulated the topic of the present dissertation that I have been working on with my advisor Andrea Weber.

Andrea supported me a lot in this project with very helpful advice and fruitful discussions. Working in her group and contributing to establish the *LingTüLab* has contributed a lot to my professional and personal development. I enjoyed and appreciate our personal discussions as well as the meetings with the group who provided valuable feedback. Andrea taught me what to mind in conducting experiments in psycholinguistics, how to formulate relevant research questions, and how to communicate findings convincingly in different channels. She made possible for me to visit big conferences where I received important feedback from other researchers from the field. I was able to travel to the University of Maryland and the MARCS Institute in Sydney for several research stays, and assist many relevant conferences and workshops with Andrea's support. Therefore, in this place, I want to thank you very much, Andrea!

I also thank Greg Poarch, who I worked with in a few projects during and after his time as Junior Professor at the English Department in Tübingen and with whom it was always a pleasure to watch soccer matches ☺. Moreover, I thank all other (current, former, and affiliated) people from the *LingTüLab* for being a great team and supporting my research projects. Thank you, Sara Beck for being a great office colleague, Maria Lazareva for great and reliable Hiwi work during workshops and being a supportive office colleague, Brock Schardin for great English language support and always being available for all kind of Hiwi work, Thanh Lan Truong for being extremely organized, reliable, and creative, Johanna Geyer for super reliably testing the participants for Monsterlesen, Sabrina Bucher for great support in eye-tracking, scripting, and analyzing data, Achim Skuta for still being a great "Computer-Hiwi" and more importantly a friend, Holger Mitterer for helping with statistics and many other questions, and Karin Klett for great coffee and conversations!

I thank Kathi Zahner, who supported my work (within and beyond this dissertation), especially during our mutual research stay at the MARCS Institute in Sydney. We had a great time in Australia and became quite good travel mates! I also thank Heather Kember not only for her extraordinary travel agency services, but for being such a reliable and clever collaboration partner and friend. Thank you also, Jenny

Yu for your last minute English language support! During my stay in Sydney, I learned a lot and got valuable feedback and started collaborations with Anne Cutler and Heather Kember from Sydney together with the SpeechNet BaWü researchers (in the Labs of Andrea Weber, Bettina Braun, Sabine Zerbian, Adriana Hanulíková). Within the scope of this dissertation, I also spent a few weeks at the University of Maryland, allowing me to collect data with monolingual American English speaking students. This was possible due to the Maryland-Tübingen cooperation and with greater support by the people in Maryland, especially Colin Phillips, Bill Idsardi, and Julia Buffinton (who did great work in coordinating participants).

I thank my friends and family, who always supported me in this project — if not with their expertise, then mentally — especially my mom Siglinde! Eventually, I thank everybody who participated in my studies in many different Labs and sometimes had to be quite patient, especially during the eye-tracking sessions. In summary, being able to write these words makes me proud of my own work, and endurance, which, however would not have been possible without other people's (also those that I forgot to mention here) support.

Contents

Chapter 1

Introduction

Due to the increasing pressures of globalization, people with diverse native language backgrounds need to communicate with one another. Frequently, English is used as the lingua franca (Crystal, 2012). By definition, non-native speakers of English did not learn English during early childhood[1] and therefore likely have an accent in their second language (L2) English — a *foreign accent*. However, native (L1) speakers of English also do not all produce the same variant of English themselves. For example, American English is different from British English, and within each of these variants many regional varieties exist as well. In the present dissertation, these varieties are referred to by the term *native accents*. Similarly, German also consists of numerous native accents, with Swabian, spoken in the southwestern state of Baden-Württemberg, being a prominent example.

Therefore, in their everyday lives, L1 and L2 listeners are frequently exposed to foreign and native accents. Both accent manifestations can make speech comprehension difficult — even native German speakers with little prior experience with the southern German Swabian accent initially have problems understanding a Swabian accented speaker. However, listeners are quite skilled at accounting for these variations: accent learning studies have shown that even brief listening experience with an accent significantly increases accent comprehension (e.g., Clarke & Garret, 2004). At the same time, many people with extensive listening experience with a specific accent are themselves speakers of that accent. Bent and Bradlow (2003), for instance, showed that the intelligibility of a high-proficiency speaker from the same L1 background as the L2 listener could be equal to that of an L1 speaker. This raises questions regarding the role of speakers' individual production experience in the process of accent learning. Both listening experience and self-production potentially

[1] Following the assumptions by Werker and Tees (1984), a child has to be exposed to a given language by the age of six months to learn its sound system in a native-like manner, i.e., being able to detect the relevant phonetic contrasts. This faculty decreases after that age, resulting in a greater likelyhood of a foreign accent in that language, when it is learned later on.

promote accent learning. The present dissertation investigates how each, production and listening experience, contributes to accent learning in both L1 and L2 listeners.

Specifically, the central question is whether and to what extent producing, compared to listening to, accented words contributes to accented word learning and accent learning more generally. Learning effects of accent production are compared with learning effects of listening to accented speech. From these studies, conclusions are drawn regarding the nature of accent learning and how learning mechanisms induced by listening and production relate to each other. Speech material recorded by L1 and L2 speakers is used to investigate whether the processes involved in foreign accent learning are the same as those of native accent learning.

The importance of listeners' native language background is investigated by testing both L1 and L2 listeners in the present experiments. A further goal is to characterize the learning effects in terms of the processing levels by using different experimental paradigms. Reaction time and eye-tracking tasks investigate the effects of learning on online processing, and memory tasks examine the effects on memory recognition. Finally, accent learning as well as learning with production versus listening is considered together with salience. Accented words might be more salient than canonical words, and produced words might be more salient than words that are listened to.

The present introductory chapter first provides an overview on prior findings regarding foreign and native accent learning with listening training. This is extended with findings on learning with self-production. Based on this foundation, the research questions that motivate this dissertation are presented together with an outline of the experiments that seek to answer these questions.

Accents

A definition of accent

An accent is an indexical cue that provides information about a talker's home region and language background. Alene Moyer defines accent as "a set of dynamic segmental and suprasegmental habits that convey linguistic meaning along with social and situational affiliation" (Moyer, 2013). This general definition of accents accounts for accents produced by both native and non-native speakers of a given language. Accents produced by native speakers are frequently designated with the term *dialect*. In the present definition, however, accent exclusively refers to the realization of speech sounds, i.e., the phonetics and phonology of a language variety, whereas dialect describes a language variety that has its own vocabulary, grammar, discursive style — and accent.

On the sound level, an accent can be defined by segmental peculiarities, as reflected in the modification of sounds, the omission or insertion of sounds, or specific sound substitutions. For example, speakers of the southern German Swabian accent (a native accent) substitute /st/ with /ʃt/ when it occurs across syllables: *Zahnbür/st/e 'toothbrush'*, is pronounced as **Zahnbür/ʃt/e* (Vogt, 1977). Sound substitutions in a foreign accent also characterize German-accented English, where the /θ/ is replaced with /s/: /θ/*eft* is pronounced as **/s/eft* (Hanulíková & Weber, 2012). Furthermore, suprasegmental peculiarities contribute to an accent (e.g., Munro, 1995; Trofimovich & Baker, 2006; van Els & de Bot, 1987); for example, intonation, rhythm, pitch, length, tempo, and loudness.

This, together with linguistic meaning, conveys social and situational affiliation. Social affiliation includes factors such as age, gender, regional background, level of education, and social class. Situation affiliation means that speakers use an accent to position themselves vis-à-vis others. A further discriminative feature of accents has been suggested by John Christopher Wells. He defines accents in English as a "pattern of pronunciation used by a speaker for whom English is the native language" (Wells, 1982, p. 1). Wells distinguishes the accent produced by a native speaker, called *native accents*, from *foreign accents* that are produced by non-native speakers. Therefore, a foreign accent is modulated by a speaker's native language, whereas a native accent represents a local variation of a language. The present dissertation adapts these definitions. The term *accent* solely refers to the pronunciation level and the following investigations refer to both accents produced by native speakers (native accents) as well as accents produced by non-native speakers (foreign accents).

Native and foreign accents

Bent, Atagi, Akbik, and Bonifield (2016) summarized relevant findings on the properties of native and foreign accents. In native accents, the difference frequently lies in vowel realizations. For example, speakers from northern cities in the United States realize the *Northern Cities Chain Shift*, i.e., fronting and lowering of /ɑ/ in *hod* and raising and fronting of /æ/ in *had* (Clopper, Pisoni, & de Jong, 2005). Native accents are also marked by sound changes in consonants; for example, word initial fortis plosives /p/ and /t/ are aspirated in Standard German, but not in Austrian German, e.g., *Pinsel 'brush'* is pronounced as *Binsel* (Siebs, Boor, Moser, & Winkler, 1969). Moreover, changes on the prosodic level (Clopper & Smiljanic, 2011; Robb, Maclagan, & Chen, 2004) are possible.

Non-native speech, on the other hand, is a result of the contact between a speaker's native and second language in their mind (Bent et al., 2016). For example, native German speakers frequently replace the voiceless interdental fricative (/θ/ as in *theft*) with an /s/ resulting in the nonword *seft* (Hanulíková & Weber, 2012). In this case, a sound that is not part of the speaker's native language sound inventory is replaced with a sound that is included in that inventory. Therefore, the speaker's native language background plays an important role in the nature of a foreign accent (e.g., Best & Tyler, 2007). However, there are also factors that are not native language specific. These factors are manifested in many different foreign accents produced by people with diverse native language backgrounds; for example, slower speaking rate, more hesitations, less F0-movements, and greater phonetic and phonological variability than native speech (Baese-Berk & Morrill, 2015; Gut, 2012; Wade, Jongman, & Sereno, 2007). This results in greater variability of foreign accents over native accents.

The present dissertation focuses on peculiarities on the segmental level of sounds. Compared to changes of the intonation pattern, segmental changes play a more important role in the perception of a foreign accent (Jilka, 2000; Sereno, Lammers, & Jongman, 2016; other studies, however emphasize the role of nonsegmental accent markers: Anderson-Hsieh, Johnson, & Koehler, 1992; Munro & Derwing, 1999). As mentioned above, segmental peculiarities include the substitution of one specific sound with a second one. Single accent markers differ in their strength resulting in varying salience of the respective marker. If several of these specific accent markers co-occur, a so-called *global accent* emerges.

For example, the German accent in English represents a combination of many different concrete accent markers. In addition to the /θ/-replacements described above, Germans frequently replace /æ/ with /e/, i.e., *sat* is pronounced as *set*, and Germans also tend to replace a voiced plosive in word final position with its voiceless counterpart, e.g., *pub* is pronounced as *pup*, *dog* is pronounced as *dock* (Swan, 2001). Many other accent markers contribute to the global accent and they can also be situated on the suprasegmental level. The present thesis considers both global accents and specific accent markers with a focus on specific accent markers.

The definition presented above includes that native accents are regional varieties of a given language that do not have a major influence of a different language (the speaker's mother tongue). Therefore, native accents may not deviate as much from the standard pronunciation and are therefore more similar to one another than foreign accents. Bent and colleagues (2016), however, point out that this is not always the case. For instance, Standard Southern British English (SSBE) and Dutch are more similar to one another in terms of rhythmic properties than SSBE and Glaswegian English (White, Mattys, & Wiget, 2012). In their study, Bent and colleagues (2016) used a free classification task where American English listeners grouped 24 different native and non-native accents (six US regional dialects, six international English dialects, 12 non-native accents) and found that their participants could differentiate between native and foreign accents. An aspect that may help to differentiate between the two is the amount of phonological and phonetic variability, which was found to be greater within non-native speech than in native speech (Baese-Berk & Morrill, 2015; Wade et al., 2007).

Accent learning

Many L2 learners are affected by a foreign accent. The earlier a language is learned, the weaker the foreign accent (Flege, Munro, & MacKay, 1995). However, foreign accents occur with great variability and even late L2 learners can reduce their accent with sufficient training (e.g., Flege, Bohn, & Jang, 1997). Therefore, training helps reduce an L2 speaker's accent. In addition, training can also help accented speech comprehension. With sufficient accent exposure, a listener can learn an accent and overcome initial processing costs (e.g., Clarke & Garrett, 2004). This is true for foreign accents, but native accent learning has also been observed. In the following sections, studies investigating the learning of both foreign and native accents are discussed, followed by an evaluation of similarities and differences between foreign and native accent processing.

In their L1, listeners adapt rapidly to foreign-accented speech with a long-lasting effect. Clarke and Garret (2004), for example, investigated accent learning with a cross-modal word verification task. In the training phase of their study, L1 English participants listened to Spanish- or Chinese-accented speakers for one minute. In the following test phase, they then listened to single accented sentences produced by the same speaker, followed by a visual word verification task. The brief exposure to an accented speaker was sufficient for L1 English listeners to overcome initial processing difficulties for sentences spoken by these speakers. In three cross-modal priming experiments, Witteman, Bardhan, Weber, and McQueen (2015) found quick, automatic, and reliable adaptation effects to both words with a global accent and words with an additional specific accent marker. L1 listeners of Dutch adapted to Dutch words with a global Hebrew accent and to Dutch words where the Hebrew speaker specificially shortened words with [i] to [ɪ], e.g., Dutch *stati:f* '*tripod*' pronounced as **statɪf*.

Rapid learning of specific accent markers was also investigated for synthesized speech. Maye, Aslin, and Tanenhaus (2008) artificially created speech stimuli where all English front vowels were lowered so that, for example, *witch* was pronounced as **wɛtʃ*. In a training phase, native English participants listened to a short story with the accent for about 20 minutes. Their recognition of single accented words was tested afterwards in a lexical decision task. The proportion of accented words that were accepted as words, was significantly higher after training with the accent than before training. Accent learning in L1 after just brief accent exposure was also observed in a further cross-modal priming study by Witteman, Weber, and McQueen (2013). Again, initial processing difficulties of words with a strong German accent marker (e.g., /œy/ in *huis 'house'* pronounced as /ɔɪ/) was quickly overcome by the Dutch participants.

Only after exposure to a four-minute story in German-accented Dutch did strongly-accented words elicit facilitatory priming.

Learning processes have also been investigated for native accents and it has been shown that longer-term experience with a native accent facilitates accent processing. For example, Sumner and Samuel (2009) measured priming effects for words with the final r-dropping that is typical for New York City (NYC) English. Listeners with extensive experience with the NYC-English accent were compared to listeners with limited experience and the experienced group showed greater priming effects for the accented words. Similarly, Adank Evans, Stuart-Smith, and Scott (2009) found that only listeners who were familiar with both Standard Southern British English (SSBE) and Glaswegian English (GE), and not listeners who were only familiar with SSBE (originating from the Greater London area), showed equal performance on both accent types in a sentence verification task.

Evidence that the familiarity advantage results from learning processes, was provided by Evans and Iverson (2007). They tested university students who were originally from northern England, but studying in southern England. The students were tested several times over a period of two years with several production and comprehension tasks. Each testing session consisted of a reading task and two perception tasks. First, the students were asked to choose best exemplar locations for vowels embedded in either northern or southern English accented carrier sentences. Second, they accomplished a sentence recognition task in which they identified words in noise spoken with either a northern or a southern English accent. The students changed their accent to educational norms after attending university. Whereas perception did not change over time, the participants chose similar vowels to the one they produced, and students with a more southern English accent showed better performance in identification of southern English speech in noise. A more recent study

(Trude & Brown-Schmidt, 2012) found that adaptation to a native accent occurs even within one experimental session (referred to as short-term adaptation or learning).

Native versus foreign accent processing

Taken together, both foreign and native accents first slow down processing, but adaptation to these variants is possible. A direct comparison between foreign and native accent processing was provided by Floccia, Butler, Goslin, and Ellis (2009). Reaction times to English words embedded in sentences with both regional (Plymouth and Irish English) and foreign accents (French accented English) were slowed down in the accented context, but the effect was greater for the foreign accent than the regional accent. Therefore, there is either a difference in processing between foreign and native accents, or other factors play a role, such as speaker-specificity. In this section, both options are discussed.

Current hypotheses on accent learning. Processing mechanisms for foreign and native accents are either the same or they differ from one another. Goslin, Duffy, and Floccia (2012) refer to these two options by the *Perceptual Distance Hypothesis* and *Different Processes Hypothesis*. The Perceptual Distance Hypothesis states that the same mechanism controls foreign and native accent processing, but this mechanism differs depending on the nature of the accent. For example, it is more refined or attuned for native accents than for a completely foreign accent. This view is in line with the *accent processing classification* by Clarke and Garrett (2004). They postulate that accent learning is solely determined by the accent's acoustic distance from native speech. The same principles are suggested for foreign and native accent learning, but the ease of accent learning is determined by accent strength. In an accent learning setting, stronger accents need more time or more intense training.

This view is supported by the following two studies. In a series of experiments, Larraza, Samuel, and Onederra (2016b) compared early Basque-Spanish bilinguals speaking the Standard Basque dialect to early Basque-Spanish bilinguals speaking the western Basque dialect. The Standard dialect differentiates between the apico-alveolar fricative /s̺/ and the lamino-alveolar fricative /s̻/; whereas the western dialect only has the merger /s̺/. Perception of the alveolar fricative sound contrast that only exists in the standard variant (/s̺/ versus /s̻/) was tested in an AXB discrimination task and the Standard Basque group's results were superior to the western Basque group's results. In a lexical decision task, acceptance rate of tokens was also tested. Tokens corresponded to real words in the western dialect merger (/s̻/ was replaced with /s̺/), but were nonwords in Standard Basque. The western group accepted significantly more nonwords than the Standard group. There was no effect in a third experiment that tested semantic priming. Larraza and colleagues concluded that dialectal effects on speech processing manifest themselves on the sound level and lexical level, and not on the semantic level. Therefore, these effects solely concern the pre-lexical level.

In a second study (Larraza, Samuel, & Onederra, 2016a), using the same material as discussed above, Larraza and colleagues found that Spanish-Basque late bilinguals (age of acquisition: 7) who were non-native speakers of Standard Basque produced a similar pattern of results. They demonstrated their sound discrimination abilities on the sound level, and on the lexical level, but not in the semantic priming task. This suggests that the native influence is situated along a graded dimension (Cutler, 2012), implying the same processes for native and foreign accents.

The Different Processes Hypothesis, on the other hand, assumes distinct processes for regional and foreign accents. Evidence for this hypothesis was found in an ERP-study by Goslin, Duffy, and Floccia (2012). They used three accents from three different areas, testing students from the South-West of England. The regional distance of these locations from the study area varied from very close to more distant (home accent: South-West of England, regional accent: South Wales and Yorkshire, Leeds; foreign accent: North of Italy and Poland). The results of the ERP study suggest different strategies for regional accent processing compared to foreign accent processing. Whereas unfamiliar regional accents are normalized at a pre-lexical/phonological level, foreign accents have a continued effect at lexical access.

These studies come to different conclusions by supporting opposing theories. Acoustic distance between different accent groups are relevant in these accounts on foreign versus native accent processing. In the studies by Larraza and colleagues (2016a, 2016b), the acoustic distance between Standard Basque and the western Basque accent is likely smaller than the distance between the Basque varieties and Spanish. Also, in the Goslin study (Goslin et al., 2012), the acoustic distance between the South Wales and Yorkshire accents (regional accents) and the southwestern English home accent, on the one hand, is likely smaller than between the Italian and Polish accents (foreign accents) and the English accent on the other hand. However, no detailed acoustic information on the accents used in the studies above is available. So, there is no objective measure that allows us to compare the single acoustic differences within and particularly across studies. Therefore, at this point, no definite conclusion is drawn on whether the processing of native versus foreign accents is related or not.

Speaker-specificity of accent learning. Speaker-specificity has been discussed in the context of both foreign and native accent learning. Accent learning is speaker-specific, when it is restricted to the training speaker and does not generalize to new speakers with the same accent. Prior research suggests that native accent learning is speaker-specific. In a short-term native accent learning study including a relatively short training phase followed by a test, Adank and McQueen (2007) tested different speakers with the same accent during training and test and did not find accent learning. Trude and Brown-Schmidt (2012) used the same native accented speaker in their two experimental phases and observed learning effects. Studies that tested speaker-specificity in foreign accent learning, on the other hand, provided mixed results. In a sentence transcription task, for instance, generalization of accent learning (Chinese-accented English) to new speakers was only possible if the listener was exposed to multiple speakers with the Chinese accent in English during training (Bradlow & Bent, 2008). Similar results were reported in Sidaras, Alexander, and Nygaard (2009), and Tzeng, Alexander, Sidaras, and Nygaard (2016). In short, native accent learning seems to be speaker-specific, and the situation is not clear for foreign accents.

Listeners' native language background

Next to the speaker's native language the native language background of the listener plays a role in accented speech processing. For example, in two different studies, simultaneous and late bilinguals of Basque completed an auditory lexical decision task with words including cases of the western Basque accent merger where/s̻/ is replaced with /s̺/ (Larraza et al., 2016a, 2016b). The simultaneous bilinguals were native speakers of Spanish and Standard Basque (not the western dialect) (Larraza et al., 2016b). The late bilinguals (Larraza et al., 2016a) were native speakers of Spanish and learned their L2 Basque by the age of seven. The early bilinguals accepted 79 % of words with the western dialect merger, whereas the late bilinguals accepted only 71 %.

This is a definite sing of differences in accent processing between L1 and L2 listeners, but it is not so clear in which way rapid learning of accented speech (foreign or native) differs between L1 and L2 listeners. Several accent learning studies tested foreign accent learning by L1 listeners (e.g., Clarke & Garrett, 2004; Witteman et al., 2013; Witteman et al., 2015) and native accent learning by L1 listeners (e.g., Adank & McQueen, 2007). Studies with L2 listeners have concentrated on the effects of long-term accent experience, showing that this experience facilitates L2 listeners' accent comprehension in a laboratory setting (Bent & Bradlow, 2003;

Hanulíková & Weber, 2012; Weber, Broersma, & Aoyagi, 2011). Therefore, no study so far has looked at short-term accent learning by L2 listeners, and the direct comparison between short-term learning by L1 and L2 listeners requires further specification.

In summary, foreign and native accents cannot be distinguished in an absolute way. However, prior research suggests that listeners can differentiate between native accents and foreign accents (Bent et al., 2016) and other studies (Adank et al., 2009; Clarke & Garrett, 2004; Goslin et al., 2012; Weber, Di Betta, & McQueen, 2014) hint at differences between foreign and native accent processing. In this dissertation, the terms and concepts of foreign accent and native accent are defined by the information whether speech is produced by a native speaker or a non-native speaker. Using both types of accents can provide important evidence for or against the distinction between foreign and native accents.

Models of spoken word recognition

From the studies discussed so far, we know that listeners are quite flexible in coping with variability in speech. How is this is possible? Which kind of structure and processes in the mental lexicon give rise to these learning effects? Models of spoken word recognition address these questions. There are two major groups of these models, abstractionist accounts (McClelland & Elman, 1986; Norris, 1994; Norris & McQueen, 2008) and episodic accounts (e.g., Goldinger, 1998) of the mental lexicon. The episodic accounts suggest the storage of every concrete exemplar of a speech unit encountered by a listener (including speaker-inherent details as for example voice and accent properties), whereas in abstractionist models, abstract representations of a word's canonical representation build the lexicon.

Variations of the canonical form, such as accents, can be accounted for by pre-lexical mapping rules. These rules are founded on a few exemplars that are no longer stored. When, for example, an accented token is encountered after accent training, the learned rule is applied to the respective abstract entry in the lexicon. This explains why learning a specific variation can generalize across many different words (McQueen, Cutler, & Norris, 2006). McQueen, Cutler, and Norris (2006) also suggest that in an abstractionist account, processing might be probabilistic. In Shortlist B (Norris & McQueen, 2008), an abstractionist model relying on Bayesian principles, word activation depends on both a potential candidate's prior probabilities, and the current evidence in favor of them. The more often a candidate is encountered, the higher is the prior probability. This means that a candidate's so-called resting activation level is increased, which accounts for frequency effects (for example, the more often a word is encountered, the faster it is recognized). Additionally, there also exist hybrid models (e.g., McLennan, Luce, & Charles-Luce, 2003) that borrow the ideas from both groups of models of spoken word recognition. They can, for example, integrate exemplars and pre-lexical mapping rules into a single account.

Which account is most applicable, can be probed with generalization effects across voices (different voices during a training and a test phase) or across words (the same accent presented in a training phase included in new words during the test phase). These learning effects are either based on listening experience, but they might also rely on an indiviual's own production experience. Learning with production, however, requires connections between speech production and comprehension that are elaborated in the following paragraph.

Connections between speech production and comprehension

The question of the relationship between language production and comprehension has become an increasingly popular object of study during the last few years. Meyer, Huettig, and Levelt (2016) recently edited a special issue on this topic in the *Journal of*

Memory and Language. In their editorial note, they summarized each study included in the special issue. For example, Kittredge and Dell (2016) investigated the effects of listening on participants' productions, and Baese-Berk and Samuel (2016) as well as Zamuner, Morin-Lessard, Strahm, and Page (2016) looked at the effects of participants' own speech productions on single sound perception and non-word learning.

From these results and other studies, conclusions about the nature of the relationship between production and comprehension representations are drawn. Kittredge and Dell (2016) distinguish three different positions. The first position says that the representations for production and comprehension are kept separate with no influence on one another. The second position includes inseparable representations, implying identical representations for production and comprehension. The third position postulates separable representations. This latter position includes different representations for production and comprehension that can, however, influence one another under certain conditions because there are strong links between production and comprehension representations. The nature of production and comprehension representations has been discussed by most papers included in the special issue, as Meyer and colleagues summarized. They observed that there is a clear tendency towards the third position (separable representations)[2].

Connections between speech production and speech comprehension was also observed in neurophysiological studies. For example, monkeys activate the same neurons when performing an action and when observing somebody performing a similar action (e.g., Rizzolatti & Craighero, 2004). These visiomotor neurons are situated in area F5 of the premotor cortex and are called *mirror neurons*. Studies with humans suggest the existence of mirror neurons also in human beings (for a review: Rizzolatti, Fogassi, & Gallese, 2000). Further evidence for the connection between

[2] None of the studies advocates the first position (separate representations), and only one study (Chater, McCauley, and Christiansen, 2016) advocates the second position (inseparable representations).

production and comprehension was found in a transcranial magnetic stimulation (TMS) study by Watkins and coworkers (Watkins, Strafella, & Paus, 2003). MEP (motor evoked potentials) responses, which are small lip movements, were provoked by stimulating the motor cortex that controls movements of the face. These MEPs were amplified when the study participants listened to speech or when they watched a second person moving their lips without producing an audible speech signal.

Models integrating production with comprehension

Despite its trending nature, the production-comprehension link is not new to the research community: in the 1950s, it was discussed specifically in phonetic models of speech perception. Both Stetson (1951) and Liberman (1957) emphasized that speech perception relates to articulation. Subsequently, different models and experimental paradigms that include the production-comprehension link were developed. These models and paradigms all assume that speech perception affects speech production. For example, Stevens' (1960) *analysis-by-synthesis* model states that the listener first derives a *spectral representation* from the speech signal. This representation is then converted into an articulatory description. In turn, the so-called *active synthesis* creates a matching spectral representation in the articulatory description.

Alvin Liberman and colleagues (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, & Mattingly, 1985; Liberman, & Whalen, 2000) developed a variation of Stevens' approach, the *motor theory of speech perception*. It claims that listeners match the acoustic signal with associated articulatory movements. When hearing a [b], a listener matches the sound with lip closing. The next step is the matching of the articulatory movement onto an *intended gestural pattern*. In the [b]-example, this gestural pattern is the abstract category /b/. This theory is supposed to explain how speech comprehension is possible despite the great variability that is naturally included in the acoustic speech signal. For successful comprehension, listeners map the acoustic speech signals onto abstract phonemic categories. Liberman and colleagues argue that in contrast to the acoustic speech signal, the motor properties of speech production are relatively invariant. The reference to these invariant properties helps the listener to match the variable acoustic signal with abstract phonemic categories. From the assumption that speech perception equals the perception of intended articulatory gestures follows that speech production and speech perception are closely connected with one another.

The experimental picture-word interference paradigm (Glaser & Düngelhoff, 1984) makes use of the effects of speech perception on speech production. In this paradigm, pictures are presented visually together with a printed word. Interference effects are observed when the printed word does not describe the picture, i.e., is used as a distractor. This research procedure was extended to spoken words instead of printed words (Schriefers, Meyer, & Levelt, 1990). The printed distractor word allows investigating semantic distractor effects, and the spoken distractor word allows investigating semantic and phonological distractor effects. Effects of speech perception on production are also made use of in the syntactic priming paradigm by Bock (1986). When participants first hear a given sentence structure provided by the experimenter and then are instructed to describe a picture, the participant uses the same syntactic form in the picture description task as presented by the experimenter beforehand.

In more recent models, the effects of speech perception on production are explained on the grounds of the concept of prediction, for example in the *P-chain model* by Dell and Chang (2014), the *CAPPUCCINO model* by McCauley and Christiansen (2011), and the *integrated theory of language production and comprehension* by Pickering and Garrod (2013). In their theoretical account, Pickering and Garrod claim that listeners facilitate the comprehension process by imitating the talker's utterance. The listener covertly imitates a speaker's utterance and builds predictions using forward models. There are two different sources for these predictions: the association route and the simulation route. The association route is based on experience with comprehending other people's utterances, and the simulation route refers to the comprehender's individual production experience.

All these theories consider the relation between speech production and comprehension, which can rely on the same or different representations. The nature of production and comprehension representations, however, have also been investigated with memory studies. One major finding of these studies is the *production effect*. This effect describes a memory advantage of produced words compared to words that are either read silently or listened to during a training phase.

The production effect ––findings from memory studies

Producing versus reading silently. Ways in which learning can benefit from production are derived from the principles discussed in early psychological theories. For example, William James (1890) emphasized that activeness crucially facilitates learning. Activeness of encoding was also stressed by Zinchenko, Vygotsky, Leont'ev (cited in Wertsch, 1979) as well as Craik and Lockhart (1972). These theories predict that learning with production results in greater learning effects than learning with a less active modality, such as reading silently or listening.

Current memory literature includes a branch of research that focuses on the effects of voicing a word out loud compared to learning a word by using other methods. Indeed, the findings suggest that self-production results in better word recollection than, for example, reading a word silently (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). This advantage is referred to as the *production effect*, and Colin MacLeod and colleagues have addressed this phenomenon in a series of experiments. In one of their studies (MacLeod et al., 2010), participants first studied a list of words and then performed a memory recognition task. During the study phase, some words were read aloud, and other words were read silently. The participants were asked not to move their lips during the silently-read trials. The memory task was an old/new recognition task, in which participants saw single printed words on a computer screen in front of them and decided with a button press whether the word was old (included in the study phase), or new (not included in the study phase). Words that were read aloud were more likely to be correctly recognized than words that were read silently.

When participants were asked to mouth the words without vocalization, instead of reading them out loud, memory was still improved compared to reading them silently. A further memory study suggests that the advantage of producing a word aloud over reading it silently can even last for at least one week (Ozubko, Hourihan, & MacLeod, 2012). The participants' word memory for words that had been studied with production or by means of silent reading was tested with a memory recognition task directly after the study phase and one week later. After the one-week delay, more produced words were correctly recognized still than words that had been read silently.

Producing versus listening. If word learning with production is more effective than with silent reading, producing might also provoke greater learning effects than listening to a word produced by someone else. Compared to listening, production involves muscle movements and is a more active process than listening. An advantage of learning with production over learning with listening would be in line with James' (1890) early assumptions. And indeed, MacLeod (2011) found that memory for self-produced words is greater than memory for words that are produced loudly by a second person during the same experimental session. In the word learning phase of that study, two participants were sitting together in front of a computer screen while single words appeared on the screen one after the other. One participant read half of the words aloud, and half were read by the second participant. In a following memory recognition test, the words that had been self-produced were more likely correctly recognized than the words that the second person produced.

An eye-tracking study (Zamuner et al., 2016) that compared nonword learning with production to nonword learning with listening shows that the production advantage not only applies to a recall task, but also to online processing. In that study, some auditorily presented words had to be repeated by voicing them out loud, and some were listened to twice. In the subsequent eye-tracking test, there was an advantage of the repetition condition[3].

Accents and the advantage of production

Up to this point, accent learning and learning with production has been referred to separately. These two topics are now discussed together, particularly in relation to how speakers learn an accent with production that is compared to listening based learning.

Production can be advantageous for accent learning. This has been demonstrated, for example, with paradigms where auditorily presented accented speech has to be imitated with the accent. Adank, Hagoort, and Bekkering (2010) tested the effects of imitating a made-up accent in Dutch on participants' subsequent accent comprehension. L1 Dutch participants first completed a pre-training accent comprehension test in Dutch. Dutch sentences with vowel conversions (e.g., /ɛ/ was pronounced as /e:/ and vice versa; /u/ was pronounced as /Y/) were played auditorily with speech-shaped background noise to them and the participants were required to transcribe them. A training phase was then completed. Participants were assigned to

[3] Note that the study by Zamuner et al. (2016) involved a repetition paradigm (auditorily presented tokens had to be repeated aloud) during training, but in order to compensate for the double input in the repetition condition, the listening tokens were presented twice.

different groups that either listened to, repeated (with their own accent), transcribed or imitated sentences with the specific accent. The experiment was concluded with a renewed transcription accuracy test (the same as in the pre-training). Between the pre-training and testing, transcription performance improved most strongly after imitation. Improvement of transcription performance did not differ significantly between the other conditions.

Accent imitation was also implemented in a memory study by Cho and Feldman (2013). In the learning phase of the experiment, all native English participants listened to single words while at the same time a visual cue instructed them to just listen to or repeat the words aloud. All participants produced some words and listened to others. One half of the English words were recorded by a native Dutch speaker, and the other half were produced by a native speaker of American English. During repetition of the Dutch-accented words, the English participants had to imitate the Dutch accent. In a subsequent memory recognition task, participants were presented with words on a computer screen and had to decide whether they were words from training or new words. A greater number of repetition items than listening items were remembered.

The role of salience

Both accent and the advantage of self-production in word learning can be explained by the concept of salience. Generally speaking, salience refers to "how easy it is to perceive a given language form" (Azaz, 2017, p. 5). Consequently, the easier a language form is perceived, the easier it is learned. The two concepts *accent* and *learning with production* are now related to salience. In language studies, salience has been described more specifically as "the property of a linguistic item or feature that makes it in some way perceptually and cognitively prominent" (Kerswill & Williams, 2002, p.

81). Accents, for example, increase a word's salience. This assumption is based on the phonetic difference between two variant forms; the greater the difference, the more a dialect speaker is aware of it (Trudgill, 1986). Therefore, such forms are learned more easily than unaccented words. In addition, factors beyond linguistic or structural properties, such as learning modality, may affect salience. The act of producing a word makes the word more salient than listening to a word or reading it silently because of a greater degree of distinctiveness. This supports the findings from the memory studies reported above (MacLeod et al., 2010; MacLeod, 2011).

Two different types of salience were differentiated in the attention literature — top-down and bottom-up salience (Awh, Belopolsky, & Theeuwes, 2012; Summerfield & Egner, 2009). While top-down salience refers to a generally greater stimulus salience, for example, due to regular practice and the resulting cognitive pre-activation, bottom-up salience describes a stimulus that itself attracts attention because of its distinct, physical characteristics. In sociolinguistics, however, Auer, Barden, and Grosskopf (1998) differentiate between objective and subjective criteria of salience. In relation to accent properties they describe *articulatory distance* for example, as an objective criterion, whereas *perceptual distance* is its subjective counterpart. In its role as objective criterion, articulatory distance describes the magnitude of acoustic deviation of a linguistic token from the canonical realization, which is concretely quantifiable. Perceptual distance, on the other hand, relates to the individual listener's perception that is shaped by their prior experience. It describes the way that a listener perceives this distance. From this information results that subjective criteria elicit top-down salience, and objective criteria refer to bottom-up salience.

I suggest that the two concepts that form the focus of this dissertation (*accent* and *learning with production)* can act as subjective and objective criteria. For example, the definition of accent salience above describes accent as an objective criterion. An accent deviates acoustically from the canonical realization. However, listeners can have individually rich prior experience with an accent, which would refer to a subjective criterion. Production learning might be objectively more salient because the act of production is a more active one than listening, for instance. It may, however, also be a subjective criterion based on the amount of prior experience with production.

The present dissertation

Research questions and overview

The findings, theories, and hypotheses presented thus far establish three major research questions that are investigated in the present dissertation. First, prior studies have shown that canonical word production can facilitate word learning compared to listening to a second person producing the words (MacLeod, 2011). However, it is not clear, whether this also applies for accented speech. Therefore, Question 1 asks to what extent accent production contributes to accent learning compared to accent listening, which, in turn, facilitates accent comprehension.

Adank, Hagoort, and Bekkering (2010) and Cho and Feldman (2013), for example, demonstrated an advantage of imitating accented speech over listening only to accented speech, but they did not investigate the effects of production alone. Therefore, in the present experiments, no repetition or imitation paradigm was applied, but participants were instructed to deliberately produce a specific accent marker. Learning effects of accent production were compared with learning effects by listening to accented speech. From this, conclusions can be drawn concerning the nature of accent learning and how learning mechanisms induced by the two modalities, listening and production, relate to each other.

Second, the exact properties of these learning effects are not clear. For example, prior findings do not provide a clear picture on whether accent learning is speaker-specific (as suggested by Trude and Brown-Schmidt, 2012) or if it can instead generalize across many different speakers (evidence for speaker-general accent learning with multiple training speakers was provided by Bradlow and Bent, 2008). Whether these learning processes differ between foreign and native accents (Perceptual Distance Hypothesis versus Different Processes Hypothesis, see Goslin et al., 2012) as well as between native and non-native listeners (Larraza et al., 2016a, 2016b for L1/L2 differences in accented speech processing) needs further investigation. Therefore, Question 2 asks what characterizes accent learning effects. Also of interest, are the processing levels where accent learning with production and listening training is possible, which would indicate generality of learning. This was examined using different experimental paradigms: reaction time and eye-tracking (similar to Trude and Brown-Schmidt, 2012) tasks investigated the effects of learning on online processing, and memory tasks (similar to MacLeod et al., 2010) examined the effects on memory recognition. The generality of learning with production and listening was further investigated by testing differently familiar accents. Moreover, the questions of how long lasting learning effects are and what the role of self-listening is in learning with production were investigated. Referring to James (1890), the activeness of the learning modality plays an important role. However, speakers may also learn through a multitude of channels because they hear themselves during production.

Finally, Question 3 scrutinizes the role of salience in accent learning and learning with production and listening. Accented words are possibly more salient and therefore more easily learned (referring to Trudgill, 1986); the same might apply for produced words compared to words that are listened to. These three big questions provide the basis for a theoretical explanation for processes behind accent learning with production and listening. The questions were investigated in ten experiments, presented in five chapters, using the various experimental methods described below.

Experimental methods

Three different experimental tasks were integrated in a training-test paradigm. This paradigm, the single tasks, and the methods for analyzing their results are briefly outlined below.

Training-test paradigm. The training-test paradigm refers to an experimental unit that is combined of two phases, a training phase followed by a test phase (as in Witteman et al., 2013). All ten experiments presented in this dissertation are based on this paradigm. In the present experiments, the training involves producing or listening to accented speech. In the following test phase, accent learning strength was tested with different methods, including lexical decision, eye-tracking, and memory tasks.

Lexical decision. In a lexical decision task, single tokens are presented serially and can either be a word in a given language or not (non-words). The presentation is either visual (the orthographic version of the words is printed on the screen) or auditory (the participants hear the tokens, usually over headphones; Goldinger, 1996). For testing accent learning, the tokens are usually presented auditorily. Accent properties are tested that are manifested on the sound level. Participants are required to decide whether the presented token is a word or not in the given language (in the present cases: English or German) using a button press. Typically, participants are instructed to answer as quickly and as accurately as possible. Endorsement rates and reaction times are measured and compared between different conditions.

Eye-tracking: visual world paradigm. In eye-tracking, participants' eye movements and fixations on a given computer screen are recorded (Allopenna, Magnuson, & Tanenhaus, 1998). In the classical version of this paradigm, four pictures are displayed on the computer screen. In the present studies, however, orthographic words were used rather than pictures (McQueen & Viebahn, 2007; Weber, Melinger, & Lara Tapia, 2007). Generally the display in this paradigm includes a target, a competitor, and two distractors.

In the present eye-tracking studies, the competitor was always a phonological competitor (e.g., target *PALME 'palm tree'* and competitor *BALKEN 'beam'*). Each word was printed with its correct orthography on the screen. The words were equally distributed across the screen, as shown in Figure 1-1. As suggested by Huettig and McQueen (2007) and supported by Weber, Melinger, and Lara Tapia (2007), the printed word variant is more sensitive to phonological manipulations than the picture version. On the other hand, its sensitivity to semantic representations is smaller (Huettig & McQueen, 2008). This well-suits to the framework of the present dissertation, as phonological, rather than semantic effects are investigated.

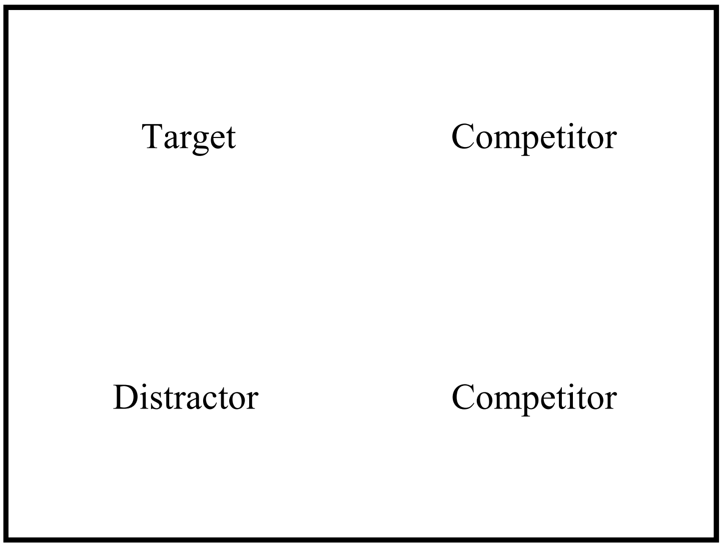

Figure 1-1. *Example display including the target, distractor, and the two competitors.*

In eye-tracking studies of this type, it is important to secure that the position of each token role varies from trial to trial so that the target and the competitor are equally distributed across all four screen positions during the whole experiment. While the screen with all four words is presented, participants are auditorily instructed to click on a target word while their eye movements are recorded. The proportion of target fixations is measured and usually compared between different conditions. For example, in the present Experiments 5–7, if the target was the word *PALME ‘palm tree’*, the competitor was *BALKEN ‘beam’*. If participants learn that *Balken* is pronounced as **Palken* during the training phrase, an increased amount of looks to the competitor *BALKEN* when hearing *Palme* suggests successful accent learning. Including phonologically unrelated distractors allows the conclusion that the competitor induces phonological competition.

For eye-tracking, different devices are available. Many researchers in psycholinguistics make use of large high-end devices that can however only be used for testing sessions in a stationary lab, as for example the *Eyelink 1000* (SR Research Ltd.). In the last few years, however, small, mobile devices have been developed. The *EyeTribe* (https://theeyetribe.com/), for instance, is one such mobile eye-tracker in pocket format and at a considerably lower price range. However, it only offers low sampling rates (up to 60 Hz) and does not have pre-installed software.

Memory tasks. In contrast to the former two methods, the last three experiments did not make use of an online processing task, but rather two different memory tasks. In a free recall task, participants are asked to write down all the words that they remember from the preceding training phase. The analysis refers to the number of correctly remembered words, termed *memory accuracy*. Moreover, an old/new recognition task was run. In this task, words from the training phase and new words are presented one after the other and the participants decide with a button press whether the word was included in the training phase (old word) or not (new word).

Hit rates and false alarms (erroneously accepted new words) are recorded based on which d-prime scores are calculated. The d-prime score is a measure developed in the signal detection theory (Stanislaw & Todorov, 1999) that takes into account hit rates (correctly detected old words) and false alarm rates (erroneous yes-answers to new words) in a memory recognition task. In either task, memory accuracy or the d-prime score is compared between different conditions. For example, in Experiments 8–10, memory accuracy of self-produced words was compared to memory accuracy of listened-to words. This allows for conclusions on whether one training condition strengthens word memory more than a second training condition does.

Statistical analysis. The program R (R development core team, 2017; versions 3.2.2 to 3.3.3) was employed for statistical analysis. In most cases, linear mixed effects regression models (e.g., Baayen, Davidson, & Bates, 2008) were run using the lme4 package (Bates, Maechler, Bolker, & Walker, 2016). In these models, there are fixed-effect factors and random effects. Random intercepts and slopes allow for a maximal random effects structure (e.g., Barr, Levy, Scheepers, & Tily, 2013). They consider differences between participants and items in size of the effect of the fixed variables (Cunnings, 2012). Random slopes reduce the Type I error rate and make the analysis conservative. For each analysis an individual, best fitting model was built that included a particular choice of fixed and random factors. The final, most parsimonious model of best fit that still converges usually includes only significant predictors and interactions. Significance of factors was indicated by t-values > |2|. Corresponding p-values were either calculated based on Satterthwaite approximation by using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2015) or with likelihood ratio tests using the anova-function in R.

In summary, mixed models preserve the maximum amount of information from the original dataset. Moreover, they are well-suited for the inclusion of continuous predictors (e.g., *items' test list position*, *frequency*), and consider random effects, such as individual variation by participants or test items. Possible dependent variables in the models that were calculated in the present experiments are *reaction time* (lexical decision task), *hit rate* (lexical decision and memory recognition task), *proportion of fixations* (eye-tracking), *recall accuracy* (free recall task), *and d-prime score* (memory recognition task).

Outline

The following paragraphs outline the chapters and experiments presented in this dissertation. In Chapter 2A (Experiments 1–3), the effects of production and listening training on subsequent comprehension of foreign-accented speech were investigated in a training-test paradigm. During training, L1 English and German L2 participants either listened to an English short story in which all /θ/s were replaced with /t/ (e.g., *theft* was pronounced by a German speaker as **teft*), they read the story out loud with the *th*-substitutions, or they had no accent training. During a test, participants made auditory lexical decisions to English words with *th*-substitutions. The test words were recorded by a different speaker than the training material to test speaker-general learning effects. Reaction times and endorsement rates of the accented tokens were measured.

Chapter 2B (Experiment 4) addresses whether it is easier to learn an accent when the speaker has an accent similar to oneself. Similar to Experiments 1–3, a training-test design was applied. During training, L1 English participants either listened to a story read by an L1 speaker (no L2 speaker as in Chapter 2A) who replaced all /θ/s with /t/, or they produced the story with the /θ/-substitutions themselves. Reaction times and endorsement rates of the accented tokens were measured.

Chapter 3A (Experiments 5–6) further investigated native accent learning and manipulated salience in the training and test phase. Two experiments tested the degree to which an accent is learned with production or listening training. A training-test paradigm was administered on native German participants utilizing an artificial

German accent. During training in Experiment 5, participants either read single German words out loud and deliberately devoiced initial voiced stop consonants (e.g., *Balken 'beam*' pronounced as **Palken)*, or they listened to pre-recorded words with the same accent. In a subsequent eye-tracking experiment, looks to auditorily presented target words with the accent (Experiment 5) or canonical German words that overlapped in onset with the accented words from training (Experiment 6; e.g., ***Pal**me 'palm tree'* overlapped in onset with the training word **Palken*) were presented as target words. Training and test words were recorded by two different L1 German speakers. Accent was the objective criterion of salience (accented test words in Experiment 5 and canonical test words in Experiment 6), and training modality was the subjective criterion (production versus listening training in both experiments).

Chapter 3B (Experiment 7) targets the methodological aspect of eye-tracking. It includes a replication of Experiment 5 with a different eye-tracking hardware being used, the *EyeTribe* tracker (https://theeyetribe.com/), a mobile low-cost eye-tracker. Experiment 7 investigated whether in a visual world experiment, the *EyeTribe* could provide similar results as the *EyeLink 1000* (SR Research Ltd.).

Chapter 4 (Experiments 8–10) also examines native accents, but refers to long-term accent familiarity. The effects of training and long-term accent familiarity on word memory were investigated in a training-test paradigm. Here, emphasis was placed on investigating the way accents affect word memory rather than accent learning itself. Single words were produced or listened to by L1 German participants (raised in southern Germany with the Swabian accent) in a training session. Training words either had a Swabian accent marker that was familiar to participants (/st/ pronounced as /ʃt/, **Zahnbür/ʃt/e 'tooth brush'*) or an unfamiliar northern German accent marker (/ʃt/ pronounced as /st/, **Blumen/st/rauß* '*bouquet*'). After training, participants completed a visual memory recognition task where they had to decide whether presented words were from the training phase (old words) or new.

Chapter 5 summarizes and discusses the findings of each experimental chapter and integrates them into existing theories and accounts on accent learning and learning with production. This chapter also elaborates on which theoretical assumptions explain the present empirical findings best and are therefore recommended to be further pursued.

Chapter 2A

Accent learning with production and listening in L2 versus L1[4]

Abstract

The effects of production and listening training on the subsequent comprehension of foreign-accented speech were investigated in a training-test paradigm. During training, English native (L1) and German non-native (L2) participants either listened to a short story in which all /θ/s were replaced with /t/ (e.g., *theft* was produced by a German speaker as **teft*), they read the story out loud with the *th*-substitutions, or they had no accent training. During a test, participants made auditory lexical decisions to English words with *th*-substitutions. L2 participants' reaction times to words from the training were significantly faster after having produced the story than after no training, and having listened to the short story also resulted in faster reaction times, but less strongly so. The effects did not generalize to untrained words. For L1 participants, the facilitatory effect of training did not differ significantly between production and listening training. Thus, only for L2 participants, the effect of accent production for adaptation was superior to accent listening. In conclusion, production training affects comprehension processes, but these effects are determined by properties related to an individual's native language background.

[4] This chapter was published in a similar version as Grohe, A.-K., & Weber, A. (2016a). Learning to comprehend foreign-accented speech by means of production and listening training. *Language Learning, 66*(S2), 187–209. doi: 10.1111/lang.12174

Introduction

With a steadily growing number of second language speakers across the globe (Crystal, 2007), listeners are increasingly often confronted with foreign-accented speech, both when listening in their first language (L1) and when listening in their second language (L2). Understanding foreign-accented speech is often perceived to be more difficult than understanding speech from a native speaker. Indeed, foreign-accented speech typically deviates from the standard pronunciation of a target language, and the deviations can obstruct the complex processes of comprehension.

Luckily, however, recent findings have shown, that at least L1 listeners can rapidly overcome these initial processing difficulties (e.g., Clarke & Garrett, 2004; Maye, Aslin, & Tanenhaus, 2008; Witteman, Weber, & McQueen, 2013). Can L2 listeners just as easily adapt to foreign accents? And does producing an accent facilitate adaptation more than listening to it does? In contrast to L1 speakers, who are unlikely to deviate from the norms of their language spontaneously, L2 speakers regularly deviate and may therefore show a production training advantage. These questions were adressed by comparing the effect of production training with that of listening training on the subsequent comprehension of foreign-accented speech, both for L2 and L1 participants.

Adaptation to foreign-accented speech

A number of recent studies have investigated how listeners handle variation in foreign-accented speech, and in particular how experience with accents affects comprehension ease. L1 listeners have been shown to adapt rapidly and long-lastingly to foreign-accented speech. Clarke and Garret (2004), for example, found in a cross-modal word verification task that L1 English listeners can overcome initial processing difficulties for sentences spoken by Spanish- and Chinese-accented talkers within one minute of exposure to these talkers. Witteman, Bardhan, Weber, and McQueen (2015) have found that L1 listeners of Dutch adapt quickly, automatically, and reliably to both globally accented words (Dutch words that did not contain any full segmental substitutions but were spoken by a Hebrew speaker) and those with a specific segmental accent marker (the Hebrew speaker shortened words with /i/ to /ɪ/, e.g., Dutch statief */stati:f/ 'tripod'* was pronounced as **/statɪf/*).

Rapid adaptation to specific accent markers was also investigated. In Maye and colleagues (2008), all English front vowels were lowered so that, for example, */witʃ/ 'witch'* was pronounced as **/wɛtʃ/*. Native English participants were familiarized with that accent in a listening task, and their recognition of accented words was tested afterwards in a lexical decision task. Endorsement rates, i.e., the proportion of accented words that were accepted as words, were significantly higher after familiarization with the accent than before familiarization. This pattern generalized to new words with front vowel lowering that were not included in the familiarization phase, and a weaker effect was also observed for new words with a lowered back vowel (e.g., */lʊk/ 'look'* pronounced as **/lok/*).

Also Witteman, Weber, & McQueen (2013) observed accent adaptation in L1 after just brief accent exposure. In their cross-modal priming study, exposure to a four-minute story in German-accented Dutch was sufficient to overcome initial processing difficulties with a strong German accent marker (e.g., /œy/ in *huis 'house'* pronounced as /ɔɪ/) for Dutch participants; only after exposure did strongly-accented words elicit facilitatory priming. Witteman and colleagues (2015) furthermore showed that such rapid adaptation can occur with various exposure tasks and the learning effects are stable for up to one week.

No study has so far investigated rapid adaptation to foreign-accented speech for L2 listeners, but a number of studies have shown that long-term experience with their own accent makes it easier for L2 listeners to understand their accent in a laboratory setting. Bent and Bradlow (2003), for instance, demonstrated in an off-line transcription task that a foreign accent does not necessarily make a speaker less intelligible. In their study, English sentences were recorded by native talkers of English, Chinese, and Korean, and native listeners of each of these languages listened to and transcribed the sentences. Analyses of correctly transcribed keywords showed that the intelligibility of a high-proficiency speaker from the same L1 background as the L2 listener could be equal to that of an L1 speaker. This effect was called the *interlanguage speech intelligibility benefit*.

Further support for this facilitatory effect of long-term experience comes from priming and eye-tracking studies. For example, Weber, Broersma, and Aoyagi (2011) found in an English cross-modal priming study facilitatory priming for Dutch and Japanese L2 listeners when the auditory primes were pronounced as is typical for the listeners' own accent, but not when they were pronounced with an accent that was different from their own accent.

Furthermore, Hanulíková and Weber (2012) found in an eye-tracking study that L2-listeners' ease of recognizing English words with various *th*-substitutions reflected the participants' preference for substitutes in their own accent. Native German and Dutch participants saw an English target word with initial /θ/ (e.g., *theft*) on the screen together with a competitor word (e.g., *left*) and two unrelated distractors. Simultaneously, the target word was being presented auditorily, but the /θ/ was substituted with /t/, /s/, or /f/ (e.g., *theft* was pronounced as **teft*, **seft*, or **feft*). All three substitutions occur in Dutch- and German-accented English, with /f/ being perceptually the closest match, /s/ the most frequently chosen substitute by German speakers, and /t/ the preferred substitute of Dutch speakers. Germans showed looking preferences for /s/ variants, whereas native Dutch participants preferred /t/, which corresponded to the production substitution preferences of each group respectively.

Interweaving accent production and accent comprehension

Obviously, L2 learners not only have experience with their L2 accent from listening to other L2 learners with the same native language background, but also from producing the accent themselves. What is then the role of own, individual accent production in the learning process? Does producing an accent help subsequent comprehension more than listening exposure does? Seminal theoretical accounts in this domain are, for instance, the motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985), the perceptual loop theory (Levelt, 1989; Levelt & Roelofs, A., Meyer, A. S., 1999), the embodiment account (for an overview: Glenberg, 2010), and most recently the integrated theory of language production and comprehension (Pickering & Garrod, 2013).

In their theoretical account Pickering and Garrod anticipate close connections and interweaving between production and comprehension, claiming that language production involves comprehension and comprehension involves production processes. During comprehension, listeners make predictions about the talker's utterance based on imitation that promotes the comprehension process. These predictions rely on the *association route*, which is based on experience with comprehending others' utterances, and/or the *simulation route*, which draws on the comprehender's individual production experience. Generally, if the comprehender is similar to the speaker, they rely on the *simulation route*; smaller similarity directs their trail towards the *association route*. Similarity can be derived from, for example, native language background, dialect, or cultural aspects.

Research on the connection between production and comprehension has provided somewhat mixed results (e.g., Cutler, 1995). Correlations between production and perception, i.e., a link between participants' acoustic realization of sound contrasts and their discrimination abilities, have been found in a number of studies both with L1 and L2 participants. Evidence for a tight production-perception link on the phoneme level has been observed for voice onset times (VOTs) of bilabilal stops (Beach, Burnham, & Kitamura, 2001). Bilingual Greek-English or Australian-English speakers, without any knowledge of Thai, who had extreme VOT productions of /ba/ and /pa/, also had native-like phoneme discrimination between Thai /ba/, /p^{h}a/, /pa/, and /p^{h}a/. Similar correlations had previously been found for vowels in L1 American English (Bell-Berti, Raphael, Pisoni, & Sawusch, 1979). Correlations between production and perception of non-native sound contrasts were also found in a number of L2-studies (Flege, 1993; Flege, Bohn, & Jang, 1997; Flege, MacKay, & Meador, 1999; Flege & Schmidt, 1995; Schmidt & Flege, 1995; but see e.g., Hattori & Iverson, 2010; Peperkamp & Bouchon, 2011).

Evans and Iverson (2007) found production-perception correlations in L1 for English university students from the North of England studying in the South of England. Over a period of two years, their vowel productions were recorded, and the students periodically completed two perceptual tasks. First, they were asked to choose the best exemplar locations for vowels embedded in either northern or southern English accented carrier sentences. Second, they identified single words in noise spoken with either a northern or a southern English accent. After attending university for two years, the students' accent changed to those of educational norms, and they chose similar vowels to the ones they produced. Students with a more southern English accent showed a better performance in identification of southern English speech in noise. Furthermore, studies with L1 listeners have shown that self-productions can affect recall as well as the comprehension of others' speech. For example, producing a word aloud can be beneficial for later memory recall compared to reading a word silently (e.g., MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010).

A number of recent studies have investigated the effects of accent imitation on learning with somewhat mixed results. In Cho and Feldman (2013), a visual cue on the screen told L1 English participants whether they just had to listen to or repeat presented English words. English words were either produced with a Dutch accent or they were produced in unaccented, native English, and participants either had to imitate the Dutch accent or repeat the words in their natural, native English accent. Subsequently, a free recall and a word recognition task tested the word memory of each group respectively. More items were remembered after repetition than after listening exposure and a memory advantage was found for Dutch-accented words in the listening condition, but not in the repeat conditions. There was no advantage of accent imitation compared to natural accent repetition.

Adank, Hagoort, and Bekkering (2010), however, did find that accent imitation improves listeners' subsequent accent comprehension. L1 Dutch participants first accomplished a pre-training accent comprehension test in Dutch (sentence transcription), were then trained on the accent by listening, repeating, transcribing or imitating it, before being re-tested on their transcription accuracy. Between pre-training and testing, transcription performance improved most strongly after imitation. There was no significant difference in improvement between the other conditions.

Imitation, however, did not help native accent learning as reflected in a study on French closed words (Nguyen, Dufour, & Brunellière, 2012). In closed words, the northern French /o/ is pronounced as /ɔ/ in southern French, e.g., /ʀoz/ *'rose'* is pronounced as /ʀɔz/. First, northern French participants imitated single closed words read by a southern French speaker, or they listened to these words, or they had no exposure at all. Lexical decisions were then made on closed and open words (words that have an /ɔ/ in both northern and southern French, e.g., *robe 'dress'*). Reaction times were not faster for closed words compared with open words when imitation preceded testing. However, reaction times to accented words were compared with reaction times to canonical words that included the same phoneme as the trained accented words. This implies a comparison of accented words presented in a short imitation exposure phase with non-accented words that participants have had regular production experience with in everyday life. The lack of training effects may have resulted from the fact that long-term L1 production experience effects are stronger than those of short-term exposure in the experimental setting.

Baese-Berk and Samuel (2016) found that imitating a new L2 sound can even inhibit learning. In their study, native speakers of Spanish were trained on the Basque three-way distinction of fricatives (lamino-dental, apicoalveolar, and post-alveolar) using a continuum in an ABX discrimination task. Participants exclusively listened to the sounds (perception only group) on the continuum, and others listened to the sounds and then imitated them (perception+production group). Then, both groups decided with a button press whether the last sound (X) was identical to sound A or B, followed by a feedback on the correctness of their decision. The training was accompanied by a perception evaluation before and after training in which sounds on the critical continuum were played auditorily and had to be discriminated (without feedback). A comparison of the pre- and post-test revealed sound learning for the perception group, but not for the perception+production group.

With the addition of feedback on the productions during training, however, Kartushina and colleagues (Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015) did find in a similar phoneme discrimination study that production training can be helpful for learning to discriminate L2 sounds. They had native French participants undergo a production training with Danish vowels accompanied by a production and perception evaluation before and after training. Perception was evaluated by an ABX vowel discrimination task, and production evaluation was done with a task similar to the training. Both phases involved the auditory presentation of the respective vowel that had to be repeated by the participants. During training, but not during the production evaluation, feedback was provided on the vowel productions. Improvements in perception of 4.56 % as well as in production of 17 % were observed.

Thus, while memory studies unequivocally suggest that production can facilitate word recall (e.g., MacLeod et al., 2010) –– independently of a word's accent (Cho & Feldman, 2013) –– most comprehension studies suggest that production training is sensitive to accent. Active word recall is a conscious, effortful process. Word and sound recognition, as measured by transcription and sound discrimination accuracy, on the other hand, can be argued to test faster lexical or sublexical access and are situated closer to the *unconscious* than the *conscious* pole on a thought-up consciousness continuum. Does this mean that there are two different processes, one for conscious retrieval ignoring concrete acoustic details and one for fast and automatic retrieval that benefits from acoustic details?

Considering the effects of production training on subsequent comprehension, the imitation studies discussed above are constrained by the fact that not just the imitation itself (i.e., production) but also the foregoing presentation of the to-be-imitated speech (i.e., listening) potentially affected the learning process. It is therefore still an open question what the effects of production alone are on subsequent comprehension, in particular for L2 participants. Producing speech that deviates from the norms of a target language is a regular experience for L2 speakers. This potentially assigns an important role of production to the adaptation of L2 accents for L2 learners. The goal of the present chapter is to (1) investigate rapid learning of foreign-accented speech for L2 learners, and (2) determine the role of listening versus production training for learning, both for L2 and L1 participants.

Present Study

Using a training-test paradigm, the present chapter investigates learning of foreign-accented speech through listening and production training, for both L2 (Experiments 1 and 2) and L1 (Experiment 3) participants. Participants were always first trained on an accent and then completed a comprehension task (i.e., a lexical decision task). In contrast to Experiment 1, participants in Experiments 2 and 3 were instructed to accept words in the newly learned accent as existing words in the comprehension task. Their performance was compared to a control group that had not undergone training. The target language was English, and the L2 speakers in Experiments 1 and 2 were German learners of English. Rather than imitating an accent, participants in the production training group produced the accent deliberately without an auditory prompt. The deliberate accent markers were t-substitutes for the English voiceless dental fricative /θ/.

Substitutions with /t/ are relatively uncommon in German-accented English (Hanulíková & Weber, 2012; Lombardi, 2003) and therefore potentially offer the opportunity to observe learning during the experiment. Furthermore, the alveolar stop /t/ is perceptually quite distinct from the dental fricatives (Cutler, Weber, Smits, & Cooper, 2004) and should be easily noticed as deviating from canonical /θ/. The stop is also part of the German (Kohler, 1999) and American English (Ladefoged, 1999) sound inventory and should not pose particular difficulties in production, neither for German learners of English nor for L1 speakers of English.

In order to have consistent accent productions, both L1 English and German L2 participants were instructed to substitute /θ/ for /t/ during production training. The listening training group was exposed to the same accent marker auditorily with pre-recorded speech from a German learner of English. The test phase included words from the training phase (old words) and words that the participants encountered during the test phase for the first time (new words). If the comprehension of new words shows effects of training, this would imply that the learning process generalizes across the lexicon (e.g., Witteman et al., 2015, Maye et al., 2008).

Experiment 1

Participants

Seventy-two female native German students from the University of Tübingen (19–29 years old, mean age = 23, SD = 2.5) participated for monetary reimbursement. Only women were tested in order to account for the fact that the recordings were exclusively made by female talkers. None of them suffered from any hearing disorders, and they all had intact or corrected vision. Participants were highly proficient in their second language English as was attested both by the *LexTALE* test, a visual lexical decision test (Lemhöfer & Broersma, 2012), and their overall good performance in the experiment. *LexTALE* performance was on average 71.2 points out of 100 (equally distributed across participant groups). On average, they had started learning English in German High School (Gymnasium) at the age of ten, minimally for eight years. None had stayed longer than eight weeks in an English-speaking environment or been regularly exposed to native English speakers at the time of the experiment.

Material

Training*.* An excerpt of the fairy tale *King Thrushbeard* (565 words) served as training material (see Appendix A, Table A1). Each *th* in the story was a digraph and corresponded in 39 cases to the English interdental fricative /θ/ (20 word-initially, 4 word-medially, 15 word-finally) and in 58 cases to /ð/ (53 word-initially, 5 word-medially). Sentences in the story were generally short with a simple structure and contained many high frequency words. For production training, participants read the story out loud from the screen with *th*s being highlighted, and deliberately substituted each *th* with /t/. For listening training, the story was pre-recorded by two female German speakers of English who also read the story from the screen with *th*s being highlighted, and deliberately substituted each *th* with /t/. During training, both voiced and voiceless sounds were replaced, which made the task consistent and easier for participants. This strategy also made it less obvious that only instances of the voiceless sound would be tested later.

Test. Twenty-four English words with a voiceless interdental fricative in initial position were chosen as critical words (e.g., *theft;* see Appendix A, Table A2; mean frequency of 163.6 per million, CELEX; Baayen, Piepenbrock, & Gulikers, 1995). When /θ/ was replaced with /t/ (**teft*), these words resulted in nonwords in both English and German[5]. Twelve of the critical words were taken from the training story (old words), and twelve were new. Old and new critical words were matched for frequency and number of syllables. An additional 120 fillers were selected, half of them words (mean frequency of 479.6) and half nonwords. Twenty-four of the fillers (12 words and 12 nonwords) contained a /t/ (two thirds of which word-initially). Half of the word fillers were taken from the training story and half were new.

Recordings. Training and test materials were recorded with an Olympus LS-11 sound recorder (44.1 kHz, 16 bit) by two German female speakers who were highly proficient in English but had a noticeable German accent (speaker A: 24 years old; speaker B: 21 years old). Two speakers were recorded so that different voices could be used for training and test in the listening group, which accounts for a comparable difference in voices in the production group. The speakers did not differ significantly in F0-range or speaking rate, and the authors judged their degree of accent to be comparable. Both speakers were instructed to pronounce each *th* as a /t/, but to speak otherwise as naturally as possible. The recordings were made sentence by sentence until each sentence was read fluently and all critical substitutions were made. The final

[5] Note that one item was potentially a word in German (*tief 'deep'*). This item did not affect the pattern of results as a separate analysis excluding the item shows (see Appendix A, Table A4.1thief for Experiment 1 and A4.2thief for Experiment 2).

recording of speaker A was in total three minutes long, and that of speaker B was three and a half minutes long.

Design

Twenty-four participants listened to the recordings of the training story (listening group), 24 read the story out loud (production group), and 24 had no training (control group). Half of the participants heard speaker A during test and half heard speaker B. For the listening group, the speaker of the training story was always different from the speaker during test. Four experimental lists (including all 144 items; 24 critical words and 120 fillers) were created for the lexical-decision task with varying, pseudo-randomized item order. Each critical word was preceded and followed by a filler. The lists were distributed equally across participants. The experiment was programmed with the software *Presentation* (version 17.0, www.neurobs.com).

Procedure

The experiment took place in a quiet room at the university. Each participant was seated in front of a computer screen and wore over-ear headphones (Sennheiser HD 215 II). All instructions were given orally in German by a female native German speaker before the experiment started. Instructions were also provided in written English. Participants read them aloud, which tested their spontaneous /θ/ productions and allowed the identification of participants who spontaneously produce many t-substitutes (none were found).

The listening group listened twice to the pre-recorded story with t-substitutes while seeing the text on the screen; the production group read the story twice out loud with the t-substitutes. The production training was designed as a reading aloud task, rather than a picture naming task for example, for two reasons. The number of English words that fitted the experiment's criteria was relatively small, and not all of the potential words were depictable (e.g., *thesis, threshold, thankful*). Furthermore, by having the orthographic form present during training, it was easy to "remind" participants to substitute each *th*. The listening group was specifically asked to pay attention to the pronunciation of the talker and to report particularities afterwards. This ensured that they were attentive and conscious of the substitutions just as the production group was naturally. After the first reading, a short feedback was provided for the production group. Overall reading performance was quite good, with a mean error number during the first reading of 1.8 (4.5 % of all voiceless *th* occurrences, max. 12.8 %) and 0.9 (2.3 %, max. 12.8 %) during the second reading. Altogether, participants with training were exposed to the voiceless *th*-substitution 78 times.

After training, English instructions for the lexical decision task were shown on the screen. Participants were instructed to decide as quickly and reliably as possible whether a presented auditory stimulus was an existing English word or not. Each stimulus was preceded for 500 ms by a fixation cross on the screen. In Experiment 1, there was no specific instruction on the required action for critical words with t-substitutes. Note that word forms with t-substitutes were not real words of English (e.g., **teft* for *theft*), but participants could consider them words after accent training. No feedback was given during lexical decision. After the lexical decision task was completed, participants filled in a language background questionnaire and the session was concluded with the English language proficiency test *LexTALE*.

Analysis and results

Statistical analyses were conducted with the statistical software R, version 3.2.2 (R development core team, 2017). Endorsement rates and reaction times for accepted critical words were analyzed with linear mixed effects regression models (e.g., Baayen, Davidson, & Bates, 2008) using the lme4 (Bates, Maechler, Bolker, & Walker, 2016) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015) packages. Endorsement rates were analyzed with generalized models (by adding *family=binominal*), which account for the binary nature of the dependent variable. Reaction times, measured from word offset, were log-normalized. In mixed-effects regression modeling there are fixed-effect factors and random effects. Random intercepts and slopes allow for a maximal random effects structure (e.g., Barr, Levy, Scheepers, & Tily, 2013). They consider differences between participants and items in size of the effect of the fixed variables (Cunnings, 2012). Random slopes reduce the Type I error rate and make the analysis conservative.

The present data include *participant* and *item* as random effects, and fixed factors are *training* (listening versus production versus control), *familiarity* (old versus new), *item duration* (in ms), *lexical frequency* (logged frequency per million occurrences), *LexTale score*, and *speaker* (speaker A versus speaker B). For each analysis an individual, best fitting model was built, that included a particular choice of fixed and random factors, applying a backward, stepwise selection procedure. The final, most parsimonious model of best fit only includes (marginally) significant predictors and interactions. In the following, the random effect structure of the models is only specified if it deviates from having simply *participant* and *item* as the random intercept.

Endorsement rates and reaction times of all accepted critical items (i.e., yes responses to **teft*) were analyzed. Reaction times faster than 100 and slower than 2,000 ms were excluded (4.1 % of the data). Endorsement rates were on average 53.7 %. Most items were accepted by the listening group (61.7 %), with an intermediate acceptance rate by the control group (53.5 %). The production group accepted fewest items (46.2 %). Mean endorsement rates of individual items ranged between 32.9 % (*third*) and 80 % (*thief*) (see Appendix A, Table A3.1).

The final analysis model for endorsement rates had *lexical frequency* and the *LexTale score* as fixed factors (*training* and *familiarity* were excluded because they were not significant: all t-values < |1.5|, p-values > .1). A higher LexTale score resulted in significantly lower endorsement rates ($ß_{LexTale}$ = -0.08, SE = 0.02, z = -4.03, p < .001), and items with higher lexical frequency were accepted less often ($ß_{Frequency}$ = 0.14, SE = 0.07, z = 1.93, p = .05).

The model for reaction time analysis of accepted words had as fixed factors *item duration* as well as an interaction of *training* with *item familiarity*. *Item* and *participant* were cross-random factors, and by-*participant* random slopes for *familiarity* were included. Figure 2-1 and Table A4.1 (Appendix A) show that there were two significant interactions between *training* and *familiarity*; first between the control and the production group ($ß_{training*familiarity}$ = 0.20, SE = 0.08, t = 2.31, p = .02) and second between the control and the listening group ($ß_{training*familiarity}$ = 0.20, SE = 0.81, t = 2.52, p = .01).

These interactions are attributed to group differences that are present for old, but not new items. Both the production group ($ß_{training}$ = -0.32, SE = 0.10, t = -3.27, p = .002) and the listening group ($ß_{training}$ = -0.27, SE = 0.09, t = -2.83, p = .006) were significantly faster than the control group for old words. This was not true for new words (all t-values < |1.2|, p-values > .2). There was no significant difference between the production and listening group, neither for old, nor new words (both t-values < |0.6|, p-values = .6)[6]. Furthermore, reaction times became faster with increasing item length (ß = -0.001, SE = 0.0003, t = -4.39, p < .001).

[6] Descriptively, reaction times were faster for the production group than the listening group, while endorsement rates were higher for the listening group than the production group. To exclude the possibility of a speed-accuracy tradeoff, the statistical model was re-run and the mean endorsement rate by participant was added as fixed factor. This analysis showed, that contrary to a speed-accuracy tradeoff, reaction times became in fact significantly faster with increasing endorsement rates ($ß_{endorsement:rates}$ = -0.32, SE = 0.13, t = -2.34, p < .02).

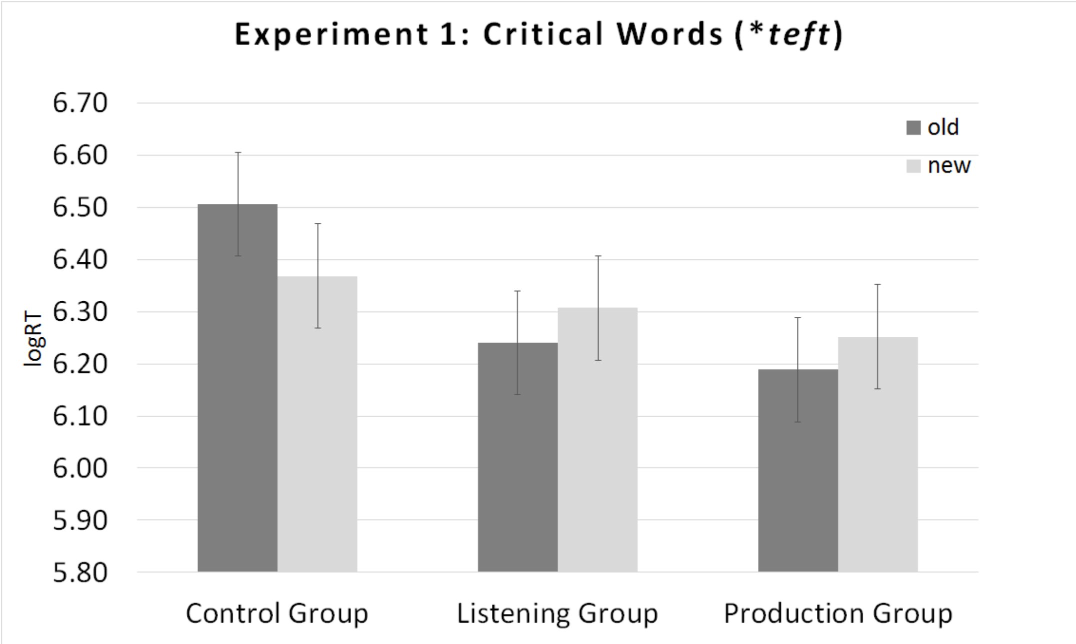

Figure 2-1. *Reaction times of accepted critical words by training group and familiarity in Experiment 1. Whiskers represent standard errors.*

Interim discussion

Overall, endorsement rates for critical words (e.g., **teft*) were not particularly high and relatively equally distributed across training groups and item familiarity. Reaction times for old critical words indicated facilitation triggered by both listening and production training. Although the effect for old critical words in the production group was numerically larger than that for old words in the listening group, this difference was not statistically significant. An analysis of rejected items may be able to show a potential difference between the training groups, however any likely effect was hidden by the small amount of data.

The generally low acceptance rate (53.7 % across groups) was probably due to participants' insecurity. A number of participants reported after the experiment that they recognized many old words but did not know whether they were supposed to accept them (**teft* is still a nonword outside of the experimental situation after all). With more security about the appropriate reaction, the difference between production and listening training might still emerge. In Experiment 2 the participants' insecurity was reduced by explicitly telling them that they can accept words with t-substitutes in the lexical decision task. The resulting increase in available data (i.e., yes responses) makes reaction time analyses more reliable, but renders endorsement rates less informative by making the choice to accept critical words less spontaneous. In Experiments 2 and 3, a focus will therefore be placed on reaction time analyses which are particularly suitable to provide information about online processing.

Experiment 2

Experiment 2 was identical to Experiment 1, with the only difference being that participants in Experiment 2 received explicit instructions to accept critical words (e.g., **teft*) during lexical decision. The control group was given one example of the accent and instructed to treat items with that accent as words. This was expected to increase participants' confidence regarding the lexical decision, which, in turn, increased the amount of correctly accepted critical words and gave reaction time analyses more statistical power.

Participants

Seventy-three female native Germans (18–31 years old, mean age = 23.3, SD = 3.0), all students at the University of Tübingen, participated. None of them suffered from any hearing disorders, and they all had intact or corrected vision. L2 English proficiency was again attested by the *LexTALE* test (mean = 70.3 out of 100 points, equally distributed across participant groups) as well as participants' overall good performance in the experiment (filler word error rates ranged between 0 and 8.3 % out of 60 words). On average, they had started learning English in German High School (Gymnasium) at the age of ten, minimally for eight years. None had stayed longer than eight weeks in an English-speaking environment or been regularly exposed to native English speakers at the time of the experiment. The analysis of voiceless *th*-substitution errors resulted in an overall reading performance of the production group similar to that of Experiment 1. The mean number of errors came to 1.4 (3.5 %, max. 18 %) during the first reading and 1.2 (3.1 %; max 12.8 %) during the second reading.

Results

Responses to accepted critical words between 100 and 2,500 ms were analyzed (3.7 % data excluded). Endorsement rates showed that 79 % of all critical words were accepted (listening group: 79.8 %; production group: 73 %; control group: 84 %; see Appendix A, Table A3.2). The mean endorsement rate of individual items ranged between 52.1 % (*third*) and 93.4 % (*Thursday*). The listening group accepted significantly more old than new items ($ß_{familiarity}$ = 1.48, SE = 0.44, z = 3.35, $p < .001$), and the production group tended to accept fewer old items than the listening group ($ß_{training}$ = -1.03, SE = 0.61, z = -1.67, $p < .1$). There was no other statistically significant difference regarding endorsement rates (all t-values < 1.4, p-values > .16).

Reaction times were analyzed with a model that included an interaction of *training* with *familiarity* and the predictors *speaker* and *item duration*. Figure 2-2 and Table A4.2 (Appendix A) show that there were two significant interactions between *training* and *familiarity*. First, between the control and the listening group ($ß_{training*familiarity}$ = 0.27, SE = 0.06, t = 4.19, $p < .001$), and second, between the control and the production group ($ß_{training*familiarity}$ = 0.21, SE = 0.07, t = 3.24, $p < .001$), which may be attributed to differences between old and new items as well as group differences. Old words were significantly faster than new words for both the listening group ($ß_{familiarity}$ = -0.24, SE = 0.1, t = -2.52, p < .02) and the production group ($ß_{familiarity}$ = -0.19, SE = 0.1, t = -1.94, p = .06), but not for the control group (t < |0.3|, p > .8).

Across groups, old items in the production group were significantly faster than old items in the control group ($ß_{training}$ = -0.18, SE = 0.08, t = -2.28, p < .03), while the listening group was only numerically faster than the control group (t = -1.33, p = .19). Reaction times for new items were significantly slower in the listening group than in the control group with no training ($ß_{training}$ = 0.17, SE = 0.08, t = 2.13, p = .04). Between the production group and the control group, there was no significant difference for new items (t = 0.3, p = 0.75). New items tended to be faster in the production group than in the listening group ($ß_{training}$ = -0.13, SE = 0.08, t = -1.64, p = .1). Reaction times were significantly faster for longer items ($ß_{wavdur}$ = -0.0014, SE = 0.68, t = -4.76, p < .001), and speaker B tended to be faster than speaker A ($ß_{speaker}$ = -0.1, SE = 0.06, t = -1.75, p < .09)[7].

[7] As in Experiment 1, a speed-accuracy tradeoff could be excluded by adding mean endorsement rate by participant as fixed factor to the final model. No significant effect of endorsement rate (t < 0.4, p > .7) was found.

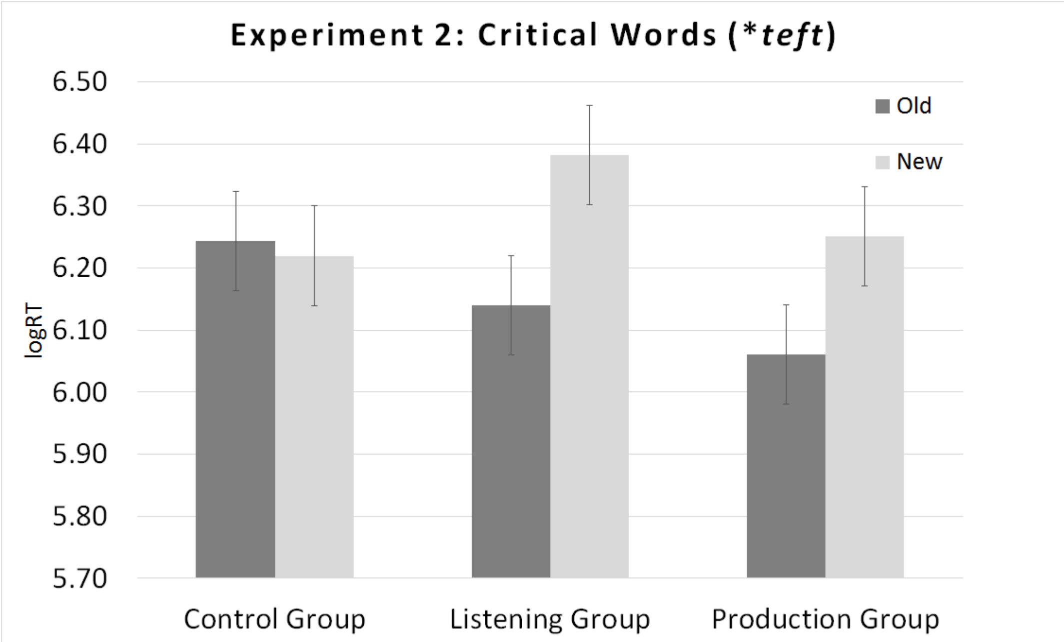

Figure 2-2. *Reaction times of accepted critical words by training group and familiarity in Experiment 2. Whiskers represent standard errors.*

Interim discussion

In comparison to Experiment 1, endorsement rates increased noticeably in Experiment 2 (from 53.7 % to 76.7 % on average). In Experiment 2, reaction times to old critical words were faster after production training than with no training. The listening group was also numerically faster than the control group, however this effect was not fully significant. Thus, production training facilitated subsequent comprehension of accented speech more strongly than listening training did. The direction of this effect was already present in Experiment 1, but the difference between production and listening training is now more pronounced with more data. For L2 participants, producing a sound substitution facilitates subsequent comprehension of familiar words containing this substitution more so than listening to the substitution does.

In principle, the advantage of production training may be a general one and also apply to L1 speakers. The act of production alone may modify mental representations more effectively than listening does, regardless of whether the speaker is native or non-native. The production advantage may, however, also be restricted to L2 speakers. L2 speakers are, after all, far more experienced with producing a variable speech signal (Wade, Jongman, & Sereno, 2007) than L1 speakers are. Greater experience with production variation may increase the effectiveness of production training in the present laboratory setting because of more flexible learning faculties. In Experiment 3, therefore L1 English participants were tested to investigate the role of the native language background on the effects of production and listening training.

Experiment 3

Methods

Sixty-seven female students with L1 American English were tested. The majority of them were students at the University of Maryland, some were tested in Montreal, and a few were students of the University of Tübingen. Six participants were excluded because they failed to follow the instructions of the experiment and one further participant because English was not her L1. The remaining sixty participants were between 18 and 34 (mean = 22.2, SD = 3.6) years old.

Twenty participants were assigned to the listening condition, twenty to the production condition, and twenty to the control group. The material, design, and procedure were identical to those in Experiment 2, including the explicit instructions to accept critical words. Hit rates of word fillers revealed errors rates between 0 and 15 % (0–9 errors among 60 word fillers). Overall, *th*-substitution performance during production training was high. The mean number of errors came to 0.9 (2.3 % of all voiceless *th*-occurrences; max. 10.3 %) during the first reading and 0.6 (1.5 %; max.

7.7 %) in the second reading. As participants in Experiment 3 were native speakers, no test of English proficiency (*LexTale*) was required.

Results

Accepted critical words with reaction times between 100 and 2,000 ms were analyzed (4.1 % data excluded). 75.2 % of all critical words were accepted (listening group: 74.5 %; production group: 74.8 %; control group: 76.4 %; see Appendix A, Table A3.3). Endorsement rates of individual words ranged between 48.3 % (*thriller*) and 98 % (*Thanksgiving*). More old than new words were accepted, significantly so for the listening group ($ß_{familiarity}$ = 1.35, SE = 0.53, z = 2.56, p = .01).

Reaction times were analyzed with a model having an interaction of *familiarity* with *item duration*, and as further predictors *training* and *speaker* in the fixed section. *Item* and *participant* were cross-random factors, and *speaker* was added as random slope by item. Reaction times tended to be faster for speaker B than speaker A ($ß_{speaker}$ = -0.14, SE = 0.07, t = -1.84, p = .07). Old words became faster with increasing item length ($ß_{wavdur}$ = -0.0016, SE = 0.0005, t = -3.22, p = .004). The same trend was observed for new items, however this was not significant (t = -1.22, p > .2). Both training conditions tended to be faster than the control group, but not significantly so (all t-values < |1.3|, p-values > .2).

Analysis of fillers. In addition, word and nonword fillers were analyzed to check for specific group differences that may also have affected the answers to critical words. Word fillers with reaction times between 100 and 1,500 ms were analyzed (5 % excluded). Endorsement rates reached 94.8 % and were lower than in Experiment 2 (see Appendix A, Table A3.4). Nonword fillers with reaction times between 100 and 1,800 ms were analyzed (5 % excluded). Rejection rates of all nonwords came to 85 %. They were higher when there was a training phase (Listening: 87.8 %, Production:

89.1 %) than without training (78 %). This difference was significant, as shown by a model with *training* as fixed factor. Both the production ($ß_{Training}$ = 1.04, SE = 0.3, z = 3.47, p < .001) and the listening group ($ß_{Training}$ = 0.92, SE = 0.3, z = 3.09, p = .002) were more accurate in rejecting nonwords.

L1-L2 comparison — critical words. In order to compare directly between L2 participants in Experiment 2 and L1 participants in Experiment 3, a combined analysis was run. Reaction times were analyzed with a model including a three-way interaction of *training*, *familiarity*, and *native language background*, as well as *item duration* as fixed factors (see Figure 2-3 and Appendix A, Table A4.3). L1 and L2 participants did not differ significantly between training groups for neither old nor new items (all t-values < |1.4|, p-values > .17). Two three-way interactions of *training* with *familiarity* and *native language background* (control group versus listening group: $ß_{Training*Familiarity*Native}$ = 0.25, SE = 0.10, t = 2.6, p = .009; control group versus production group: $ß_{Training*Familiarity*Native}$ = 0.26, SE = 0.1, t = 2.9, p = .004) support the finding that the interactions for the L2 participants did not apply for the L1 participants.

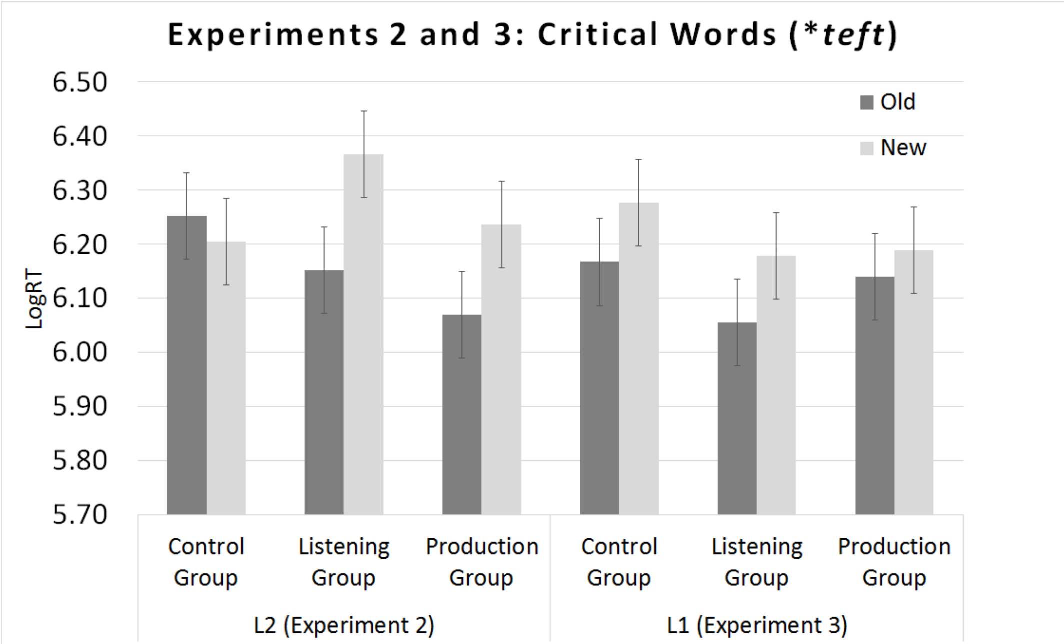

Figure 2-3. *Reaction times from Experiment 2 and 3 in one analysis. Whiskers represent standard errors.*

Interim discussion

In Experiment 3, more old than new critical words were accepted by the listening group, and both training groups were much better rejecting nonword fillers than the control group was. The production group did not experience increased training effects compared to listeners. Both training groups tended to accept critical words more quickly than the control group. The production and listening groups accepted significantly fewer nonword fillers. It is not clear whether this tendency is reflected in reaction times because there are only few data points for wrongly accepted nonwords which does not provide sufficient data for representative reaction time analyses.

Increased rejection rate of both training groups does, however, hint at less specific and more general learning effects for the participants with training. The direct comparison between L2 and L1 participants shows that within those groups, old words were accepted by the listening group more often than new words, as should be expected. Thus, joint reaction time analysis supports the results reported in the separate analyses.

Discussion

The present chapter investigated if L2 learners can learn as quickly L2 speech as L1 speakers can, and if producing an accent facilitates learning more than listening to it does. In three experiments with German L2 and L1 English participants a training-test paradigm was used in which participants first produced or listened to an English short story in which all instances of voiceless dental fricatives were replaced with /t/ (e.g., **teft* for *theft*). Then, participants accomplished an auditory lexical decision task on words with the same accent. Experiment 1 tested German learners of English and did not include any specific instruction on the required action for critical words during lexical decision. In Experiment 2, German learners of English were specifically instructed to accept critical words as real words, and Experiment 3 tested L1 American English participants with the same instructions.

In Experiment 1, without explicit instructions regarding critical words, participants were not highly confident in their decisions, resulting in relatively low and variable endorsement rates of critical words during test. Reaction times nevertheless showed significantly faster responses after listening and production training than after no training, with a descriptive advantage of production training over listening training. The explicit instruction to accept critical words in Experiment 2 increased endorsement rates considerably, and the reaction time advantage for the production training became more pronounced. This pattern, however, did not hold for L1 participants who showed generally quite weak training effects. Thus L2, but not L1, participants learned more easily a new, salient foreign accent with production than with listening training. This group difference excludes a general comprehension advantage triggered by production as such and assigns an important role to the native language background of the participants. The difference may go back to variability in effectiveness of training and/or more speaker-specific learning in L1 than in L2.

The amount of variability distinguishes L2 productions from L1 productions in general, which may affect training effects. Individual talkers' productions are known to be more inconsistent and variable in L2 than in L1 speech. For example, dental fricatives in a speaker's L2 English, depending on their L1 background, have a preferred substitute (e.g., /s/ for L1 German), but this is discontinued by other substitutes (e.g., /f/) or the canonical pronunciation (Hanulíková & Weber, 2012). Wade et al. (2007) found that L2 speakers of English with L1 Spanish produced more variable English vowel realizations than L1 English speakers did. Assuming that the properties of mental representations depend on regular input from listening and speaking, highly variable L2 speech can possibly cause processing to be more flexible for L2 speakers. Previous production variability thereby makes L2 speakers more sensitive to training effects from production, but not listening training.

At first glance, this contradicts Baese-Berk and Samuel (2016) who found that sound production inhibits learning of new L2 sound categories, while listening training helps. There is, however, a crucial difference between the design of the Baese-Berk and Samuel study and ours. While the new accent in the present study (i.e., pronouncing /θ/ as /t/) was easy to produce, participants in Baese-Berk and Samuel had to imitate sounds from a continuum ranging from /s̪a/–/ʃa/. It is feasible that the imitations of the artificial steps of the continuum were not perfect. The acoustic discrepancy between the presented sound prompt and the participants' own productions may have inhibited learning effects. Baese-Berk and Samuel argue against this option because in their Experiment 3, the production group had to produce unrelated sounds rather than imitating the new L2 sound and learning was still disrupted. Disruption may, however, result from the event of two incompatible tokens first being listened to and then produced (which involves self-listening). The discrimination study by Kartushina, Hervais-Adelman, Frauenfelder, and Golestani (2015) supports such an interpretation. In their study, production accuracy was increased by concrete feedback on productions, which in turn resulted in better sound discrimination performance after production than after listening training.

Also attentional differences between training groups could have had an influence in the present study. Possibly, producing an accent draws attention more intensely on mispronunciations than listening does. A greater learning effect for the production group would then be due to increased attention. Note that to enhance participants' awareness of the accent, participants in the listening groups were specifically instructed to focus on the talker's speech properties and to report particularities afterwards. While the possibility cannot be excluded that attentional differences affected the results, it needs to be pointed out that it would be difficult to explain the lack of a difference between training groups for L1 participants based on attention alone.

Generally speaking, training effects only emerged for items that were also present during training in the present study. This lack of generalization is in line with an episodic account of the mental lexicon, i.e., the storage of exemplars (e.g., Goldinger, 1998). It is also possible, however, that generalization simply takes more time and greater variability of individual speaker-related factors for L2 learners to emerge. In support hereof, Bradlow and Bent (2008) found that accent training (Chinese-accented English sentences) with multiple speakers results in significantly better performance in a subsequent sentence transcription task by L1 English listeners than the training of the same accent pronounced by a single speaker different from the test speaker. If the training and test sentences were pronounced by the same speaker, listeners' performance reached a similar level to that after multiple-speaker accent training. The role of production variability is underlined by Nguyen, Dufour, and

Brunellière (2012). In their study, not even imitation training facilitated L1 participants' recognition of words with a trained accent in a way that made it more easily accessible than canonically pronounced words that included the same sound as the accent substitute.

I now turn to possible reasons for why production training was more helpful for L2 participants than for L1 participants. One factor may have been a closer match between own productions and speech input during test for the L2 participants. Participants in the production training always learned from their own productions and then had to apply this knowledge to an L2 talker during test. For L1 participants, but not L2 participants, the change of speaker also included a change in nativeness of the speaker. Quite likely, the realizations of critical words were overall more similar across different L2 speakers than across L1 and L2 speakers. Thus acoustic differences between the sounds encountered during training and the ones during the test could have influenced the results.

Similarities between talker and listener are also emphasized by Pickering and Garrod (2013), who suggest that imitation supports the comprehension process by creating predictions about what an interlocutor is going to say. The *association route* and the *simulation route* modulate this process. Similarity between the speaker and comprehender determines which route is chosen. In the test phases of the present experiments, the native language background establishes high similarity between the recorded speakers and comprehenders (in all conditions) in Experiment 2 because both are native German speakers with L2 English.

The recorded speakers and comprehenders in Experiment 3, however, are less similar to one another because L1 participants listened to an L2 speaker during the lexical decision task. This predicts the selection of the *simulation route* by L2, and the *association route* by L1 participants. As the *simulation route* relies on the comprehender's individual production experience, facilitation by production (compared to listening training) in L2 is in line with this assumption. The *association route* relies on experience with comprehending the utterances of others, resulting in the prediction of facilitation by listening and not production exposure for L1 participants. The L1 listening group was not, however, faster than the L1 producers, indicating that there must be other factors that may have prevented the listening training from being fully effective.

In contrast to Maye, Aslin, and Tanenhaus (2008), L1 participants in the present study did not show strong learning effects during listening training either. One reason may be the explicit instruction to accept all critical words, which could have blurred possible spontaneous differences between training conditions. Secondly, Maye and colleagues presented more instances of accented words during the training phase than the present experiments did. They included 24 different training items half of which occurred 7–10 times (high frequency items) and half occurred 1–2 times (low frequent items), resulting in an estimated occurrence of 92–144 times (mean 118). The present story (presented twice) included 78 accented (critical) words which, based on the calculated mean, is about 50 % less than in Maye and colleagues. Significant differences may require greater training input.

Thirdly, Maye and colleagues had their participants perform the same lexical decision task once before and a second time after accent training, taking the first attempt as reference point, whereas the present experiments included an inter-subject design. Finally, in the listening conditions, different voices were presented during training and testing, while Maye and coworkers used the same synthesized voice in all experimental stages. Prior research has indeed found that adaptation to foreign-accented speech is more difficult across voices than within voices (Bradlow & Bent, 2008; McLennan & González, 2012; Weil, 2001), indicating that the voice difference may have reduced learning effects in the present study.

L1 participants in the present study did not experience facilitation by production, whereas imitators did in imitation studies reported earlier (Adank et al., 2010; Kartushina et al., 2015), suggesting facilitatory, and not inhibitory effects as argued by Baese-Berk and Samuel (2016), via the double input it includes. Whereas production alone only involves self-listening, imitation adds listening to a prompt. The present study also did not reflect a general advantage by producing a word over simply listening to it in terms of MacLeod's proceduralist account (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010), according to which producing a word aloud adds an additional source of discrimination to its memory record, facilitating later recall. The observed effects were limited to critical words in the L2 group and did not apply to word fillers or L1 participants.

What MacLeod and colleagues call the *production effect*, seems specific to active and conscious recall tasks and seemingly does not apply to online word recognition or for accented or canonically pronounced words. This is an argument for two different processes of lexical access depending on a listener's concrete task during the test phase: one for active and conscious word recall, probably relying on an additional source of discrimination (MacLeod et al., 2010), and one for fast and automatic word recognition. The active recall process is very general and does not refer to episodic detail (Cho & Feldman, 2013), whereas the faster word recognition process benefits from more detailed information that may have been obtained during production training. Learning from production may be episodic or abstract in nature, assuming that abstraction requires more intense training with greater talker variability than in the present study.

Conclusion

In conclusion, production and comprehension processes are not independent from one another, but, under certain circumstances, learning in one modality can generalize to the second one. Learning properties based on both modalities as well as the properties of the actual comprehension process strongly depend on the talker's native language background. The production advantage was found in L2, but not in L1, because learning from production training is possibly modulated by the participant's regular production variability.

Higher variability in L2 productions results in more flexible representations prone to modifications by brief production training. It probably also depends on the faculty of generalizing learning to new talkers with properties different from the training talker that was found for L2, but not L1 speech. Whether own experience is accessed during the comprehension test may rely on talker-comprehender similarity, which is higher if both have the same native language background. This hypothesis is further tested in Chapter 2B.

Chapter 2B

Global accent similarity in accent learning[8]

Abstract

Is it easier to learn an accent when the speaker has an accent similar to oneself than when not? This question was addressed in a training-test study. During training, L1 English participants either listened to a story read by an L1 speaker who replaced all /θ/s with /t/, or they produced the story with the /θ/-substitutions themselves. Learning, as demonstrated by faster lexical decision times to accented words, was observed after both production and listening training. Speaker-listener similarity was ensured by having L1 speakers during training and test. Without this similarity, no learning had been observed in Chapter 2A.

[8] This chapter was published in a similar version as Grohe, A.-K., Downing, M., & Weber, A. (2016). Similarity in global accent promotes generalized learning of accent markers. In Australasian Speech Science and Technology Association (Ed.), *16th Speech Science and Technology Conference.*

Introduction

Native as well as foreign accents are pervasive in everyday speech. While both accent types deviate from the standard pronunciation of a language, they are produced by speakers with different native language backgrounds — foreign accents are produced by second language (L2) speakers; native accents are produced by native (L1) speakers of the target language. Both accent types can initially slow down comprehension, but L1 listeners are able to adapt to them.

In Clarke and Garret (2004), L1 English listeners overcame initial processing difficulties for sentences spoken by Spanish- and Chinese-accented speakers within one minute of listening to these speakers. In Witteman, Weber, and McQueen (2013), listening to a four-minute story in German-accented Dutch was sufficient for Dutch L1 participants to subsequently show facilitatory priming for words with a strong German accent marker (/œy/ as in *huis 'house'* pronounced as /ɔɪ/); see Witteman, Bardhan, Weber, & McQueen, 2015 for comparable results). Native accent learning has been investigated for example in a study by Evans and Iverson (2007). Students originally from northern England adapted to Standard Southern British English within two years of their university studies in southern England. Native accent learning is also possible with a training phase that only lasts several minutes and can even affect cohort activation of unaccented words, as found in a recent eye-tracking study by Trude and Brown-Schmidt (2012). Learning was, however, speaker-specific, i.e., it was restricted to the training speaker and did not generalize readily to new speakers.

Thus, learning of foreign- and native-accented speech occurs, but it is not clear yet if the underlying processes are the same. One possibility is that learning follows the exact same principles in both cases and only the accent's acoustic distance from standard speech, i.e., accent strength, is the determining factor for learning ease. Support for the role of accent strength comes, for example, from Witteman, Weber, and McQueen (2013) who found that priming effects were smaller for strongly-accented words than for weakly-accented words. On the other hand, Eisner, Melinger, and Weber (2013) observed that native English listeners generalize learning of a position-specific accent marker (devoicing of final /d/) to a new position only when the accent marker was learned from an L2 speaker, and not when the same accent marker was learned from an L1 English speaker. This suggests more tolerant learning of foreign-accented speech than native-accented speech. Further support for this notion was found by Witteman, Weber, and McQueen (2014). L1 Dutch listeners learned the same German-accented words more quickly when the speaker was perceived to be an L2 speaker of Dutch in the filler items than when they were perceived to be a native speaker of Dutch. Thus, not only accent strength, but possibly also the nativeness of the speaker, can influence the learning process.

In all of the above studies, learning occurred through listening. However, producing an accent can also form the basis for learning. This was tested in Chapter 2A of this dissertation. In a training-test paradigm, the effect of an individual's own accent production was compared to that of listening to someone produce an accent. Participants first either listened to an English short story recorded by a German learner of English who replaced all dental fricatives ('*th*', /θ/) with /t/ (e.g., *theft* was pronounced as **teft*), or they read the same story aloud themselves with the instruction to substitute all *th*s with /t/. Neither German learners of English (Hanulíková & Weber, 2012) nor the tested population of L1 English speakers typically replace /θ/ with /t/, but they have no difficulties producing /t/s when instructed to do so. After the training, participants completed a lexical decision task on words with *th*-substitutions spoken by another German learner of English. Surprisingly, L1 English participants showed no learning, while German L2 participants accepted, both after production training and after listening training, accented words more quickly than a control group.

One possible explanation for the lack of observable learning effects for L1 participants is based on the fact that, in addition to testing generalization of learning across two speakers, L1 participants did not share a language background with the L2 test speaker. Possibly, this combination resulted in acoustic differences between training and test items that were too large for learning to generalize across two speakers. Support for this explanation comes from the eye-tracking study reported in Chapter 3A in which L1 German participants learned an accent marker (devoicing of word-initial voiced stops; e.g., *Balken ‘beam’* was pronounced as **Palken*) both after having listened to an L1 German speaker produce the accent and after having produced the accent themselves.

If this pattern of results transfers to L1 English participants and to a lexical decision task with reaction times as dependent measure, then L1 English participants should learn an accent marker well enough to generalize effects to a new speaker using the materials from Chapter 2A, but only if the materials are produced by L1 English speakers. In summary, prior research has shown that learning foreign and native accents is possible, with both listening and production training. It is not clear, though, what is the role of a speaker’s nativeness in accent learning.

This was investigated in the present study. L1 English participants performed the same production and listening training and subsequent word recognition task as in Experiments 2 and 3 of this dissertation (reported in Chapter 2A). That is, they either listened to an English story in which all /θ/s were pronounced as /t/, or they produced the story with the accent themselves before responding to English words with the accent marker in a lexical decision task. However, both the training story for the listening group and the test items were recorded this time by two female native speakers of American English rather than by German learners of English. Thus, the accent marker was the same as in Chapter 2A, but in both training conditions the language background of the training speaker (L1 participants from the production training and the pre-recorded L1 speakers from the listening training) now matched the language background of the test speaker.

The present study thus tested whether the similarity between one's own accent and the accent to be learned facilitates learning. This would be in line with the assumption that L2 participants in Chapter 2A generalized accent learning across speakers because their own accent was similar enough to the pre-recorded speaker's accent. It is expected that L1 English participants in the present study learn the accent and generalize learning to a new speaker both after listening and after production training because the language background of the speaker during test is the same as that of the participants, thereby increasing the global similarity of the speech stimuli. Such a result would show that in an individual's L1, speaker-listener similarity in terms of global accent can promote speaker-general learning of a specific accent marker.

Experiment 4

Participants

Fifty-nine female students (18–26 years old, mean age = 20.8, SD = 2.0) from the University of Maryland, all American L1 English speakers, were tested. None of them suffered from any hearing disorders, they all had intact or corrected vision, and none had participated in Experiment 3.

Material

The training text was 565 words long and was based on the fairy tale *King Thrushbeard*. Each *th* in the text was a digraph and corresponded to the English interdental fricative /θ/ in 39 words. For the lexical decision test, 24 English words (see Table 2-1) with /θ/ in initial position were chosen as critical words (mean frequency: 163.6 occurrences per million according to the CELEX word form dictionary: Baayen, Piepenbrock, & Gulikers, 1995). When /θ/ was replaced with /t/, the resulting word forms were nonwords (*theft* was pronounced as **teft*). Twelve of the critical words were taken from the training story (old words), and 12 were new. Old and new critical words were matched for frequency and number of syllables. New critical words were included in the study to test if accent learning can generalize across the lexicon or is specific to the trained old words. An additional 120 filler words (24 contained a /t/, none contained a /θ/) were selected, half of them were words and half nonwords.

The selected material was the same as in Chapter 2A, but the stimuli were recorded by L1 English speakers rather than by German learners of English as in Chapter 2A. Training and test materials were recorded with a professional recording device (Focusrite Scarlett 2i2 and a Rhode NT1-Kit; 44.1 kHz, 16 bit) by two L1 American English female speakers (speaker A: 29 years old, from New Jersey, USA; speaker B: 35 years old, from Georgia, USA). Two speakers were recorded so that different voices could be used for training and test in the listening training group, which accounts for a comparable difference in voices in the production training group. The speakers did not differ significantly in F0-range or speaking rate. They were instructed to pronounce all *th*s as a /t/, but, otherwise, to speak as naturally as possible. Every *th*-instance was highlighted in yellow in the text for the recording. The final story of both recorded speakers was about three minutes in length with no significant difference between speaker A and B.

Table 2-1. *Critical old words from the training story and critical new words in Experiment 4.*

old words	new words
thankful	Thanksgiving
theft	thematic
therapist	theory
thing	thesis
think	thickness
thinner	thief
thirsty	thirty
thrifty	threaten
throughout	threshold
throw	thriller
thumb	throat
Thursday	thunder

Design and Procedure

Twenty participants listened to the recordings of the training story (listening group), 20 read the story out loud and deliberately substituted all *th*s with /t/ (production group), and 19 had no training (control group). Half of the participants heard speaker A during test, and half heard speaker B. For the listening group, the speaker of the training story was always different from the speaker during test. Four experimental lists (including all 144 items; 24 critical words and 120 fillers) were created for the lexical decision task with varying, pseudo-randomized item order. Each critical word was preceded and followed by at least one filler. The lists were distributed equally across participants. The experiment was programmed with the software *Presentation* (version 17.0, www.neurobs.com).

The experiment took place in a soundproof room at the University of Maryland. Each participant was seated in front of a computer screen and wore noise-canceling headphones. All instructions were provided in written English. The listening group listened twice to the pre-recorded story with *th*-substitutions while seeing the story on the screen; the production group read the story twice out loud with the *th*-substitutions. The listening group was asked to pay specific attention to the pronunciation of the talker and to report oddities afterwards. This ensured that they were just as attentive and conscious of the substitutions as the production group naturally was. The production group followed the substitution instructions quite consistently (mean error rate of all voiceless *th*-occurrences: 2.9 %).

After the training, English instructions for the lexical decision task were presented on the screen. Participants were told to decide as quickly and reliably as possible whether a presented auditory stimulus was an existing English word or not. Each stimulus was preceded by a fixation cross on the screen for 500 ms. As in Experiments 2 and 3 (Chapter 2A), the participants received explicit instructions to accept critical words (e.g., **teft*) during lexical decision in order to clarify any uncertainty about the decision. The control group was given one example of the accent and also instructed to treat items with that accent as words. Explicit instructions were necessary because word forms with *th*-substitutions were not real words in English (e.g., **teft* for *theft*), but participants might consider them words after accent training. No feedback was given during lexical decision. After the lexical decision task was completed, participants filled out a language background questionnaire.

Results

Analyses were conducted with the software R (R development core team, 2017). Endorsement rates and reaction times for accepted critical words (i.e., yes responses to **teft*) were analyzed with linear mixed effects models (Baayen, Davidson, & Bates, 2008). Reaction times, measured from word offset, between 85–1,900 ms long were included in the analyses (5 % outliers) and were log-normalized. For each analysis, an individual, best fitting model was built that included only significant fixed factors as well as random factors (*participant* and *item* as random intercepts). This was done with a backward stepwise selection procedure starting with the most complex model including all possible main effects and interactions that still converged. Significance of level comparisons was indicated by t-/z-values>|2|. Corresponding p-values of factors and interactions, as reported in the text below, were determined with likelihood ratio tests using the anova-function.

Endorsement Rates. Endorsement rates were on average 86 % (listening group 86.1 %; production group 91.6 %; control group 81.5 %, see Appendix A, Table A3.5). Mean endorsement rates of individual items ranged between 63.6 % (*thematic*) and 100 % (*thankful, Thanksgiving, thirsty, threshold*). Statistical analyses show that the effect of training group was significant (χ^2 = 9.3, $p < .01$). The production group accepted significantly more accented tokens than the control group (ß = 1.1, SE = 0.34, z = 3.3). The explicit instruction to accept all accented tokens as words, however, renders endorsement rates less informative by making the choice to accept critical words less spontaneously. As in Chapter 2A, focus will therefore be placed on reaction time analyses, which are particularly suitable to provide information about online processing.

Reaction Times. Reaction times for accepted critical tokens (i.e., yes responses to **teft*-items) were analyzed with a model having as fixed factors an interaction between *training* (with the levels production, listening, and no training) and *familiarity* (with the levels old words and new words), as well as *item duration* (in ms) and *list position*. *Item* and *participant* were cross-random factors, and by-*participant* random slopes for *list position* as well as by-item random slopes for *list position* and *speaker* (with the levels speaker A and speaker B) were included (see Appendix A, Table A4.4). Figure 2-4 shows that the *training*familiarity* interaction was significant (χ^2 = 10.9, $p < .005$). This interaction stems from differences between both training groups and the control group (a main effect of training) for old items. These differences were less strongly pronounced for new items. The listening group accepted old items faster than the control group (β = -0.31, SE = 0.09, t = -3.4), and there was a strong trend to accept new items more quickly in the listening group than in the control group as well (β = -0.16, SE = 0.09, t = -1.8). The production group tended to be faster than the control group in accepting old items (β = -0.16, SE = 0.09, t = -1.7), but there was no effect for new items (t < 1).

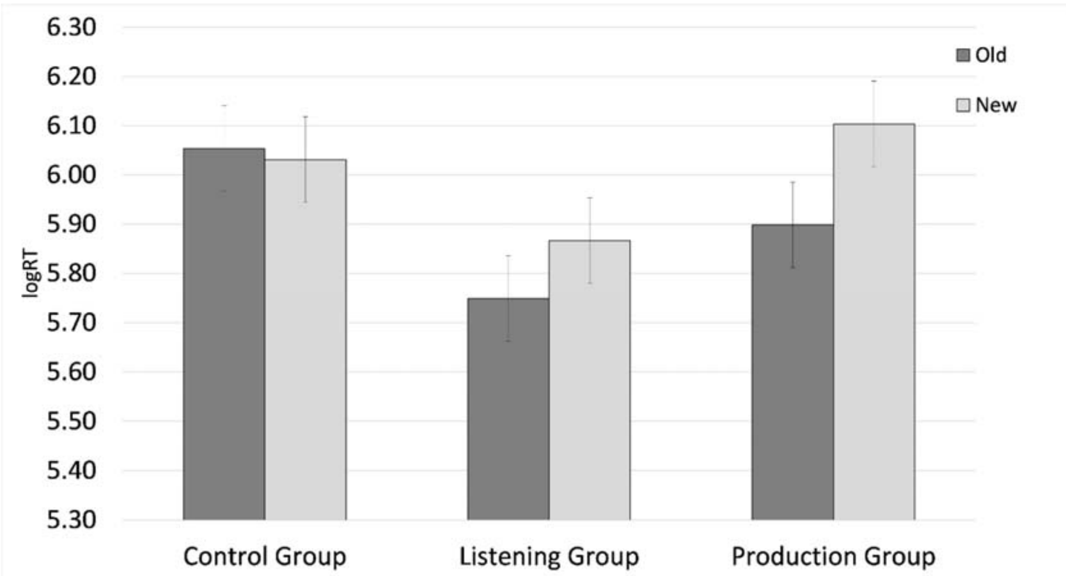

Figure 2-4. *Original reaction times to old and new critical items in the listening group, the production group, and the control group without training in Experiment 4. Whiskers represent standard errors.*

In order to further ensure that the observed effects indeed reflect a manipulation of the training conditions rather than more random differences in general processing speed between participant groups, reaction times to correctly accepted word fillers (60–1,500 ms) were analyzed as a baseline comparison. Word fillers were canonical word forms without specific accent markers, and reaction times of the listening group were on average 14 % faster than those of the control group, while the production group was 2.4 % slower than the control group (see Figure 2-5). Statistical analyses (see Appendix A, Table A4.5) showed a significant main effect of *training group* ($\chi^2 = 9.25$, $p < .01$). To correct for this bias, reaction times for accepted critical tokens in the training groups were adjusted to the processing speed of the control group (i.e., reaction times of the listening group were increased by 14 % and those of the production group reduced by 2.4 %).

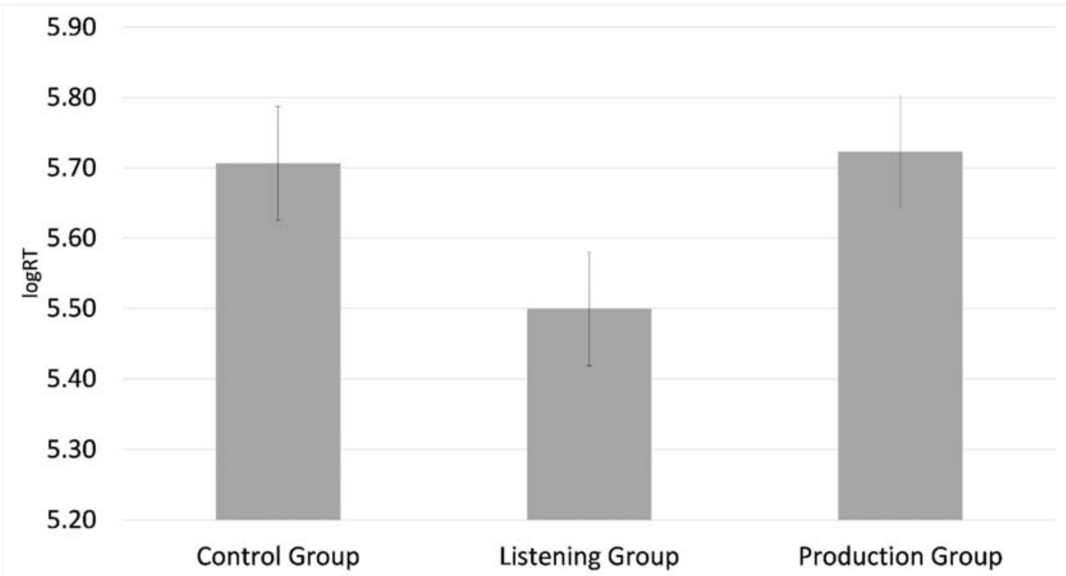

Figure 2-5. *Original reaction times to filler items in the listening group, the production group, and the control group without training in Experiment 4. Whiskers represent standard errors.*

With the new, adjusted reaction times (see Appendix A, Table A4.6), the *training*familiarity* interaction was still significant ($\chi^2 = 10.9$, $p < .005$). Training effects were observed for old, but not new items (see Figure 2-6). The listening group accepted old items faster than the control group ($\beta = -0.17$, $SE = 0.09$, $t = -1.9$), and the production group was significantly faster than the control group ($\beta = -0.18$, $SE = 0.09$, $t = -2$). Reaction times for new items were not affected by accent training (all t-values < |0.5|).

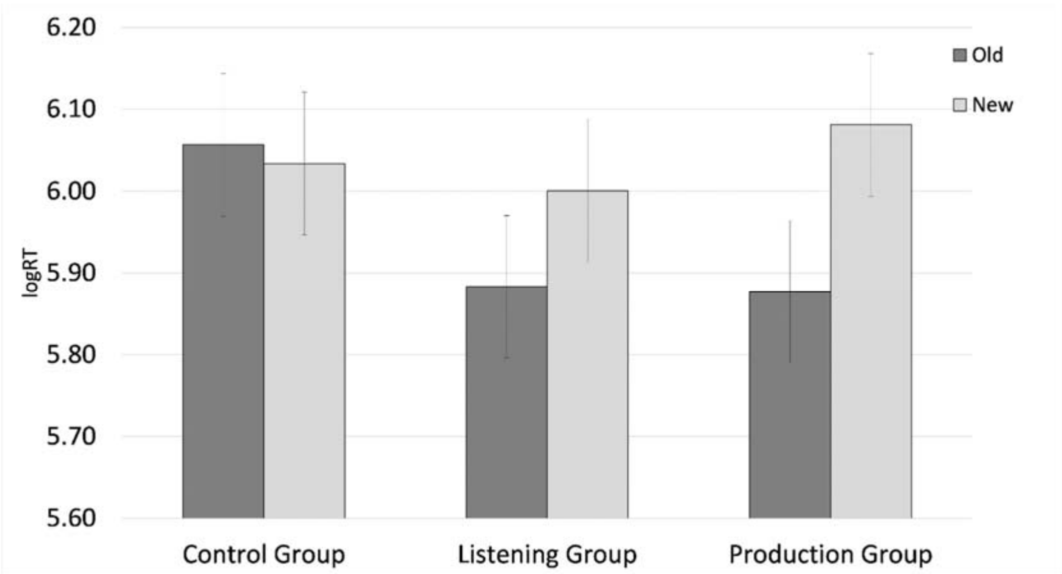

Figure 2-6. *Adjusted reaction times to old and new items in the listening group, the production group, and the control group without training in Experiment 4. Whiskers represent standard errors.*

Next, the present data were analyzed together with the L1 data from Experiment 3 (Chapter 2A). The same material and design was used in the latter study, but the pre-recorded tokens were spoken by two female L2 learners of English. However, no training effects were observed with L1 participants in Experiment 3. The new, large dataset consequently included the additional variable *speaker-nativeness*

(L1 versus L2 English). The same model as above was run, replacing the two-way interaction with the three-way interaction of *training*familiarity*speaker-nativeness* (see Appendix A, Table A4.7). Adding *speaker-nativeness* significantly improved the model ($\chi^2 = 16.5$, $p < .02$). This confirms that speaker nativeness was the critical factor that provoked the training effects.

Discussion

In the present study, L1 English participants learned an accent marker in their L1 well enough to generalize it to a new speaker through brief listening and production training. The listening training and test material was recorded by L1 English speakers. The same material, though recorded with a second language speaker, had previously not induced speaker general learning in either training condition (reported in Chapter 2A). Joint analyses of the present data and the analogous L1 experiment in Chapter 2A (Experiment 3) confirmed that accent learning depends in this case on the pre-recorded speaker's native language background. This underlines the importance of similar accent properties between test speakers and participants.

Accent similarity was created by having a test speaker with the same native language background (American English) as the participants. L1 participants learned the accent with the L1 test speaker. In addition to greater speaker-listener similarity, L1 speakers also had a smaller degree of overall accent strength. In contrast to the L2 speakers in Chapter 2A, they did not have a global L2 accent. Thus, not only speaker-listener similarity but also generally weaker accentedness could have driven the present findings. The role of similarity, however, is strengthened by the fact that L2 participants in Chapter 2A did generalize learning of the accent when it was produced by L2 speakers with a global L2 accent.

The present findings, moreover, emphasize the role of global accent markers. Even if a global accent does not inhibit processing as much as a specific accent marker might (Witteman, Bardhan, Weber, & McQueen, 2015: Experiment 1), it can still play an important role for generalization of accent learning across speakers. Adding a global accent to a specific accent marker can increase or reduce speaker-listener similarity and thereby affect generalized accent learning. The reason why accent similarity between speaker and listener is so important in accent learning likely lies in participants' prior experience with the accent in question. A language user is more experienced in both producing and listening to (also by self-listening) their own accent than other accents, which facilitates accent learning.

In contrast to many prior studies on accent learning, the present study tested accent learning across speakers, meaning, whether an accent can be learned from one speaker and be applied to a second speaker with the same native language background. It was found that generalization across speakers is possible as long as both speakers have similar accent properties. Learning across speakers was also found with L2 participants in Chapter 2A, suggesting that accent learning is not necessarily speaker-specific as was suggested in Trude and Brown-Schmidt (2012). This is in line with further studies (Bradlow & Bent, 2008; McLennan & González, 2012; Weil, 2001) that suggest that accent learning across speakers is more difficult than within one speaker, but is possible.

Possibly for this reason, accent learning in the present study only occurred for items that were included in the training phase (old tokens) and did not generalize to new words. In line with abstractionist accounts of the mental lexicon (e.g., McClelland & Elman, 1986; Norris, 1994), it is suggested that the amount of training was not enough for full abstraction. Intensifying the training would probably have evoked training effects also for new tokens. Note however, that in prior research (Witteman, Bardhan, Weber, & McQueen, 2013) single Dutch words with both a global and a specific Hebrew accent induced priming effects after only 3.5 minutes of phoneme monitoring training. However, unlike the present study and Chapter 2A, Witteman and colleagues presented the same speaker during training and test.

Interestingly, learning effects did not differ between the production and the listening groups; L1 English participants learned an accent in their L1 equally well with both production and listening training. This finding is in line with the results presented in Chapter 3A. The eye-tracking study with single accented German words revealed similar proportions of looks to the target by L1 German participants after production and after listening training. Why then was there a production advantage for L2 participants in Chapter 2A? This is probably because accent strength still plays a role in accent learning; accent strength modulates accent learning together with speaker-listener similarity.

The role of accent strength in accent learning with listening training is supported by Witteman, Weber, and McQueen (2013). Mild accents are more easily learned than stronger accents. Accent strength is further emphasized in the *accent processing classification* that was postulated by Clarke and Garrett (2004). In this account, the accent's acoustic distance from native speech is the only decisive factor in accent adaptation. Foreign and native accents follow the same principles, but the strength of an accent determines the ease of accent learning. When learning through listening, stronger accents need more time or more intense training. In Chapter 2A, the speakers were L2 speakers, which involves a stronger accent than in the present study, where L1 speakers were recorded. The L2 speakers featured both global accent markers and the specific, manipulated accent marker, whereas the L1 speakers in the present study only produced the specific accent marker, implying a smaller distance from canonical pronunciation.

Conclusion

Can differences in accent learning from L1 and from L2 speakers be explained by accent strength differences alone? The present results suggest that to probably not be true: Accent similarity between speaker and listener facilitates accent learning. Typically, an L1 user's accent is more similar to a second L1 user's accent than it is to that of an L2 user. This assigns an important role to the speaker's native language background. Still, the L1–L2 speaker comparison has shown that accent strength *per se* co-determines accent learning.

Chapter 3A

The role of salience in native accent learning[9]

Abstract

In two eye-tracking experiments, the effects of salience in accent training and speech accentedness on spoken-word recognition were investigated. Salience was expected to increase a stimulus' prominence and therefore promote learning. A training-test paradigm was used on native German participants utilizing an artificial German accent. Salience was elicited by two different criteria: production and listening training as a subjective criterion and accented (Experiment 5) and canonical test words (Experiment 6) as an objective criterion. During training in Experiment 5, participants either read single German words out loud and deliberately devoiced initial voiced stop consonants (e.g., *Balken 'beam'* pronounced as **Palken*), or they listened to pre-recorded words with the same accent. In a subsequent eye-tracking experiment, looks to auditorily presented target words with the accent were analyzed. Participants from both training conditions fixated accented target words more often than a control group without training. Training was identical in Experiment 6, but during test, canonical German words that overlapped in onset with the accented words from training were presented as target words (e.g., *Palme 'palm tree'* overlapped in onset with the training word **Palken*) rather than accented words. This time, no training effect was observed; recognition of canonical word forms was not affected by having learned the accent. Therefore, accent learning was only visible when the accented test tokens in Experiment 5, which were not included in the test of Experiment 6, possessed sufficient salience based on the objective criterion *accent*. These effects were not modified by the subjective criterion of salience from the training modality.

[9] This chapter was published in a similar version as Grohe, A.-K., & Weber, A. (2016b). The *Penefit* of salience: salient accented, but not unaccented words reveal accent adaptation effects. *Frontiers in Psychology, 7*(864). doi:10.3389/fpsyg.2016.00864

Introduction

Languages typically consist of a number of regional dialects — that is, native accents. In the southwestern German state of Baden-Württemberg, for example, one does not have to travel very far to encounter various native accents, as Spiekermann (2008) has documented. This variation can pose a problem for non-locals. When non-locals hear a native accent for the first time, they often do not understand what is being said as easily as do locals, who are experienced with the regional varieties. Recent studies have indeed shown that listeners process accents in their native language more easily when they are familiar with the accents than when they are unfamiliar with them (e.g., Adank, Evans, Stuart-Smith, & Scott, 2009). Adaptation by non-locals to native accents is, however, possible. Adaptation has been found for longer periods of exposure to a novel accent (Evans & Iverson, 2007), but it can even be observed after only four minutes of listening to a new accent (Trude & Brown-Schmidt, 2012).

This is also true for second language (L2) learners. Chapter 2A of this dissertation suggests that producing a new accent for only seven minutes can facilitate subsequent accent understanding for L2 learners, even more so than listening to the accent does. The act of speech production arguably makes an accent more salient than listening to that accent does. Can the advantage of production experience also be observed in a listener's native language (L1)? Next to signal modality (production and listening), salience can also emerge from concrete properties of the speech signal itself. Acoustic distinctiveness of a speech signal can enhance its salience (e.g., Cho & Feldman, 2013). The present study used a training-test paradigm in German in which salience was induced by either a subjective or an objective criterion and looked at the role of salience in native accent learning. The subjective criterion was implemented through two different accent trainings (production and listening) and the objective criterion through the featuring of either accented (Experiment 5) or canonical (Experiment 6) target words during test in an eye-tracking study.

Learning native accents

Familiarity with a native accent facilitates accent processing. For example, listeners with extensive experience with the New York City (NYC) English accent show greater priming effects for words with the NYC-English-typical final r-dropping than listeners with limited experience (Sumner & Samuel, 2009). Similarly, Adank and colleagues (2009) found that only listeners who were familiar with both Standard Southern British English (SSBE) and Glaswegian English (GE) showed equal performance on both accent types in a sentence verification task. The familiarity advantage probably results from adaptation processes, as demonstrated by Evans and Iverson (2007). In their study, students who were originally from northern England adapted to SSBE over the course of their university studies in southern England, as shown through production and comprehension tasks.

Processing advantages for participants' own accents over an unfamiliar accent were also found for French listeners (Floccia, Goslin, Girard, & Konopczynski, 2006). In a lexical decision task, reaction times to items in long sentences were faster when sentences were presented in the participants' own accent (Northeastern French) than when they were presented in the unfamiliar southern French accent. Furthermore, participants did not adapt to the unfamiliar accent during the course of the experiment (see also Floccia, Butler, Goslin, & Ellis, 2009). Additionally, Adank and McQueen (2007) found no short-term adaption in a study with regionally-accented Dutch. In their study, Dutch participants who were not familiar with the Flemish accent had to make animacy decisions on single words spoken by two different speakers, one with a Flemish accent and one with the same accent as the participants. Then, participants were exposed to another speaker with the Flemish accent before having to repeat the animacy decision task. Decision times in the second animacy task were not faster than in the original task.

Short-term adaptation was, however, found by Trude and Brown-Schmidt (2012). Participants were first trained on a native English accent and then tested in an eye-tracking paradigm. During training, participants listened to scripted dialogues with an accented male speaker and an unaccented female speaker. The male speaker raised the /æ/ before /g/ to /eI/, i.e., *bag* was pronounced /beIg/. During test, target words were either spoken by the male or the female speaker. When *back,* a word unaffected by the accent, was the target and *bag* the competitor, *bag* was ruled out more quickly as a candidate word for trials with the male speaker than it was with the female speaker. When *bake*, a word with /eI/ in its canonical form, was the target and *bag* acted as competitor, *bake* was fixated less often when it was spoken by the male. This effect, however, was less strongly pronounced, i.e., competitor inclusion was more difficult than competitor exclusion.

Specific properties of the tested accents could account for the missing effects of adaptation in the studies discussed above, but, more importantly, speaker-specificity can explain it, too. In contrast to Adank and McQueen (2007), who had different speakers with the same accent during training and test and did not find accent learning, Trude and Brown-Schmidt (2012) used the same accented speaker in both of their two experimental phases. Short-term learning of native-accented speech may therefore be rather speaker-specific. This problem has also been addressed in studies on foreign accent learning with mixed results. Using a training test paradigm, Bradlow and Bent (2008) found that generalization of accent learning (Chinese-accented English) to new voices is only possible if the listener is exposed to multiple speakers during training (for similar findings see also Sidaras, Alexander, & Nygaard, 2009). Kraljic and Samuel (2007), on the other hand, found with L1 listeners that the nature of the tested material has an effect on whether perceptual learning can generalize to new speakers. They found generalization effects for plosives but not for fricatives.

Learning with production

Speaker-specificity raises the question of whether there is a way to make training more effective, i.e., allowing for generalization across speakers, and potentially rendering competitor inclusion more robust. This might be possible through production training. In Experiment 2 (Chapter 2A), the production of a foreign accent in participants' L2 promoted learning of that accent in a subsequent lexical decision task. Participants first either listened to an English short story that featured the replacement of all dental fricatives /θ/ with /t/ (e.g., *theft* pronounced as **teft)*, or they read the same story aloud with the same substitutions. A control group had no accent training. Afterwards, all participants completed a lexical decision task on words with the *th*-substitutions.

The production group accepted the accented words significantly more quickly than the control group did. The listening group produced only a weak training effect. When the same experiment was run with L1 participants (Experiment 3), no training effect was observed. Referring to speaker effects, L1 participants in the production group produced the critical accent marker, but they were listening to an L2 speaker in the test phase. According to Pickering and Garrod (2013), listeners are more likely to refer to their own previous production experience if it is highly similar to the speaker they are listening to (e.g., in terms of sex, L1 background, dialect). Less similarity leads listeners to draw on their experience with others' speech. Since only the L2 participants in Chapter 2A had the same L1 background as the recorded test speakers, speaker-listener similarity was smaller for L1 participants than for L2 participants. The findings from Experiment 4 (Chapter 2B) support these assumptions. When the test tokens were recorded with an L1 speaker, a training effect was also observed for the L1 participants.

Facilitatory effects of producing an accent were also found in an accent imitation study with L1 speakers of Dutch (Adank, Hagoort, & Bekkering, 2010). Baese-Berk and Samuel (2016), however, found that imitating a newly learned L2 sound can even inhibit learning. In their study, participants had to imitate sounds from a sound continuum ranging from /s̪a/–/ʃa/, which is arguably difficult for speakers to imitate correctly. A potential acoustic discrepancy between the sound prompt that was presented and the participants’ productions may therefore have inhibited learning effects. A discrimination study with Danish vowels (Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015) supports this interpretation. In that study, production accuracy was increased by concrete feedback on productions, which in turn resulted in better sound discrimination performance after production than after listening training.

Salience in learning

This section focuses the concept of salience, which can potentially explain both the results on accent learning and the advantage of produced compared to listened-to tokens. Salience has been generally defined as “the property of a linguistic item or feature that makes it in some way perceptually and cognitively prominent” (Kerswill & Williams 2002, p. 81). An important question, however, is what exactly makes a linguistic item salient. First, sociolinguistic research on salience suggests that an accent can increase a word’s salience. As suggested by Trudgill (1986), the phonetic difference between two variant forms affects their salience; the greater the difference, the more a dialect speaker is aware of it. Phonetic distance can also be considered within the framework of distinctiveness which assumes that isolated, i.e., distinct, words are more salient than others during encoding — provoking additional processing and, therefore, better memory (for a review: McDaniel & Geraci, 2006).

Geraci and Manzano (2010), for example, had participants study a list of semantically related words that also included a few semantically unrelated, i.e., distinct, words. In ensuing tests, more unrelated than related words were recalled.

Accordingly, Siegel (2010) claims that salience requires a listener to notice a contrast between two linguistic tokens. In terms of phonetic variability, words that carry an accent are distinct from their unaccented counterparts and bear greater salience. Therefore, they can be learned more easily than unaccented words[10]. This was tested in a different memory study (Cho & Feldman, 2013). L1 English participants listened to either Dutch-accented or native-accented English words during a training phase. In a subsequent word recognition task, there was an advantage for Dutch-accented words.

Second, factors beyond linguistic or structural properties may also affect salience. For the case of dialect accommodation, Kerswill and Williams (2002) suggest intensity of dialect contact as one of several factors. Considering the findings on production effects on accent adaptation, the list of extra-linguistic factors can be extended towards cognitive mechanisms by introducing accent learning modality (production versus listening) as an additional factor. Several studies have found an advantage of production over listening for dialect accommodation; this has been named the *production effect*. It predicts that overt production facilitates word recollection when compared with studying a word silently (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010) and also when compared with listening to others producing a word aloud (MacLeod, 2011). It has been suggested that produced words are more easily recalled because they are more distinctive and therefore more salient. Distinctiveness results from listeners focusing more on their own than on others' productions, which are, in the sense of the embodiment hypothesis (for an overview: Glenberg, 2010), more embodied than others' productions.

Salience, as described above, has been further specified in sociolinguistic research. Referring to Schirmunski (1928) and Trudgill (1986), Auer, Barden, and Grosskopf (1998) differentiate objective and subjective criteria of salience. For

[10] For example, in their account on social weighing in speech perception, Sumner, Kim, King, and McGowan (2013) predict better memory for accented words — but only if the accent is socially prestigious.

example, *articulatory distance* is described as an objective criterion and *perceptual distance* as its subjective counterpart. The two relate in that articulatory distance describes the magnitude which a linguistic token deviates acoustically from the canonical realization, whereas perceptual distance describes which way a listener perceives this distance.

Based on this information, it can be concluded that subjective criteria increase the salience of a stimulus, for example, due to regular practice, and the resulting cognitive pre-activation. Objective criteria refer to properties of a stimulus that itself attracts attention because of its distinct, physical characteristics. Under this view, the production effect relies on the presence of an objective criterion. A self-produced word can be physically more distinct compared to a word read silently or a word that is produced by others because these words were only tested in mixed lists, i.e., one participant had to listen to and silently read and produce words in the same session.

In summary, prior research has shown that native accents are more easily processed if they are familiar to a listener than if they are new. Short-term adaptation to native accents is possible, and production alone can positively affect foreign accent learning, at least in L2 learners. Both robust accent learning and the role of production in accent learning may be related to salience. The role and concrete nature of salience in learning accented versus canonical words through different forms of training, however, is not yet clear.

Present study

The present study takes a closer look at this issue by investigating subjective and objective criteria of salience separately, using *modality* and *accent* as criteria. In a training-test paradigm, German participants first underwent native accent training before adaptation was tested by a printed word eye-tracking task. A subjective criterion was established by having two different types of training (production and listening), while the objective criterion featured accented versus canonical test words. Accented test words had their initial voiced bilabial or velar stop devoiced. In Experiment 5, accented words (**Palken* for *Balken 'beam'*) were presented during training and test. In Experiment 6, the same accented words were presented during training (**Palken* for *Balken*), but target items during test were canonical words that overlapped in onset with the trained accented word (***Pal**me 'palm tree'*). Old word pairs from the training phase as well as new word pairs that had not been included in the training list were tested. This manipulation was included to test generalization of learning, i.e., whether the accent is only learned for trained words or also for new accented words.

A subjective criterion of salience was tested by comparing effects of individuals' accent productions with that of listening during training. In contrast to MacLeod (2011), the current study did not manipulate training modality in mixed lists within participants but rather between participants. This permits the comparison of the magnitude of salience based on a subjective criterion of the production modality with that of the listening modality. Individual participants are exclusively trained with one modality. If producing an item in fact constitutes a subjective criterion for salience compared to listening to an item, production training with that item would result in greater salience than listening training.

An objective criterion of salience, on the other hand, was manipulated by the presence of both accented and canonical test tokens. In Experiment 5, the presentation of accented words assigned salience to the test tokens due to their great degree of inherent distinctiveness. Effects of accent as an objective criterion have been previously shown (Cho & Feldman, 2013), but with a memory experiment in which generalization effects were not examined. In the learning phase, the accented words were embedded in a list of filler target words that featured no particular accent marker. This made the accented words distinct from the fillers. Contrarily, in the present Experiment 6, the canonical test words were expected to be less salient. Experiment 6 tested whether the salience inherent in the learned accent can modify the processing of words that include the accent target sound in their canonical form.

A difference in learning effects can be reflected in the activation differences of canonical target words starting with the manipulated accent's target sound. Learning that *Balken* is pronounced as **Palken* potentially increases lexical competition for the canonical *Palme*, which, in turn, slows down recognition of *Palme*. This is based on Trude and Brown-Schmidt (2012), who found that accent learning can imply the inclusion of new competitors. In the present study, *Balken* could be included as a new competitor for *Palme* after training, resulting in fewer target looks to *Palme*.

The pattern of target and competitor activation is especially important during the segmental overlap of target and competitor words. Referring to the principles of an abstract mental lexicon, it is assumed that accent learning is based on learning pre-lexical rules. When hearing **Palken* in Experiment 5, successful word recognition requires the application of a specific accent rule (/b/ → /p/). If the accent rule is learned robustly during training, it is applied by default as soon as the auditory input potentially matches the accent, i.e., from initial /p/ presentation onward. When, in an eye-tracking experiment, the display includes *PALME* and *BALKEN* and **Palken* is the auditory target, both *PALME* and *BALKEN* should be fixated from word onset until disambiguation (including /pal/). Only after disambiguation should *BALKEN* be fixated more often than *PALME*. If the accent rule is not learned robustly enough, the candidates that require the rule (*BALKEN*) have a weaker activation than those that do not require the rule (*PALME*). Consequently, during the overlapping word portion, *PALME* will still be more strongly activated than *BALKEN*, and *BALKEN* will only be preferred after disambiguation.

Successful recognition of a canonical word (*Palme*), as in Experiment 6, does not require accent rule application. However, successful accent learning should result in increased competitor activation of words with a /b/ in initial position. This increase in competitor activation might adversely affect canonical target activation. The rule should be applied by default as soon as the auditory input potentially matches the accent, i.e., also when *Palme* is presented. Having *PALME* and *BALKEN* on the visual display, both words should be equally fixated during /pal/. Only after disambiguation should *PALME* be preferred. The same predictions as above emerge if the accent rule is not learned strongly enough — the candidates that require the rule (*BALKEN*) are activated less strongly than canonical words (*PALME*). Consequently, *PALME* will be more strongly activated than *BALKEN* even from the beginning of word processing.

The accent in the present study is an artificial accent that centers on one specific phonological accent marker and therefore must be differentiated from a dialect. Participants and pre-recorded speakers are not L2 speakers and all used Standard German pronunciation in the experimental context. *Standard* here means that the pre-recorded speakers did not have a noticeable dialect that could allocate them to a specific region in Germany, and the participants' speech did not include specific local (e.g., Swabian) accent properties during the experiment. The tested accent affected German stop consonants and has, to my knowledge, not been documented as an existing native accent of German. It refers to the lenis/fortis-contrast in German bilabial and velar plosives (/b, p/ and /g, k/). In Standard German, fortis plosives are always aspirated in word initial position, while lenis plosives are never aspirated (Jessen, 1998; Kleiner & Knöbl, 2015). The present accent neutralized this contrast, i.e., lenis velar and bilabial stops were aspirated (/g/ pronounced as [k^h] and /b/ pronounced as [p^h]: *Gitter* pronounced **Kitter*, and *Balken* pronounced **Palken*). The accented sound was always aspirated, similar to the canonically fortis stops. For simplification, aspirated, fortis plosives (***P**alme*) are referred to as *voiceless* and lenis plosives with the additional aspiration in the accented version as *devoiced* (*****P**alken*).

An accent with a target sound that is included in the German sound inventory (Kohler, 1999) was opted for. This makes it easy to produce for native German participants and promises relatively stable acoustic properties of the target sounds across participants. The accent under investigation has to be differentiated from middle-Bavarian dialects where bilabial, alveolar, or velar plosives are not realized with an aspiration contrast before /r, l, n, m/; they are always voiceless and unaspirated and therefore neutralize with their lenis counterpart, e.g., *Preiselbeeren 'cranberries'* pronounced as *Breiselbeeren* (Moosmüller & Ringen, 2004). Likewise, in Austrian German, the fortis plosives /p/ and /t/ are not aspirated (Siebs, Boor, Moser, & Winkler, 1969), e.g., *Pinsel 'brush'* pronounced as **Binsel*. In contrast, the accent presented in this study neutralized all bilabial plosives to [p^h] and all velar plosives to [k^h]. Since the accent tested in the present study describes a voicing shift in the opposing direction of existing native German accents, it can be assumed that none of the present participants had had experience with the accent. This ensured the observation of only laboratory-specific training effects.

The prediction is that accent training will result in accent learning effects. The training modality can determine the amount of salience based on one subjective criterion. This would be in line with prior findings where producing rather than listening to a word resulted in better memory (MacLeod, 2011). Salience that relies on an objective criterion of the target token is expected to affect looking patterns such that the learned accent affects processing of highly salient, accented devoiced tokens more than that of canonical voiceless tokens.

Experiment 5

Experiment 5 tested if salience can result from training as subjective criterion. Critical test words had a native accent and were assumed to be highly salient based on the objective criterion *accent*. During training, native German participants either read aloud or listened to single German words that had their initial /b, g/ devoiced to /p, k/, e.g., *Balken* pronounced as **Palken*, while the control group had no training.

In the test phase of the experiment, participants accomplished the printed word variant (McQueen & Viebahn, 2007; Weber, Melinger, & Lara Tapia, 2007) of the eye-tracking task (Allopenna, Magnuson, & Tanenhaus, 1998). Participants saw four printed words in their canonical spelling (including a target, a competitor, and two distractors) on a computer screen and were auditorily instructed to click on a target word while their eye movements were recorded. They listened to devoiced words (**Palken*) and had to click on a visual display that included the target word (*BALKEN*) and a competitor (*PALME*). The competitor allows the investigation of whether activation of the devoiced token can be as strong as activation of voiceless word forms without an accent. The proportion of target fixations was measured and compared between the production, listening, and control groups.

Participants

Seventy-four native German speaking female students (19–30 years old, mean age = 23.8, SD = 2.7; 5 left-handed) from the University of Tübingen participated for a small monetary remuneration. Only women were tested in order to account for the fact that the recordings were exclusively made by female speakers. German was their only mother tongue[11], they did not suffer from any hearing disorders, and had normal

[11] Fifty participants indicated dialect familiarity (42 specifically with a southern German dialect, mostly Swabian). As most dialect speakers had exposure to a southern German dialect, the variable *Southern Dialect* was tested in

or corrected-to-normal vision. Two participants were excluded due to unsuccessful calibration, resulting in the collection of data from 72 participants (26 production group, 22 listening group, and 24 control group).

Material

Words during the test phase. Ninety-two word quadruplets were presented during test, each containing four German nouns. Twenty-eight quadruplets were based on critical word pairs; Sixty-four quadruplets were based on filler word pairs. The 28 critical word pairs (see Appendix B, Table B1) were each composed of a target word with an initial voiced stop and a competitor word starting with the respective voiceless stop. Only target words were presented auditorily during the experiment. Fourteen had a bilabial onset (e.g., target *BALKEN 'beam'* — competitor *PALME 'palm tree'*), and 14 had a velar onset (e.g., target *GITTER 'grid'* — competitor *KITTEL 'tunic'*).

Plosives were opted for, because it has been shown that perceptual learning of plosives can generalize across speakers (Kraljic & Samuel, 2007), arguably because they contain hardly no talker-specific information in comparison to fricatives, for example. This was important because participants in the training groups were trained with a different voice than was heard during test. Voiced stops occurred only in the initial position of target words. The initial stop consonant was always followed by a vowel[12]. Apart from the initial consonant, target and competitor overlapped in at least two segments. When the target words were presented auditorily, the initial voiceless plosives were devoiced (*Balken* was pronounced as **Palken*), resulting in overlapping

initial data analyses, resulting in no significant effect. Participants were not selected based on dialect competence, and it was not counter-balanced across conditions; therefore this factor was not included in the methods section.

[12] In some varieties of German, speakers tend to devoice initial voiceless stops when they are followed by a liquid (e.g., *grillen 'to barbecue'* can be pronounced as *krillen*). By always having a vowel following the initial consonant, potential previous experience with the accent was avoided.

word onsets of target and competitor for at least three segments. Auditory words with the native accent (**Palken*) were never existing words of German.

Mean log-frequencies of target words were 0.61 per million for velar stop words, 0.85 for bilabial stop words, and of competitors 0.67 for velar stop words, and 0.88 for bilabial stop words according to the CELEX word form dictionary (Baayen, Piepenbrock, & Gulikers, 1995). In order to form quadruplets, each of the 28 critical target-competitor pairs was paired with two semantically unrelated distractor words that matched in frequency with the target-competitor pair. Distractor words never had a stop in initial position but could contain stop consonants in other word positions.

The 64 filler word pairs also had a target and a competitor. There were 8 targets with initial /k/, 8 with initial /p/, 16 with initial /t/, and 32 targets with no initial stop in onset position (the *no-stop targets*). For the total of 32 targets with initial /k/, /p/, and /t/, half of the competitor word onsets overlapped with the target word onset by at least three segments, and half were phonologically unrelated. Two phonologically and semantically unrelated distractors were added to each target-competitor pair. The 32 no-stop targets were paired with competitors that also did not have stops in initial position. However, half of them overlapped in onset with the target for at least two segments (e.g., target *Seife 'soap'* — competitor *Seite 'side/page'*). There were four types of distractors for the 32 no-stop target-competitor pairs, each containing eight distractor pairs. The bilabial (b/p), velar (g/k), and alveolar (d/t) distractor pairs followed the same prerequisites as the corresponding critical target-competitor pairs. As they were not presented auditorily, stop+consonant onsets (e.g., *Brosche 'brooch'* — *Prospekt 'brochure/leaflet'*) were allowed. The fourth group had two semantically and phonologically unrelated initial sounds that were never stops.

Altogether, the test included 92 critical trials and four practice trials. Half of the critical targets and their corresponding competitors had been included in the preceding training phase, and half were new to participants. Likewise, half of the targets not starting with a stop (other-group) were new to participants and half were familiar from the training. Every participant had her own experimental list, each starting with the same four practice trials. Filler and critical trials were equally distributed across the lists, and a critical trial was always followed by at least one filler trial. There were never more than two old and not more than five new trials in a row. The various filler conditions were equally distributed across the lists.

Words during the training phase. Seventy-two single words from the above described target-competitor pairs were used for training. They included half of the devoiced targets (seven targets with bilabial onset, e.g.,**Palken*, and seven targets with velar onset) and their respective competitor (*Palme* for target **Palken*). The devoiced and voiceless items were included twice in the training list, resulting in 28 devoiced and 28 voiceless trials. Additionally, 16 filler targets from the no-stop targets were included, resulting in 72 training trials in total. Training trials with the same initial sound did not occur more than twice in a row, and each devoiced item was followed by at least one canonical item.

Recordings of test and training tokens. All tokens used for training and test were recorded by two female native speakers of Standard German without a noticeable regional accent (speaker A: 23 years; speaker B: 28 years). The speakers did not differ significantly in F0-range or speaking rate, and the authors judged their pronunciation to be comparable. Two different speakers were recorded to have different voices for both the training and test phases in the listening group. This permitted constant conditions across the training groups because the production condition always involved a different speaker during the training (the participant) and the test (the pre-recorded talker). Nevertheless, speaker-listener similarity was granted by having participants and pre-recorded speakers with the same sex and L1 background in the test phase. Acoustic differences between the training and test tokens are held as small as possible. Moreover, it can be tested whether speaker-specific effects, as observed by Trude and Brown-Schmidt (2012), can generalize to new speakers of the same sex (both female).

Recordings were carried out in a sound proof cabin with an Olympus LS-11 sound recorder (44.1 kHz; 16-bit). Every target word was recorded in the context of the carrier sentence *Klicken Sie jetzt auf 'Now click on'*. The devoiced target words (**Palken*) and the voiceless words (*Palme*) were all recorded naturally, that is, the speakers were explicitly instructed to pronounce the /b/ in *Balken* the way they would pronounce the /p/ in *Palme*. The best exemplar of the carrier sentence was chosen for each voice, and the duration of the carrier sentence was matched between both voices. Then, the carrier sentence was added to each target word recording.

Procedure

An *SR-Research Eyelink 1000* set-up was used for data collection with a sampling rate of 1,000 Hz, and the experiment was programmed with *Experiment Builder* (version 1.10.1025, SR Research Ltd.). Before the experiment started, the dominant eye of each participant was determined. Then, participants were seated in front of a computer screen and placed their chin on a chin rest. They were brought to a position in which they could stay comfortably for the duration of the experiment (~30 minutes). The eye-tracker was calibrated; then written instructions were shown on the screen. Participants had as much time as they needed to read them and initiated the experiment with a mouse click.

Training. The same training list was presented to each participant from the two training groups, while the control group received no training. The training tokens were either presented visually and auditorily (listening group) or visually only (production group). The listening group first saw a fixation cross for 1,000 ms, then the orthographic transcript of the training word (black Arial font, size 24) appeared in the center of the screen. It corresponded to German spelling rules (*BALKEN*). The initial letter that corresponded to the devoiced sound was colored red. Five hundred milliseconds later, the training word was played (**Palken*). Participants listened to the single words (devoiced, voiceless, and fillers) through over-ear headphones (Sennheiser HD 215 II) and at the same time fixated the transcript on the screen. There was a 2,000 ms inter-trial interval. Participants in the listening group were explicitly told to listen attentively to the words and to be aware of the speaker's accent while fixating the orthographic version of the words. Witteman, Weber, and McQueen (2013) showed that a single word context is sufficient for listeners to learn an accent. In their cross-modal priming task participants without previous accent training had increased priming effects in the second half of the experimental list compared to the first half.

The production group did not wear headphones during the training. They saw the same orthographic transcript of the words on the screen and had to read every single word out loud while their productions were recorded. Participants were asked to pronounce the initial red letter 'B' as /p/ and the initial red letter 'G' as /k/. Before every single trial, there was a fixation cross for 1,000 ms, and then the word was shown for 3,500 ms (accounting for the timing in the listening condition: 500 ms before the sound + 1,000 ms mean word duration + 2,000 ms pause). The next trial would then start. Between training and test, the written instructions for the eye-tracking task were shown on the screen. The production group had about five seconds to put on their headphones, and the listening group waited for five seconds until the initiation of the test phase. Overall, the training took about seven minutes for each participant.

Test Phase. The test phase started with four practice trials. A fixation cross preceded each trial for 1,000 ms, then four printed words from a word quadruplet were shown on the screen for 500 ms. The words were printed in black *Times New*

Roman font, size 34 on a screen with a white background. Screen resolution was 1,024×768 pixels, and the words were distributed across four different positions (256×576, 768×576, 256×192, and 768×192 pixels), see Figure 3-1. Display positions of target and competitor were pseudo-randomized, and the target never appeared in the same display position more than three times in a row. The mouse cursor (represented by a small circle) was located in the center of the screen at the beginning of each trial. Then the carrier sentence (about 1,300 ms) followed by the target word was played auditorily. Participants clicked on the target word with the mouse. Visually, participants saw the target word in its correct spelling (*BALKEN*); auditorily, it had the same accent as presented during the training phase (**Palken*). A small fixation circle appeared on the screen after every six trials to initiate an automatic drift correction in the calibration of the eye-tracker. The experiment concluded with a language background questionnaire based on the *LEAP-Q* (Marian, Blumenfeld, & Kaushanskaya, 2007).

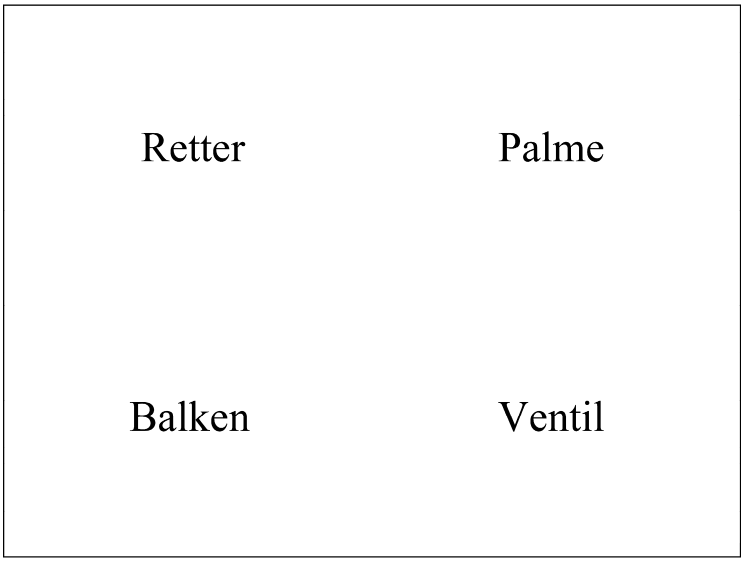

Figure 3-1. *Example display of a test trial in Experiments 5 and 6. In Experiment 5, BALKEN was the target and PALME the competitor. In Experiment 6, PALME was the target, and BALKEN the competitor. RETTER and VENTIL were distractors in both experiments.*

Analysis and Results

During training, the production group performed the instructed accent quite well. The experimenter decided based on perceptual judgments whether the critical training tokens were devoiced as communicated in the instructions. Every instance where the devoicing was not clearly perceivable was documented and subsequently validated by means of acoustic measurements of the recordings. On average, only 0.7 out of 28 critical trials were not devoiced as instructed. The proportion of correct clicks on the target during the test phase was 94.3 % (equally distributed across the training groups). However, five participants did not see the mouse cursor due to technical problems.

Fixation reports were extracted with the software *EyeLink Data Viewer* (version 2.4.1, SR Research Ltd.) and then the data was further processed with the software *R* (R development core team, 2017). The data from each participant's dominant eye was used to determine the coordinates and timing of fixations. Only fixations that fell within a cell of one of the four interest areas — target, competitor, and two distractors — were analyzed (exclusion of 3.4 % of the data). The interest areas each had a cell size of 472×344 pixels with a distance of 40 pixels between vertical cells and 60 pixels between horizontal cells. Saccades (20.8 % of the data) were not added to fixation times.

Then, the fixations for the four interest areas were analyzed in 20-ms steps in a time window from 0 to 1,000 ms after target word onset. The dependent variable *target* indicated whether in the respective 20-ms step a participant fixated the target; *competitor* indicated a competitor fixation, and *distractor* a distractor fixation. This resulted in three variables with binary values. Target and competitor fixation proportions were calculated with the empirical logit function. The plotted fixation proportions were inspected visually to determine the critical time window to which linear mixed effects regression models (Baayen, Davidson, & Bates, 2008; Bates, Kliegl, Vasishth, & Baayen, 2015) were then applied. For each analysis an individual, best fitting model was built that included a particular choice of fixed and random factors. Random effect structure included random intercepts for participants and items as well as those random slopes that significantly improved the model fit as tested by likelihood ratio tests. Significance of factors was indicated by t-values>|2|. Corresponding p-values, as reported in the text below, were determined with likelihood ratio tests. As fixed effects, *training* (production versus listening versus no training), *familiarity* (old, i.e., included in the training, versus new), *sound condition* (bilabial versus velar word initial sound), and *speaker* (speaker A versus speaker B) were considered. Proportion of target fixations was the dependent variable.

Descriptive analysis. Not surprisingly, the distractors were ruled out as potential target words very early by the participants (from about 200 ms, see Figure 3-2), i.e., the fixation proportion of distractors decreased very quickly. As launching a programmed eye movement usually takes about 200 ms (e.g., Altmann & Kamide, 2004), word processing is reflected in fixation proportions from this point in time on. Competitors were preferred over targets by all three groups from about 280 ms on until about 700 ms.

Target fixations show that the two training groups started to fixate the target more often than the control group from about 250 ms on. The advantage of both training groups became more pronounced and started being robust from about 350 ms on. Visually, there was no difference between the production and listening groups. Statistical analyses were run for the time window 250-750 ms because it included the whole process of target-competitor disambiguation, and here it became evident that the two training groups had a stable advantage of target fixations compared to the control group. As can be seen in Figure 3-2, the actual advantage lasted much longer — at least until 1,000 ms.

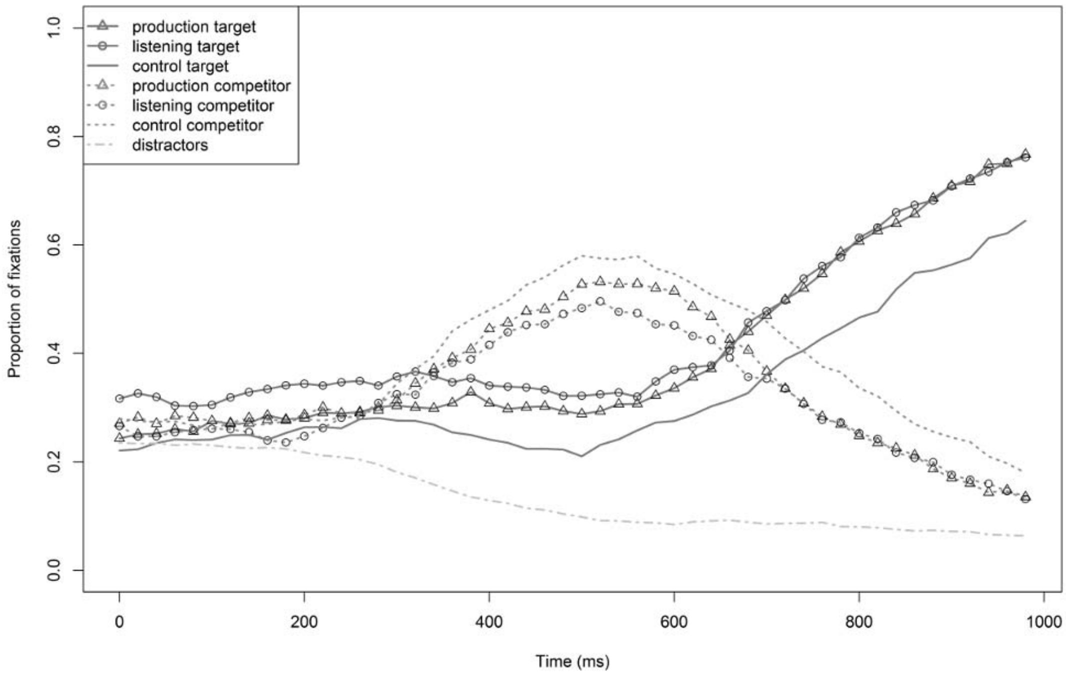

Figure 3-2. *Proportions of target (BALKEN) and competitor (PALME) fixations of the production, the listening, and the control group in Experiment 5. The bottom line describes the mean number of distractor fixations of all three groups.*

Statistical analysis. First, a model with data in the time window 0–200 ms was run. This tested looking biases before processing of the actual target word began. *Training group* was the fixed effect, and *participant* and *item* were random intercepts. There was a significant effect by training ($\chi^2 = 7.2$, $p < .03$); the results of the mixed model show that the listening group had more target fixations than both the control group ($ß_{training} = 0.39$, SE = 0.15, t = 2.6) and the production group ($ß_{training} = 0.31$, SE = 0.15, t = 2.1), hinting at a target bias for this group.

The second model analyzed data between 250–750 ms. It included *training group* and *sound condition* as fixed effects as well as *participant* and *item* as random intercepts. Training was significant ($\chi^2 = 10.7$, $p < .005$); both the listening group ($ß_{training} = 0.48$, SE = 0.15, t = 3.2) and the production group ($ß_{training} = 0.33$, SE = 0.14,

$t = 2.3$) fixated the target more often than the control group. There was no difference between the two training groups ($t = 1.0$). Furthermore, there was a main effect of sound condition ($\chi^2 = 7.5$, $p < .007$), resulting in more target fixations for bilabial than velar items ($ß_{condition} = 0.35$, $SE = 0.13$, $t = 2.8$). Due to the bias for the listening group found from 0–200 ms, the critical time window was further examined. On average, from 0–200 ms the proportion of target fixations was 8 % higher for the listening group than for the control group. To account for this early bias, 8 % were subtracted from listening group data between 250–750 ms and the same model was re-run with the modified data. Despite the reduction of the listening group's target fixation data, training was still significant ($\chi^2 = 6.2$, $p < .05$): the listening group still fixated the target more often than the control group ($ß_{training} = 0.30$, $SE = 0.15$, $t = 2.0$), and there was no difference between the production and listening groups ($t = 0.2$). This suggests robust differences between the control group and the two training groups.

Interim discussion

Experiment 5 found that accent learning was possible after both listening and production training. The proportion of target looks in both training groups was higher than in the control group. The listening group, however, fixated targets more often than the other groups, even before actual target word processing began, which might be argued to have affected the listening group advantage in the subsequent critical time window. This, however, can be excluded because the pattern of results persisted even when the fixation data of the listening group in the larger, later time window were penalized for its advantage in the initial, smaller time window.

There were no effects of speaker, i.e., learning occurred equally well with speaker A and B. The main effect for sound condition may be related to specific sound properties but does not further affect the general pattern of results. Moreover, the same pattern was observed for old tokens from the training phase and new tokens, indicating learning of a rule that generalizes to new words. The present results suggest accent learning for the production and listening groups, with no difference between the two training groups. Thus, robust effects of accent training when testing single accented words were found, hinting at a great effect by target words' accent as objective criterion of salience. Production and listening training seemingly do not differ from one another for L1 in terms of salience.

Experiment 5 provides evidence for successful accent learning after listening to or producing an accent. However, the canonical competitors (*Palme*) were activated for a very long time (until about 700 ms) before the devoiced target word was fixated more often. This time window covers the entire initial portion of the word before disambiguation (average disambiguation point: 280 ms + 200 ms for launching the eye movement = 480 ms; earliest disambiguation point: 150 ms + 200 ms = 350 ms; longest disambiguation point: 420 ms + 200 ms = 620 ms) and even longer. This suggests that, despite successful accent learning, canonical word forms still remained more easily accessible than accented word forms. There was potentially not enough accent exposure for the accented forms to be able to fully compete with canonical word forms. It is suggested that if a learned accent were to be able to have effects on the access of canonical, voiceless words with the same onset as the accent's target form, a greater amount of training is required.

Experiment 6 examines whether accent learning can be strong enough as to affect the processing of voiceless, canonical words with double the amount of accent training. Successful accent learning could imply competition effects from words that were previously not included as competitors. Thus, accent training has potentially not only effects for understanding accented word forms, but accented forms can function as competitors and affect the recognition of canonical word forms. As opposed to Experiment 5 where highly salient accented target words were tested, Experiment 6, focuses on test words that are expected to have a much smaller degree of salience based on the objective criterion *accent*, i.e., Standard German canonical words. Training effects of devoiced words (**Palken*) were tested on words that canonically start with the accent's target sound (*Palme)*. In order to increase the likelihood that accented forms could influence target recognition in their function as competitors, the training was doubled. If the accent is robustly learned, fewer target fixations would be expected by the training groups than without accent training.

Experiment 6

Again, three participant groups were tested. The training involved the same tokens as in Experiment 5, but the amount of training with the devoiced tokens was doubled. During test, participants did not hear the devoiced words (**Palken*) this time, but voiceless, canonical words (*Palme*), while seeing the same printed words on the display as the participants from Experiment 5.

Participants

Seventy-eight female students from the University of Tübingen participated for monetary reimbursement. Six had to be excluded due to calibration problems, resulting in 72 participants (18–31 years old, mean age = 23.2, SD = 3.2; 14 left-handed) who successfully completed the experiment. None of them suffered from any hearing disorders, all had normal or corrected-to-normal vision, and German was their only mother tongue[13]. The participants were randomly assigned to one of the three experimental groups (24 production, 24 listening, and 24 control group).

[13] Fifty-three participants indicated dialect familiarity, 50 of whom had exposure to a southern German dialect (mostly Swabian). The variable *Southern Dialect* was tested in initial data analyses, resulting in no significant effect, as in Experiment 5.

Methods and Material

The training list was based on that of Experiment 5. However, devoiced (**Palken)* items were presented twice in a row (rather than just once), resulting in 100 training trials in total (twice the amount of training with the devoiced tokens compared to Experiment 5). Due to the greater amount of training, the training phase took one minute longer.

During test, the same word quadruplets were presented on the screen — 92 critical trials and 4 practice trials with the same properties as in Experiment 5. However, the roles of target and competitor words were switched. Targets were now voiceless tokens (*Palme*) in their canonical form, and competitors were words that have a voiced onset in their canonical form (*Balken*). Auditorily, voiceless words were presented (*Palme*) that matched in their onset with the target word on the screen (*PALME*). Voiced tokens (*BALKEN*) that had been devoiced during the training (**Palken*), were visually presented competitors. All target words had already been recorded in the recording session for Experiment 5 by the same female speakers.

Analysis and Results

The same procedure for analysis as in Experiment 5 was applied. During training, the production group performed quite well in accomplishing the substitutions (mean: 0.8 errors out of 56 devoiced word trials). The accuracy of clicks during the test phase was 99.8 % (equally distributed across training groups). Saccades (17 % of the data) and fixations that did not fall into one of the four interest areas (3 %) were removed prior to analysis.

Descriptive analysis. Figure 3-3 illustrates the proportions of target, competitor, and distractor fixations of the production, listening, and control groups. The distractors were ruled out from the beginning of word processing (200 ms) on, meaning that the proportion of fixations decreased. Target (*PALME*) preference started very early (at about 250 ms), and competitors (*BALKEN*) were quickly ruled out as potential target words. The competitors stayed at a relatively stable level of activation between 200–400 ms, and then fixations decreased noticeably. This represents approximately the interval where target and competitor still overlap (mean overlap: 273 ms).

Target fixations by training group did not differ from one another from the beginning until the overall mean end of word processing (measurements of the voiceless target words resulted in a mean word duration of 767 ms). Disambiguation between target and competitor occurred relatively early, and there was no clear advantage of one of the training groups in target fixations. Statistical analyses were run for the time window between 250–750 ms, which included the entire disambiguation process between targets and competitors and parallels analyses in Experiment 5.

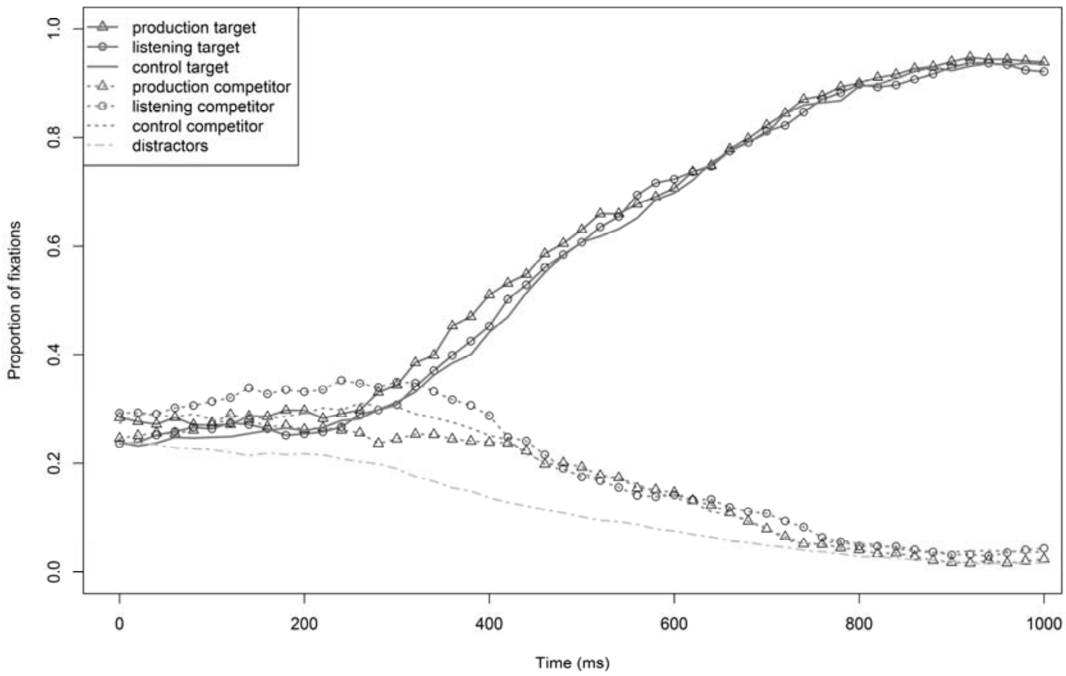

Figure 3-3. *Proportions of target (PALME) and competitor (BALKEN) fixations of the production, the listening, and the control group in Experiment 6. The bottom line describes the mean number of distractor fixations of all three groups.*

Statistical analysis. The baseline model for target fixations (0–200 ms) revealed no significant effect by training (t < 1). Mixed effects models revealed no significant effect of any of the considered fixed effects (all t-values < |1.3|) in the critical time window (250–750 ms). Auditorily presented voiceless words (*Palme*) that start with the same onset as the trained, devoiced words (**Palken*) triggered strong target activation from the beginning of word processing on. There was no effect of learning, neither by the production nor the listening group. In contrast to Experiment 5, the test words did not have the critical accent, but the voiceless paired words with the same sound onset as the devoiced, accented words were tested. As the devoiced training words included a sound substitution, the question is, especially for the production group: How much did the acoustic realizations of the devoiced tokens encountered during training differ from those of the voiceless tokens encountered during test? In other words, did the participants' own productions of the accent differ enough from the productions of the test speaker to prevent generalization across speakers? The missing training effect for both groups reinforces the question of effects of single tokens' acoustic properties. Therefore, acoustic properties of both the training materials and the test materials were analyzed in a next step.

Acoustic analyses. Pre-recorded target and training stimuli as well as the tokens produced by the production group during training were analyzed acoustically. This tested if the difference between training and test stimuli was too great for adaptation effects to be observed. Particularly in the production group, the acoustic properties of the accented plosives were likely to vary individually. The stops that mark the manipulated accent were focused on in the analysis. Voice onset time (VOT) and burst intensity (relative to total word intensity) were measured for each token that was part of an old critical word pair, i.e., word pairs that were included in both the training and the test phase. Only old word pairs were included in analyses, because there was no reference to the training phase for new words.

Each critical voiceless word (*Palme*) and its devoiced paired word (**Palken*) was considered for analysis. Both instances were taken as separate reference points in order to calculate the differences of the respective acoustic property value between the training and the test token (*Palme*). In the following, the *Palme-Palme* comparison is referred to as the voiceless word pair and the **Palken-Palme* comparison as the devoiced word pair. During training, one word was presented several times (devoiced: four times, voiceless: twice). This did not pose any problem for the listening group items because the same recording was presented several times. For the production group, however, single tokens differed individually. This issue was solved by taking average values. Two VOT- and two burst intensity difference values were assigned to each critical word for each participant — one with the values from the voiceless word as a reference point (*Palme*) and one with the values from the devoiced word as a reference point (**Palken*). Voices differed between training and test in both the listening and the production condition, so minor differences were inevitable.

First, the absolute training-test differences of the acoustic properties of the initial stops were compared:

$$\text{dif(stop value)} = |\text{stop value}_{\text{Test(Palme)}} - \text{stop value}_{\text{Training(*Palken or Palme)}}|$$

Measurements for all old word pairs were made for VOT (min = 0.14 ms, max = 71.8 ms, mean = 19.9 ms, SD = 16.6) and intensity ratio of the burst (min = 0, max = 0.35, mean = 0.08, SD = 0.06). These values were compared between the listening and production groups as well as between the devoiced and voiceless training words that included a stop.

For each VOT difference and burst intensity difference mixed effects models were run. Each model included the acoustic variable of interest as the dependent variable. *Training* (listening versus production) and *word pair* (devoiced versus voiceless) were considered fixed effects, and *participant* and *item* were random effects. The model for VOT differences also included by-*participant* random slopes for *training* and *word pair*, as well as by-*item* random slopes for *training*. None of the factors resulted in significant effects for VOT difference (all t-values < |0.7|).

The model for burst intensity included by-*participan*t random slopes for *word pair* and by-*item* random slopes for *training*. There was a significant interaction between training and word pair ($\chi^2 = 5.6$, $p < .02$) illustrated in Figure 3-4. Examining the results of the mixed model (see Table 3-1), this interaction is based on the smaller burst intensity difference for devoiced word pairs in the production group than the listening group (t =-2.15), and there was no difference for voiceless word pairs between training groups (t = 1.23). Within training groups, there was no training-test difference between devoiced and voiceless word pairs (t-values < 1.8).

Table 3-1. *Results for burst intensity differences between training and test words in Experiment 6 as calculated by the model lmer(burst difference~word pair*training + (1+word pair|participant) + (1+training|item)). The factors were releveled in order to calculate the model with different intercepts. This allows displaying t-values for all relevant level comparisons. ß = Estimate, SE = Standard Error*

Predictor	ß	SE	t
Intercept (devoiced, listening)	0.10	0.01	7.51***
word pair = voiceless	-0.03	0.02	-1.75
training = production	**-0.03**	**0.01**	**-2.15***
word pair*training	**0.05**	**0.02**	**2.49***
Intercept (voiceless, listening)	0.07	0.01	5.00***
word pair = devoiced	0.03	0.02	1.75
training = production	0.02	0.01	1.23
word pair*training	**-0.05**	**0.02**	**-2.49***
Intercept (devoiced, production)	0.07	0.01	9.31***
word pair = voiceless	0.01	0.01	1.61
training = listening	**0.03**	**0.01**	**2.15***
word pair*training	**-0.05**	**0.02**	**-2.49***

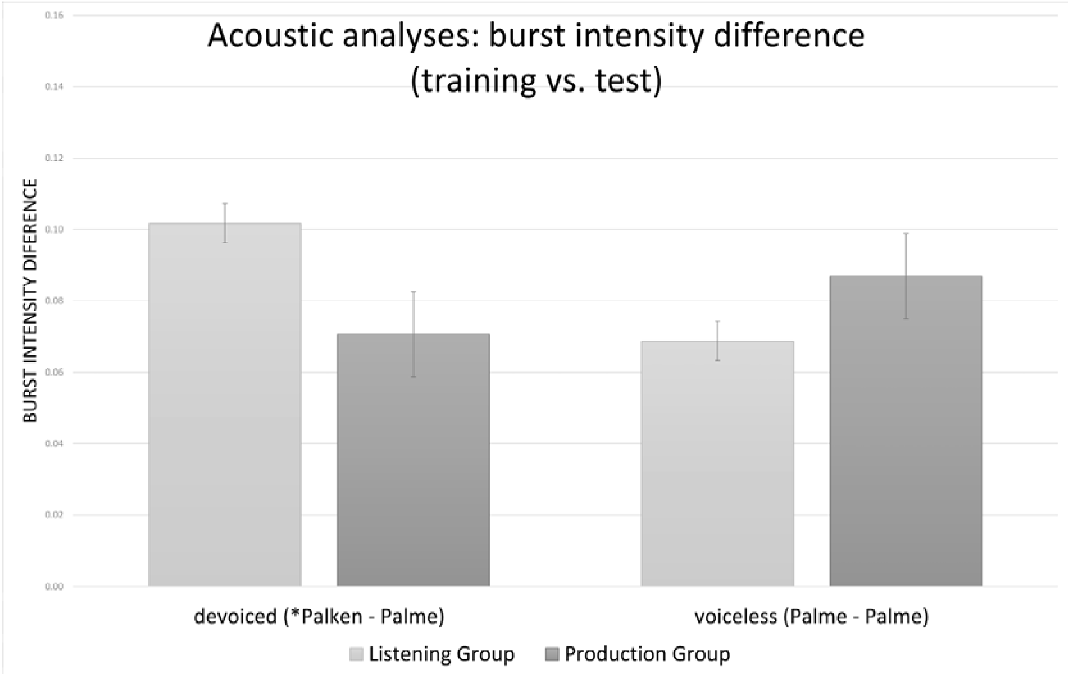

Figure 3-4. *Acoustic differences of relative burst intensity for devoiced (*Palken) versus voiceless (Palme) word pairs and listening versus production training in Experiment 6. Whiskers represent standard errors.*

Interim Discussion

Neither training group fixated the target less often than the control group without training did. They did not differ from one another in their amount of target fixations. The recognition of voiceless *Palme* was not affected by previously having learned that *Balken* is pronounced as **Palken*. This occurred despite the fact that accent training was intensified by presenting devoiced tokens twice as often as in Experiment 5. This is good news for native accent listeners, because it shows that learning a new accent does not immediately distort comprehension of canonical forms. Concrete acoustic analyses tested whether this effect was due to greater inherent salience based on an objective criterion of devoiced (as tested in Experiment 5) compared to voiceless tokens or rather because of greater acoustic differences between the devoiced training and the voiceless test tokens. There was no VOT difference between training groups, thus the production group was quite good at accomplishing the substitutions. The few production errors that occurred did not affect the overall pattern. This was also supported by the observation that burst intensity differences were even smaller for the production group than the listening group.

Discussion

The present study investigated whether different forms of native accent training and different token realizations (unaccented versus accented) differ in salience for L1 participants. This was measured by the amount of learning of the native accent. As a subjective criterion of salience, the training phase was varied by having production and listening accent training (versus no training), and an objective criterion was tested by the nature of the test tokens (accented/devoiced words in Experiment 5 versus canonical/voiceless words in Experiment 6). In Experiment 5, native German participants produced or listened to single German words that featured the devoicing of initial voiced stops (*Balken* pronounced as **Palken*).

In the subsequent eye-tracking task, participants from both training groups fixated the devoiced target more often than participants without training did, with no difference between the two training groups. This was true whether the accented target word had been included in the preceding training or if it was presented for the first time. Experiment 6 started with the same accent training and in the test — Standard German canonical words with the same onset as the devoiced tokens (*Palme*) were targets. The proportion of target looks was not affected by training. Acoustic analyses showed that devoiced training words (**Palken*) and voiceless test words (*Palme*) did not differ strongly in their onset.

Salience and learning

In Experiment 5, there were significantly more looks to devoiced targets after production and listening accent training than without training. In Experiment 6, which featured voiceless target words, target looks did not reveal accent learning. This can be explained by the role of salience in accent learning. Two criteria for salience were manipulated and tested in the present study. First, an objective criterion was tested through the nature of the test tokens (accented/devoiced test words in Experiment 5, canonical/voiceless test words in Experiment 6). The devoiced test words were predicted to be more salient than the voiceless words, thereby resulting in greater learning effects. Second, a subjective criterion was tested by having an accent training session based on different modalities (production and listening).

In terms of the objective criterion, learning only showed effects for devoiced, and not for voiceless target words that had the same word onset (**Palken* versus *Palme*). This suggests that devoiced tokens are more salient than voiceless tokens. Acoustic analyses of Experiment 6 support the present interpretation. There was no evident acoustic difference between devoiced and voiceless word onsets that could have inhibited learning. Training was still effective, though not visible, because the test tokens were not as salient as in Experiment 5. Test tokens in Experiment 5 and 6 therefore only differ perceptually from their disambiguation point onward (after /pal/ for **Palken* and *Palme*).

This implies that training effects emerged only in later stages of processing, after word disambiguation. This is supported by the analysis in Experiment 5, where the training group advantage admittedly was already detectable from about 250 ms on (see Figure 3-2); however, the plot of fixation proportions suggests that the two training groups' advantage increased over time and became stable from about 350 ms on. The shortest duration of the ambiguous word section (i.e., overlapping with the competitor) measured approximately 150 ms in Experiment 5 (*ger* in *Germane*

'Teuton', speaker A). The moment where the information after the disambiguation point is processed is then reflected from about 350 ms after the stimulus onset onward (150 ms + 200 ms eye movement launching).

Cho and Feldman (2013) have found a memory advantage for accented compared to canonical words. They argue that accented speech is more variable in terms of acoustic and phonetic detail, and, based on an episodic account of the mental lexicon, they suggest that the difference between accented speech input and stored exemplars is greater than the difference between unaccented input and stored exemplars. Accordingly, this greater difference enriches the form-meaning relationship. This reasoning essentially follows the same principles as the distinctiveness account of salience. More distinct tokens are more salient, which results in memory advantages.

It can be argued that salience of accented tokens in the present study was artificially increased by the fact that there was only one specific accent marker and no more natural, global accent. However, a cross-modal priming study by Eisner, Melinger, and Weber (2013) found that L1 English listeners adapt to final devoicing in English (*seed*, pronounced as */si:t^h/*) when it was produced either by a native British English speaker or by a native Dutch speaker with L2 English (with global Dutch accent features). Moreover, the findings from the Cho and Feldman study are in line with the present findings. They incorporated a global accent (Dutch-accented English) and still found a memory advantage of accented over canonical tokens.

A subjective criterion of salience, on the other hand, was implemented through the training session. The production group was compared to the listening group as well as the control group without training. Accent learning worked equally well with both listening and production training in Experiment 5 (target **Palken*), and effects were not visible with voiceless (*Palme*) targets in Experiment 6. There was no difference between the two training groups in either experiment. This suggests that both production and listening accent training imply a similar amount of salience in the fostering of accent learning, and learning effects become visible only when the test token receives sufficient salience through an objective criterion.

Interestingly, in L1, salience elicited by the subjective criterion of producing an accent was as large as that of listening to the accent. In Chapter 2A, the effects of production versus listening training on accent learning were tested for both L1 and L2 participants. L2 participants learned the accent most easily with production training. L1 participants did not learn, neither with listening nor production training. Importantly, all speech in the present study was produced by L1 speakers, but in Chapter 2A, test items were always produced by an L2 speaker of English. Thus, for L1 participants in the production training group there was a switch in nativenesss of the speaker between training (L1) and test (L2). L2 learners likely involve a greater amount of variability (Wade, Jongman, & Sereno, 2007) in their speech than L1 speakers, including more accent markers which probably require additional processing.

Moreover, the similarity between listener and speaker is emphasized by the *integrated theory of language comprehension and production* (Pickering & Garrod, 2013), according to which a listener's previous production experience can affect comprehension. This experience is predicted to have greater effects with increasing speaker-listener similarity. The present results, however, do not necessarily support this suggestion. In spite of greater speaker-listener similarity (same sex, same L1 background, mostly similar dialects), the production group did not have greater training effects than the listening group. Nevertheless, having an L1, not L2, speaker produce the accent helped L1 participants to learn an accent after both listening and production training. This is in line with the findings from Experiment 4 (Chapter 2B), where the L1 participants and the pre-recorded speakers had the same native language background and learning was observed. Contrary to L2 participants in Chapter 2A, however, accent learning was not stronger after production training.

Producing an accent is only a more important subjective criterion of salience than listening, because of specific L2 properties (e.g., greater perceptual flexibility). There is no general advantage exhibited by producing compared to listening.

Taken together, there was arguably no advantage of production over listening training for L1 listeners, because production might only make a linguistic token more salient if it can act as objective, not subjective, criterion of salience. This would additionally include that the concrete situation determines salience. Furthermore, the studies that have found robust production effects (Cho & Feldman, 2013; MacLeod, 2011) were all memory studies that tested active and conscious word recall, thus later stages of processing. Contrarily, the present eye-tracking study tested online word processing. It is therefore also possible that the production advantage may not arise in the earliest stages of processing. Other studies conducted a repetition experiment rather than a listening-only task as in the present experiments (e.g., Cho & Feldman; Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015). Repetition includes listening and producing the critical token, possibly implying a greater amount of salience than only production. Finally, concrete feedback may have affected the results of the study by Kartushina and colleagues.

Referring again to the definition of salience established in the beginning of this chapter, MacLeod (2011) suggests that for mixed lists (including items both listened to and produced), produced items are more distinct and therefore more salient. This kind of salience likely relies on an objective criterion — the stimulus itself attracts attention because of its distinct physical characteristics. In the present study, on the other hand, it was asked if the nature of training (production versus listening) could act as a subjective criterion of salience. The present results do not support a production advantage *per se*, but they also do not exclude the possibility of a production advantage. The production advantage may function within the scope of salience that relies on an objective, but not subjective, criterion, even with L1 participants and in an online task. In summary, salience effects that rely on an objective criterion were found and no effects that rely on a subjective criterion. Previous studies that support the production effect have always tested salience arising from an objective criterion. It is hypothesized that the nature of salience is the crucial factor in the learning process and that, in short-term learning, objective criteria are more powerful than subjective criteria.

This contrasts at first sight with findings on dialect accommodation by Auer, Barden, and Grosskopf (1998), who emphasize the importance of subjective criteria of salience. Note, however, that the researchers refer to change in production over the long term rather than to comprehension in the short term, as was tested in the present study. Therefore, different criteria of salience might result in salience that is most efficient at different stages of learning and in different modalities. On the other hand, these results are good news regarding short-term comprehension learning in language change contexts. These contexts mostly involve new and old pronunciation variants, resulting in contrasts between the two. This provides well-suited conditions for an objective criterion of salience in terms of contrasts in phonetic realizations. Learning will be easier in contact situations than it would be in potential accent-only situations.

At the same time, as accent comprehension improves, comprehension abilities of the canonical pronunciation are not impaired. If the present results (and those from Chapter 2A) are applied to concrete L2 learning situations, it can be concluded that for learning new variations, L2 learners, thanks to their greater cognitive flexibility, can still achieve reasonable results without switching between production and listening. It would, however, probably be even more beneficial to integrate variation between the two modalities.

Competitor inclusion as a further step in learning

Learning was observed not only for old words that had been part of the training phase; it also generalized to new words with the same accent and furthermore from the voice of the training speaker to the unfamiliar voice of the test speaker. This finding supports abstractionist accounts of the mental lexicon (e.g., McClelland & Elman, 1986; Norris, 1994) rather than episodic accounts (e.g., Goldinger, 1998). Whereas episodic accounts suggest the storage of every concrete exemplar of a speech unit encountered by a listener (including speaker-inherent details, for example voice properties), in abstractionist models, abstract representations of a word's canonical representation build the lexicon.

Variations of the canonical form, such as accents, can be accounted for by pre-lexical mapping rules. These rules are built on the basis of a few exemplars that are no longer stored. When, for example, an accented token is encountered after accent training, the learned rule is applied to the respective abstract entry in the lexicon. This can explain why learning a specific variation can generalize across many different words (McQueen, Cutler, & Norris, 2006). However, existence of exemplars in the lexicon is not ruled out it. Hybrid models (e.g., McLennan, Luce, & Charles-Luce, 2003) attempt to integrate exemplars and pre-lexical rules into a single account.

In contrast to Bradlow and Bent (2008) and Sidaras and coworkers (2009), who observed speaker generalization only if training was conducted with multiple voices, one voice was sufficient for generalization in Experiment 5. The globally accented speakers in those studies likely featured many different accent markers, resulting in a stronger accent than the accent presented in the present experiments. With only few accent markers, it is easy to build pre-lexical accent rules allowing for generalization to new talkers. With many different accent markers, however, multiple exemplars from multiple talkers might help successful rule-building as argued by Bradlow and Bent (2008).

Moreover, Trude and Brown-Schmidt (2012) tested competitor exclusion and inclusion and found talker-specific adaptation effects. Competitor exclusion and inclusion describes modifications in the cohort of words initially activated when a word starts to be processed. Potential candidates can be excluded (competitor exclusion), or new candidates can be added (competitor inclusion). Effects of accent training on competitor activation are indirect training effects — the effects of the accent on other words (presented as targets) are then tested. These tokens seem less salient than accented tokens. It seems that if less salient targets are tested, the role of aspects such as talker specificity increases.

In other cases, these aspects may be training intensity or prior accent familiarity, as shown in Trude, Tremblay, and Brown-Schmidt (2013). The design of their study was similar to the eye-tracking study discussed earlier (Trude & Brown-Schmidt, 2012). Talker-specific effects of accent learning on competitor exclusion were again tested, but this time with a Québec-French accent that participants had never been exposed to before the experiment. The talker replaced every /i/ with an /I/ in English words, for example, *weak* was pronounced as *wick*. An accent training session did not help participants rule out unlearned competitors more easily if pronounced by the accented talker than the unaccented talker.

As suggested by Trude and colleagues (2013), competitor exclusion failed seemingly due to the accent being completely new to participants. Considering the small amount of target word salience, more previous accent exposure (as shown in Trude & Brown-Schmidt, 2012) or greater training intensity could have helped. This interpretation is also supported by Experiment 6 of the present study. In contrast to the accented, devoiced targets from Experiment 5, the canonical, voiceless targets in Experiment 6 implied smaller overall objectively induced salience. Additionally, the accent was completely new to the participants.

The present experiments demonstrate that after a few minutes of training, an accent can be learned so that it is more easily processed than without training. Only highly salient target tokens made learning effects visible. Therefore, accent training does not always exhibit robust accent learning. As shown by Trude and Brown-Schmidt (2012), this does not mean that more robust accent learning is not impossible. They found effects of both competitor exclusion and inclusion when non-salient target tokens were tested. The effect was talker-specific, and the participants already had prior (pre-experimental) experience with the accent.

Accent learning seems to occur in various steps, ranging from unadapted to partially adapted (effects can be observed for accented, salient words) all the way to fully adapted (effects can be observed for unaccented, non-salient words). Full accent learning would mean that the way that accented word forms function as competitors is similar to the functioning of canonical word forms. However, the amount of looks to the targets in Experiment 6 was the same with and without training, indicating that full adaptation had not occurred. It likely requires more intense training, pre-experimental accent familiarity, identical talkers during training and test, or even multiple talkers during training (Bradlow & Bent, 2008).

The present learning effects seem to reflect partial accent adaptation, which is still important because it allows a listener to better understand the accented form itself. The reason why no full adaptation was found can also lie in the native language background of the present listeners. Bent and Bradlow (2003) found that non-native listeners performed equally well in a sentence recognition task while listening to a speaker with the same L1 as when listening to a native speaker. This advantage has even generalized to unrelated accents that were new to the listener. Native listeners, on the other hand, as shown in a training-test study by Baese-Berk, Bradlow, and Wright (2013), are only able to generalize accent learning to a new accent if they are trained on many different accents. This finding is in line with the results on generalization of accent learning to new voices by Bradlow and Bent (2008).

Basic assumptions from abstractionist accounts on lexical processing support the conclusion that the accent rule was not learned strongly enough to be applied to all tokens from word onset onward. In Experiment 5, the voiceless competitors (*PALME*) of the target word **Palken* were considered as potential candidates for a long period, and only after disambiguation was the target *BALKEN* fixated more often than the competitor. With the auditory target *Palme* in Experiment 6, the pre-lexical rule was not learned strongly enough to establish additional competition by *BALKEN* during the /pal/-segment. One could assume that the results of Experiment 6 are due to increased competitor (*BALKEN*) activation. Participants learned that *Balken* is pronounced as **Palken*, so they might have concluded that *Palme* is pronounced as **Balme*. This is rather unlikely, however, because the training also included canonical words starting with the voiceless sound (*Palme*). Therefore, when hearing *Palme,* they did not interpret the word input as **Balme* and thus did not fixate the competitor more often than the target.

Native and foreign accent learning

In the present discussion, studies were included that tested learning of native accents as well as studies on foreign accent learning. Research on foreign accent learning clearly shows that in their L1, listeners quickly learn foreign accents produced by L2 speakers and maintain long-lasting processing advantages (e.g., Clarke & Garrett, 2004; Maye et al., 2008; Witteman et al., 2015; Witteman et al., 2013). Similar findings arise from native accent studies (Trude & Brown-Schmidt, 2012). It is therefore possible that a dichotomy between native and foreign accents is unjustified. Similar mechanisms could apply to both native and foreign accent processing.

Clarke and Garret (2004) suggested an accent processing classification that depends on the accent's acoustic distance from native speech. Foreign and native accents follow the same principles, but the strength of an accent could determine the nature of accent adaptation. Arguably, native accents can be closer to standard native speech than foreign accents. Processing of regional and foreign accents could then rely on similar mechanisms, but stronger accents induce greater processing effort than mild accents do. As a consequence, learning regional accents would be easier than learning foreign accents.

This account would explain why, on the one hand, similar results were found if the same accent was produced by an L2 or L1 speaker (Trude et al., 2013), and, on the other hand, greater processing difficulties were found for foreign than for native accents (Floccia et al., 2009; Floccia et al., 2006). Likewise, learning for L1 participants was found when an L1 speaker produced the contrived accent in the present chapter as well as in Chapter 2B, but in Chapter 2A, learning was not found when an L2 speaker produced the accent. It is suggested that accent strength is very likely linked with the amount of different accent markers that a speaker produces, which varies among individuals. Some L2 speakers do not exhibit many accent features, whereas others do. Therefore, concrete acoustic features could be an important variable which the magnitude of accent learning depends upon.

Conclusion

In conclusion, Chapter 3A suggests that native accent learning can be fast and easy, including generalization to new voices and new lexical tokens as well as learning through individual production. However, the accent requires salience that relies on an objective criterion during test in order to display its learning effects. The strength of accent learning is therefore limited; an accent is not learned well enough to affect the processing of other, non-salient canonical tokens. It is not integrated as strongly into the lexicon as canonical tokens. Learning was not affected by the training manipulation, which relied on a subjective criterion of salience. There are, however, studies that have found an advantage of production over listening when training functioned as objective criterion of salience. Therefore, in short-term accent learning listeners might be more sensitive to objective than to subjective criteria of salience.

Chapter 3B

Testing accent learning with the mobile *EyeTribe* eye-tracker

Abstract

Eye-tracking is frequently used in psycholinguistic research, but the hardware is quite expensive and not always suited for mobile use. Recently, however, a few low-cost, mobile trackers have been developed. In the present methodological study, the performance of the low-cost *EyeTribe* tracker was tested in a visual world experiment by replicating the accent learning study reported in Chapter 3A (Experiment 5). Target fixations recorded with an *Eyelink 1000* from the original experiment were compared with the newly collected *EyeTribe* data. Overall, the *EyeTribe* provides an acceptable replication of the pattern of results found in Experiment 5, suggesting that the *EyeTribe* is well-suited for use in visual world experiments with some restrictions.

Introduction

In eye-tracking, participants' eye-movements are recorded, usually while they look at a computer screen. Several eye-tracking devices exist, and they differ in terms of possible sampling rate, remote or non-remote tracking, mobility, experimental software, and importantly price. The present study tested the *EyeTribe* tracker (https://theeyetribe.com/), a newly developed device that is quite affordable and small (pocket-sized). The *EyeTribe* has relatively low sampling rates (either 30 or 60 Hz) and does not come with an integrated software. The aim of this study was to evaluate the performance of this device in a typical visual world paradigm.

The *EyeTribe* has gained attention from eye-tracking researchers and has been checked for use in a few non-linguistic scientific studies before. For example, the *EyeTribe*'s fixation accuracy and precision has been compared with well-established eye-trackers (Dalmaijer, 2014; Ooms, Dupont, Lapon, & Popelka, 2015; Popelka, Stachon, Sasinka, & Dolezalova, 2016). Accuracy is the mean difference between the position of a visual test stimulus and the measured gaze position (Holmqvist, Nyström, & Mulvey, 2012), that is, the participant fixates points on the screen and the distance of the recorded fixations from the actual location on the screen is measured.

Precision determines the eye-tracker's ability to produce consistent measurements (Holmqvist et al., 2012). Dalmaijer (2014) compared the *EyeTribe* with the *Eyelink 1000* (SR Research Ltd.). From his analyses of precision and accuracy, Dalmaijer concluded that the *EyeTribe* is well-suited for fixation analyses, point-of-regard analyses, and pupilometry, but not for analyzing saccades. This makes sense because saccadic movements are typically completed within 20–40 ms (Keating, 2013), but the *EyeTribe* captures eye gaze only every 16.67 (60 Hz) or 33.33 ms (30 Hz). The quality of results was also approved by Popelka and colleagues (2016) (for cartographic research) as well as Ooms and colleagues (2015) who compared the *EyeTribe* with the

SMI RED 250 (SensoMotoric Instruments). The latter, however, also concluded that the 30 Hz version is less suited for scientific research as a result of the low accuracy and large time interval between recorded gaze positions (33.33 ms).

In psycholinguistics, eye-tracking is frequently used in the visual world paradigm (Allopenna, Magnuson, & Tanenhaus, 1998). Typically, in the printed word variant of the paradigm (McQueen & Viebahn, 2007; Weber, Melinger, & Lara Tapia, 2007), participants' eye fixations to four possible display positions on the screen are measured. Participants see four printed words (including a target, a competitor, and two distractors) on a computer screen and are auditorily instructed to click on a target word while their eye movements are recorded. In this paradigm, researchers are mainly interested in fixations aggregated over a given time window and not in saccades as this task does not elicit eye movements that are as linear as in reading studies, for example (Dussias, Valdés Kroff, & Gerfen, 2013). Therefore, the low sampling rate of the *EyeTribe* is not necessarily a limitation for the visual world paradigm.

In summary, the *EyeTribe* has been tested for its accuracy and precision, but no study so far has done so in a visual world paradigm. Fixations on written text (Keating, 2013) typically range from 50–500 ms, implying that even with a sampling rate as low as 30 Hz, the *EyeTribe* should represent them accurately. Analysis methods in prior visual world studies support this idea. For example, in Chapter 3A as well as in Hanulíková and Weber (2012) data were recorded at 1,000 Hz but down-sampled in the subsequent analysis to 50 Hz (data points in 20-ms steps).

Moreover, the paradigm does not require high accuracy or precision because the regions of interest are determined by only four positions on the screen that are placed quite far from one another (see example screen in Figure 3-5). This was the impetus for the present study: in a visual world experiment, does the *EyeTribe*, with a temporal resolution of only 30 Hz, provide results comparable to a high-end eye-tracker? Low sampling rates imply smaller amounts of data and less processing time in analysis.

Lower sampling rates are also able to account for more participant head movement than higher sampling rates, which require the participant's head to be restricted to a highly stable position during the eye-tracking (Dussias et al., 2013).

The visual world experiment on accent learning in German from Chapter 3A (Experiment 5) that was run with an *Eyelink 1000* (desktop mount, not remote, with a chin rest) was replicated. In the training-test study, participants in the listening group were first exposed to single words that were produced with an accent. The control group did not have training and started the experiment directly with the test phase. In the test phase of the experiment, participants from both groups completed the eye-tracking task. They saw four printed words in their canonical spelling (a target, a competitor, and two distractors) on a computer screen and were auditorily instructed, using the same accent as in the training phase, to click on a target word. The proportion of target fixations was measured and compared between the listening and control groups. The listening group fixated the accented targets significantly more often than the control group. Whether the same pattern can be shown with the *EyeTribe* in its lowest sampling rate (30 Hz) was investigated in the present study.

Experiment 7

Participants

Sixty-four native German speaking female students from the University of Tübingen participated. Sixteen were excluded because the eye-tracker did not record any fixation data (n=4), more than 50 % of the data points were missing (n = 10), or because they had been already participants in the original study presented in Chapter 3A (n = 2). This resulted in the collection of data from 48 participants (19–29 years old, mean age = 23.1, SD = 2.5) from which 25 were assigned to the listening group, and 23 to the control group. German was their only mother tongue, they did not suffer from any hearing disorders, and all had normal or corrected-to-normal vision (19 wore glasses or contacts). Note that the *Eyelink* and the *EyeTribe* data were not collected simulatenously (i.e., every participant's eye movements are tracked with both devices) because this is not feasible with the available technical options.

Material

Words during the test phase. The same material as in Experiment 5 was used, which comprised 92 German word quadruplets during test, each containing four German nouns. Twenty-eight quadruplets were based on critical word pairs; sixty-four were based on filler word pairs. The 28 critical word pairs were each composed of a target word with an initial voiced stop and a competitor word starting with the corresponding voiceless stop. Only target words were presented auditorily during the experiment. Fourteen had a bilabial onset (e.g., target *BALKEN 'beam'* — competitor *PALME 'palm tree'*), and 14 had a velar onset (e.g., target GITTER *'grid'* — competitor *KITTEL 'tunic'*).

Apart from the initial consonant, target and competitor words overlapped in at least two segments. When the target words were presented auditorily, the initial voiceless plosives were devoiced (*Balken* was pronounced as **Palken*), resulting in overlapping word onsets of target and competitor in at least three segments. Auditory words with the native accent (**Palken*) were never existing German words. In order to form quadruplets, each of the 28 critical target-competitor pairs was paired with two semantically unrelated distractor words that matched in frequency with the target-competitor pair.

The 64 filler word pairs also had a target and a competitor, which are described in detail in Chapter 3A. Altogether, the test included 92 trials and four practice trials. Half of the critical targets and their corresponding competitors were included in the preceding training phase, and half were novel to participants. Likewise, half of the targets not starting with a stop were new to participants and half were familiar from the training. Every participant had their own experimental list, each starting with the same four practice trials. Filler and critical trials were equally distributed across the lists.

Words during the training phase. Seventy-two single words from the aforementioned target-competitor pairs were used for training. They included half of the devoiced targets (e.g., **Palken* and **Kitter*) and their respective competitor (*Palme* for target **Palken* and *Kittel* for **Kitter*). The devoiced and voiceless items were included twice in the training list, resulting in 28 devoiced and 28 voiceless trials. Additionally, 16 filler targets from the no-stop targets were included, resulting in 72 training trials in total.

Recordings of test and training tokens. The recordings of the test and training tokens were the same as in Experiment 5. All tokens were recorded by two female native speakers of German without a noticeable regional accent and with a similar F0-range. Two different speakers were recorded to have different voices for both the training and test phases in the listening group and thereby testing whether accent learning can generalize to new speakers.

Recordings were carried out in a soundproof cabin with an Olympus LS-11 sound recorder (44.1 kHz; 16-bit). Every target word was recorded in the context of the carrier sentence *Klicken Sie jetzt auf 'Now click on'*. The devoiced target words (**Palken*) and the voiceless words (*Palme*) were all recorded naturally; i.e., the speakers were explicitly instructed to pronounce the /b/ in *Balken* the way they would pronounce the /p/ in *Palme*. The best exemplar of the carrier sentence was chosen for each voice, and the duration of the carrier sentence was matched between both voices. Then, the carrier sentence was added to each target word recording.

Procedure

The experiment was programmed with *PsychoPy* (version 1.84.0; Peirce, 2008), and eye movements were recorded with an *EyeTribe* tracker with a sampling rate of 30 Hz. The experiment took place in a quiet, artificially illuminated room without windows. The *EyeTribe* was placed below the computer screen on its tripod. Then, participants were seated in front of a computer screen so that they were facing the eye-tracker. They were brought to a position in which they were sitting in front of the center of the eye-tracker and the screen and where they could remain comfortable for the duration of the experiment (~30 minutes). The eye-tracker was adjusted so that the participant's eyes were displayed in the center of the calibration screen and both eyes could be tracked. Then, the calibration with a 9-grid display began. After successful calibration, written instructions were shown on the screen. Participants were given as much time as needed to read them and initiated the experiment with a mouse click.

Listening Training. A training list was presented to each participant from the listening group, while the control group received no training. Training tokens were presented both visually and auditorily. The listening group first saw a fixation cross for 1,000 ms, then the orthographic transcript of the training word (black Arial font, size 24) appeared in the center of the screen. It corresponded to German spelling conventions (*BALKEN*), and the initial letter that corresponded to the devoiced sound was colored red. 500 ms later, the training word was played (**Palken*). Participants listened to the single words (devoiced, voiceless, and fillers) through over-ear headphones (Sennheiser HD 215 II) and at the same time fixated the transcript on the screen. There was a 2,000 ms inter-trial interval. Participants in the listening group were explicitly told to listen attentively to the words and to be aware of the speaker's accent while fixating the orthographic version of the words. Overall, the training took about seven minutes for each participant.

Test Phase. The test phase started with four practice trials. Before each trial, first a blank screen (500 ms) and then a fixation cross (1,000 ms) were shown. Four printed words from a word quadruplet were then shown on the screen (white background) for 500 ms. Screen resolution was 1920x1080 pixels (1024×768 pixels in Experiment 5). The words were distributed across four different positions (480×810, 1440×810, 480×270, and 1440×270 pixels), see Figure 3-5.

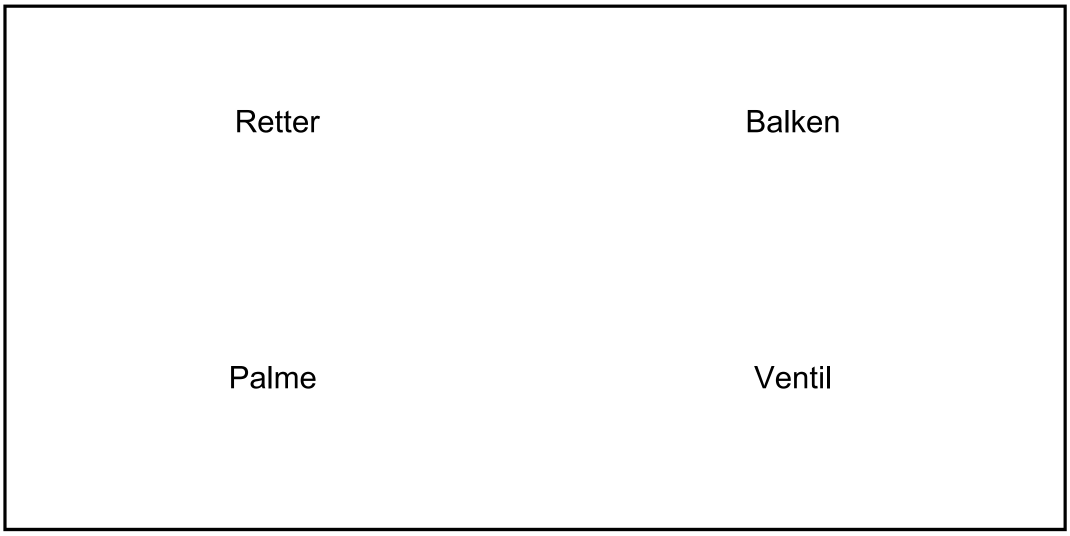

Figure 3-5. *Example display including the target (BALKEN), distractor (PALME), and the two competitors.*

Display positions of the target and competitor were pseudo-randomized, and the target never appeared in the same display position more than three times in a row. Then the carrier sentence (about 1,300 ms) followed by the target word was played auditorily. Participants clicked on the target word with the mouse. The mouse cursor (represented by a small arrow) automatically returned to the center of the screen at the beginning of each trial. Visually, participants saw the target word in its correct spelling (*BALKEN*); auditorily, it had the same accent as during the training phase (**Palken*). A small fixation circle appeared on the screen after every six trials to simulate the drift correction in the calibration of the eye-tracker in Experiment 5. Actual drift corrections between single trials are not possible with *PsychoPy*. The simulated correction, however, does bring participants' view back to the center of the screen in the same moments of the experiment as in Experiment 5. The experiment concluded with a language background questionnaire based on the LEAP-Q (Marian, Blumenfeld, & Kaushanskaya, 2007).

Analysis and Results

Data preparation. The *PsychoPy* output provided fixation reports for each participant. The data were processed with a Python script that extracted a data point every 30 ms. Then the timing and coordinates of fixations were determined. Only fixations that fell within a cell of one of the four interest areas — target, competitor, and two distractors — were analyzed. Altogether, 29.7 % of the data were excluded. Although 14 participants with few or no fixation recordings were excluded before, non-usable data for the other participants still ranged between approximately 20–40 %. The interest areas each had a cell size of 848×484 pixels (472x344 in Experiment 5) with a distance of 56 pixels (40 in Experiment 5) between vertical cells and 112 pixels (60 in Experiment 5) between horizontal cells.

Then, the fixations for the four interest areas were analyzed in 30-ms steps in a time window from 0 to 1,000 ms after target word onset. The dependent variable *target* indicated whether a participant fixated the target; *competitor* indicated a competitor fixation and *distractor* a distractor fixation. This resulted in three variables with binary values. These data were then compared with the *Eyelink* data from the original study in Experiment 5. Using the software R (R development core team, 2017), the *Eyelink* and the *EyeTribe* data were first inspected visually in a plot of fixation proportions, and then analyzed statistically.

Filler analysis. In the communication between *PsychoPy* and the computer sound card, the delay in sound presentation differs from that observed with *Experiment Builder* (SR Research Ltd.) which was used in Experiment 5 together with the *Eyelink 1000* and the respective computer sound card. As a consequence, the onset of the auditory sentence in a trial varied somewhat between experiments. As it is difficult to determine the exact delay, the filler trials were compared between the two experiments. Visual inspection of the fixation plot of fillers suggested that the fixation pattern was similar for both data sets, but in the *EyeTribe* data, target-competitor disambiguation was situated about half a second later than in the *Eyelink* data (*Eyelink*: 150–400 ms; *EyeTribe*: 650–900 ms).

The exact values of the delay were determined with a time point in the disambiguation area of each experiment (*Eyelink* versus *EyeTribe*) that fulfilled the following conditions: First, the mean proportion of target fixations by the control and the listening group were between 25–30 %. This value was observed directly before target-competitor disambiguation in all groups. Second, immediately preceding target-competitor disambiguation, the fixation proportions by the control and listening groups were quite close to one another.

As such, the respective time points were supposed to display the minimal distance between the fixation proportions of the control and listening group target, respectively. The analysis revealed this point to be at 160 ms (*Eyelink*) and 668 ms (*EyeTribe*), a difference of 508 ms. When accounting for this difference, that is, plotting eye movements at the actual onset of the sound presentation for each experiment, the *EyeTribe* data were shifted to the left by 508 ms, resulting in an *EyeTribe* filler data pattern that approximately corresponded to the *Eyelink* data (Figure 3-6).

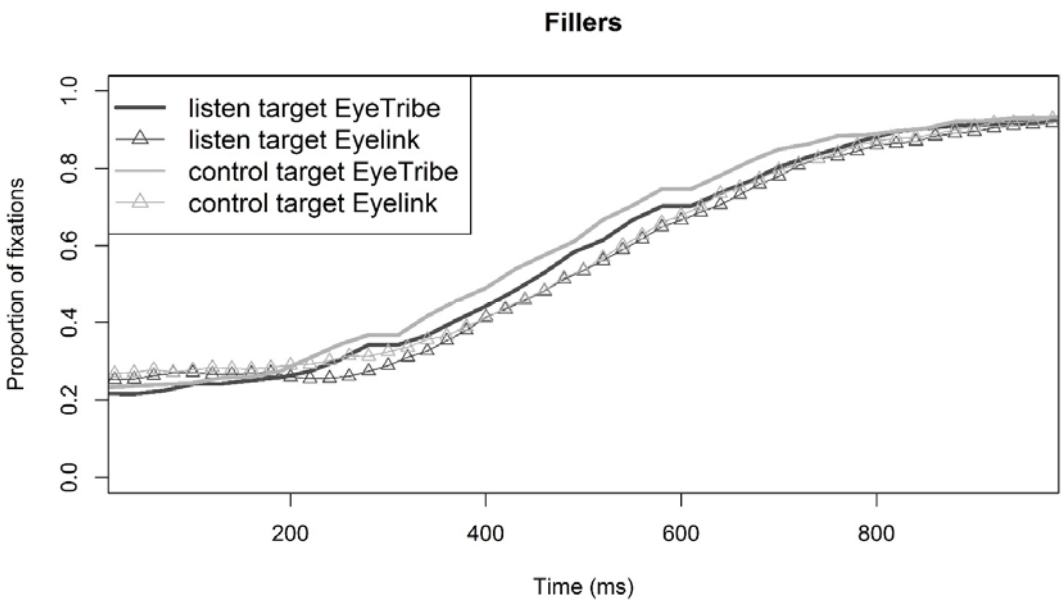

Figure 3-6. *Target fixations of fillers by the listening and the control group recorded with the EyeTribe (Experiment 7) and the Eyelink 1000 (Experiment 5).*

Accented word analysis. The accented data from both the *Eyelink* and *EyeTribe* experiments were then plotted with the same method as the fillers (Figure 3-7). Target fixation proportions as calculated with the empirical logit function, from the two trackers were then analyzed. The same time window as in Experiment 5 (250–750 ms) was analyzed with linear mixed effects regression models (Baayen, Davidson, & Bates, 2008; Bates, Kliegl, Vasishth, & Baayen, 2015). Significance of factors was indicated by p-values that were determined with likelihood ratio tests.

The final model had *experimental group* (listening group versus control group) as a fixed factor, and *participant* and *item* as random intercepts. Proportion of target fixations was the dependent variable. The listening group fixated the target significantly more often than the control group ($\chi^2 = 10.5$, $p < .002$). However, this advantage can also be observed between 0–200 ms, which is the time that a programmed eye movement needs to be launched (Altmann & Kamide, 2004). This difference was accounted for by adding the mean listening advantage in the time window 0–200 ms to the control data points from the critical time window (i.e., shifting the control data to a higher level, as was done in Experiment 5). The advantage shown by the listening group was no longer significant (χ^2 = 0.6, p = .45). The data were further inspected in a direct comparison with the *Eyelink* data from Chapter 3A.

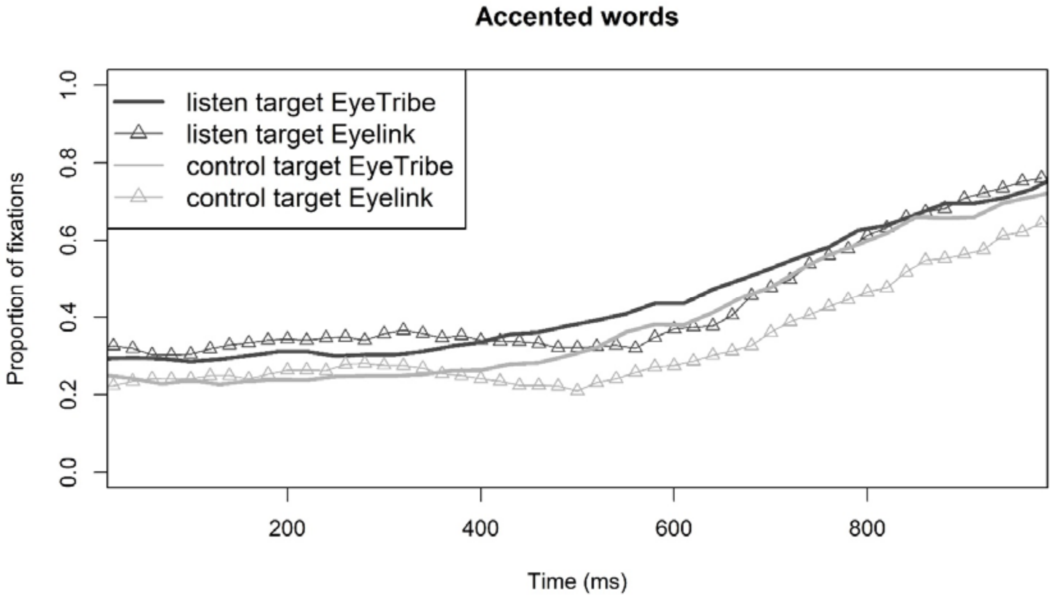

Figure 3-7. *Target fixations of accented words by the listening and the control group recorded with the EyeTribe (Experiment 7) and the Eyelink 1000 (Experiment 5).*

Target fixations between 250–750 ms were compared between the two experiments (*Eyelink* data from Experiment 5 versus *EyeTribe*) for each experimental group in a separate model. In both models, *participant* and *item* were random intercepts. In the listening group, the fixed factor *eye-tracker* (*Eyelink* versus *EyeTribe*) was not significant (χ^2 = 1.2, p = .27). In the control group, *eye-tracker* also did not significantly improve the model (χ^2 = 0.7, p = .41). This means that there was no difference in target fixations by the two listening training groups using either eye-tracker (*Eyelink* versus *EyeTribe*).

Discussion

The replication of Experiment 5 described in Chapter 3A with the *EyeTribe* shows that, particularly for paradigms where the research question does not require precise saccadic analysis or a higher sampling rate for more precise timing, the *EyeTribe* can be used as an alternative to high-end eye-trackers. The present results do not significantly differ from the *Eyelink* results in Experiment 5. In the visual world paradigm, the *EyeTribe* can provide reliable results, even at its lowest resolution of 30 Hz and when participants wear glasses or contacts. Despite some small differences between Experiment 5 and the present Experiment 7, the general pattern was still replicated. For example, in the present study, a larger screen was used than in Experiment 5. This implies longer distances between the single words printed on the screen. Moreover, the font size and type varied slightly in this experiment compared to Experiment 5.

The *EyeTribe*, however comes with several disadvantages. First, it is not accompanied by pre-installed software or any sort of tech support in case of problems. The observation that the group difference is not as strongly pronounced as in Experiment 5 (i.e., equally strong from 0–200 ms as in the subsequent longer time window), likely lies in the longer display preview that resulted from the delayed target word presentation in *PsychoPy*. For future reference, it is important to consider methods for reducing the sound replay latency in *PsychoPy*.

For example, it is suggested choosing the audio library *Pyo* instead of the *PsychoPy's* default *Pygame* (Landaulab, 2017). Moreover, the audio driver should be set to ASIO, leading to greater timing reliability, and time should be set in frames, which is more accurate than seconds. A second disadvantage is that the *PsychoPy–EyeTribe* combination does not offer the option of drift correction during an experiment. Drift correction warrants a consistent degree of accuracy throughout an experimental session. Due to paradigm-specific display properties, accuracy does not play an important role in visual world studies, however.

Thirdly, it was not always transparent what caused good or bad calibration results. Prior testing experience (e.g., in Chapter 3A), has shown that aspects like glasses or dark mascara can impair calibration with the *Eyelink 1000*. Those aspects, however, were not clearly identified as calibration issues when using the *EyeTribe*. Specifically, a number of participants with glasses varied between very good and very bad calibration results. Four more participants had to be excluded from the present data because no fixation coordinates were recorded at all, and for an additional 10 participants, coordinates were missing for more than 50 % of the data, despite successful calibration.

Conclusion

In conclusion, the *EyeTribe* provides acceptable results for visual world eye-tracking experiments, even with a sampling rate as low as 30 Hz. However, the device does not come with a pre-installed software and does not offer the option of drift correction during an experimental session, and a large amount of the collected data could not be included in the analysis because the eye-tracker did not always record fixation coordinates.

Chapter 4

Memory advantages for produced words with a familiar or unfamiliar accent

Abstract

Individuals' own overt word productions facilitate word memory more than listening to others producing the words does (MacLeod, 2011). This effect is not demonstrated as clearly, however, with accented speech or when production is not preceded by acoustic word presentation (i.e., no imitation). The present study tested whether native accent familiarity modulates the advantage of production without imitation. Words were produced or listened to in a familiar or an unfamiliar accent during a training session that was followed by a visual memory recognition task. D-prime scores were significantly higher for produced words than listened-to words in both the familiar and the unfamiliar accent condition. The same pattern was observed when participants did not hear themselves during production. When word memory was tested one week after the training, more produced words were still recognized. These results demonstrate the robustness of the advantage of production alone across accents and time.

Introduction

When trying to keep a list of words in memory, for example a shopping list or a list of names (as for example teachers frequently must), a common strategy is to read the words out aloud. This should indeed be helpful in remembering as prior research has repeatedly found that producing a word aloud helps to remember it more than just reading or hearing it (e.g., MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010), even when the word has to be recalled a week later (e.g., Ozubko et al., 2012). But does this production advantage arise because a speaker uses their own familiar accent, or is the advantage more general in nature, applying regardless of familiarity of the accent?

It has been shown that recognizing a word is easier when the word is spoken with a familiar rather than unfamiliar accent (Clopper, Tamati, & Pierrehumbert, 2016). A comparable effect for remembering words in a familiar accent would suggest that word memory is also modulated by previous linguistic experience. If the production advantage is furthermore found to be stable over time, the strengthening effect of production must be long lasting. The present memory study addresses these issues by testing German participants' memory for German words that were produced in either a familiar or unfamiliar native accent. Participants either produced the words or listened to them, and recognition memory was tested either immediately afterwards or with a one week delay.

How production influences memory

In the memory literature, the advantage of having produced a word aloud oneself is called the *production effect*. In a typical experiment (e.g., MacLeod et al., 2010), participants study a list of words before they take part in a memory recognition task. During the study phase, participants read some words aloud, and other words are read silently. In the following memory task, participants decide with a button press whether the word printed on the screen had been included in the study phase or not. Words that had been read aloud by the participants are more likely to be correctly remembered than words that had been read silently. Even when the words were not read aloud but just mouthed silently without vocalization, memory is better for silently mouthed words than for entirely silent words. This advantage of speech production is also present when memory for self-produced words is compared with memory for words that a second person spoke (MacLeod, 2011).

The production effect has been found to be stable over time. When the memory test was administered with a week delay, recognition memory was still better for words that had been read aloud during the study phase than for words that had been read silently (Ozubko, Hourihan, & MacLeod, 2012), suggesting that word production must strengthen its memory trace. The theoretical explanation for the production effect is based on the concept of distinctiveness (MacLeod et al., 2010). Produced words are more easily recalled because they stand out and are therefore more distinctive (e.g., Dodson & Schacter, 2001), and distinctive words are more salient than others during encoding –– provoking additional processing and, therefore, better memory (for a review: McDaniel & Geraci, 2006).

Traditionally, a within-participants design in which participants alternate between producing and reading words silently had been considered a prerequisite for the production effect to occur (MacLeod et al., 2010). Recently, however, a meta-analysis by Fawcett (2013) showed that the production effect can also be observed in a between-participants design where some participants only produced words while others only read silently. The production effect is, however, stronger in within-participants designs than in between-participants designs, which Fawcett (2013) explains as the result of retrieval processes being particularly strong in within-participants designs due to primary distinctiveness, while encoding and retrieval processes play a comparable role in within- and between-participants designs due to secondary distinctiveness. This view is in line with proceduralist assumptions where the advantage for produced words emerges during the recognition, i.e., the retrieval stage (Kolers, 1973; for areview: Roediger, Gallo, & Geraci, 2002). Produced words are assumed to have an additional source of discrimination, that is, the information that they had been produced. In a memory recognition task, this information is accessed, and the advantage for produced words is therefore assumed to emerge during the recognition stage, and not during encoding.

How accent influences memory and comprehension

In the studies described so far, the exact realization of the produced or listened to words was not under examination. Pronunciation, however, varies considerably between speakers and their accents, and listeners are often more familiar with some of this variation than with other forms of variation. Part of the production effect may possibly lie in the fact that the speakers are more familiar with their own accent than with that of the speakers they are listening to. Being familiar with an accent can certainly aid the understanding of words spoken in this accent. For example, listeners with extensive experience with the New York City (NYC) English accent show greater priming effects for words with the NYC-English-typical final r-dropping (e.g., *baker* pronounced as */beɪkə/*) than listeners with limited experience with the NYC accent (Sumner & Samuel, 2009).

Similarly, Adank, Hagoort, and Bekkering (2009) observed that only listeners who were familiar with both Standard Southern British English (SSBE) and Glaswegian English (GE) showed equal performance on both accent types in a sentence verification task. Processing advantages for participants' own accents over an unfamiliar accent were also found for French listeners (Floccia, Goslin, Girard, & Konopczynski, 2006). In a lexical decision task, reaction times to items in long sentences were faster when sentences were presented in the participants' own accent (Northeastern French) than when they were presented in an unfamiliar southern French accent.

While the above studies show a facilitatory effect of familiarity on the comprehension of accented words, a few studies have also investigated the role of production training on subsequent comprehension of accented speech. Adank, Hagoort, and Bekkering (2010), for example, found that accent imitation improves subsequent accent comprehension. Using an artificial accent, Dutch participants first completed a sentence transcription task before being trained on the artificial accent either by simply listening to accented sentences, by listening and then repeating the sentences in their own accent, or by listening and then imitating the sentences in the artificial accent. Training was followed by a re-test that measured transcription accuracy, and it was found that transcription performance improved most strongly between pre-test and test for participants who had to imitate the accent during training. Thus, imitation training facilitated subsequent comprehension of an accent more than listening training did.

Imitation by definition involves listening to a prompt before repeating it. When comparing the effects of imitation versus listening, it is implied that participants hear the input twice during imitation training, first from the prompt and then from their own productions, but only once during listening training. Whether production training facilitates accent comprehension without this double input was investigated in Chapter 2A, a word recognition study in which reaction times to lexical decisions were measured. In that study, at least for second language (L2) learners, producing a new accent for a few minutes facilitated subsequent accent comprehension more than listening to the accent did. Crucially, accented words were shown on the screen during the study phase, and participants either had to read them out loud or listened to a pre-recorded speaker producing the accented words. Thus, a production advantage was found in the absence of double listening input during imitation.

How much memory, rather than comprehension, is affected by accent has been investigated in only a few studies to date. Notably, Clopper, Tamati, and Pierrehumbert (2016) found an advantage for a familiar over an unfamiliar accent. Their participants were raised with a Midland accent (close to standard American English) and lived in a Midland-accent area at the time of testing. The experimental task was first to transcribe single words spoken in the Midland standard accent or a northern accent (with dialect-specific vowel shifts; cf. Labov, 1998). Then, a memory recognition task on old words from the transcription task and new words was completed. Words were either presented in the same accent as before or the other accent (Midland versus northern accent). In the memory task, words in the Midland standard accent were recognized more quickly than words in a northern accent, implying an advantage for the familiar standard accent. Moreover, with a second group of participants who grew up with the northern accent but now resided in the Midland-accent area, there was still an advantage for the Midland standard targets. Thus, the standard Midland accent was easier to process, even for listeners who had grown up with a northern accent.

Cho and Feldman (2013) recently combined the memory advantage for familiar accents with that of production training. American English participants listened to single English words while a visual cue simultaneously instructed them to either only listen or to repeat the words aloud. Half of the words that were played to the participants were produced with a global Dutch accent that was unfamiliar to them, and half were produced by a native speaker of American English, thus with a familiar standard accent. In their own productions, American English participants were instructed to imitate the unfamiliar Dutch accent in trials with a Dutch-accented auditory prompt. In a subsequent memory recognition task, more imitation items were remembered than listening items; in the listening condition, there was an additional memory advantage for Dutch-accented words over words produced by the American speaker. However, as emphasized in Chapter 2A, this advantage was not observed in the imitation condition.

Present study

In the present study, a memory advantage for German words with a familiar accent is compared to a memory advantage for German words with an unfamiliar accent; German participants were required to either produce words in the two accents or to listen to words in the two accents before finally testing their word memory. Rather than comparing a standard variety to a regional accent as Cho and Feldman (2013) and Clopper and colleagues (2016) did, two regional accents are being compared. By juxtaposing these regional accents, the role of accent familiarity for word memory can be determined when neither of the tested accents represents the standard variety in the participants' native language.

Both regional accents in the present study are marked with a specific segmental accent marker that deviates to a comparable extent from Standard German. These specific accent markers allow participants to produce the accents easily even when the accent is unfamiliar. Specific rather than global accent markers are used to ensure that possible production effects are not being attenuated by difficulties in reliably producing the accent. In order to avoid double listening input in the imitation task, all participants saw the words on the screen during the study phase in canonical orthographic transcription and either had to listen to another speaker producing the words or had to produce the words themselves. The robustness of a memory advantage for produced words is further scrutinized in two experiments testing its stability over time (e.g., Ozubko et al., 2012) and its stability in the absence of self-listening during production (e.g., MacLeod et al., 2010).

In the present study, a southern German accent is being compared to a northern German accent. Although both accents exhibit various typical deviations from Standard German, they both have a peculiarity involving the production of fricative stop sequences which is highly noticeable and usually directly associated with the two accents in question. In Standard German, the orthographic sequence ‘st’ is pronounced as /st/ when it occurs across syllables or in syllable final position, but /ʃt/ when it occurs in syllable initial position. Thus, a word like *Zahnbürste ‘toothbrush’*, is pronounced as *Zahnbür/st/e* (across syllables) and *Obstmesser ‘fruit knife’*, is pronounced as *Ob/st/messer* (syllable final position). However, *Blumenstrauß ‘bouquet’*, is pronounced as *Blumen/ʃt/rauß*.

Rather than pronouncing *Zahnbürste* canonically with /st/, speakers of the Swabian accent, spoken in southern Germany, regularly pronounce *Zahnbürste* with /ʃt/ (Vogt, 1977). This segmental substitution does not occur in the northern German accent. In the northern accent, however, the canonical /ʃt/ in syllable initial position as in *Blumen/ʃt/rauß* is regularly pronounced as /st/ (Schirmunski, 1962). Again, this is a segmental substitution that speakers of the Swabian accent would not make. All participants in the present study grew up with the Swabian accent and are currently enrolled as students at the University of Tübingen (in southern Germany). They are all familiar with the Swabian accent, but not familiar with the northern German accent.

In Experiment 8, words with the familiar accent marker and with the unfamiliar accent marker were produced or listened to in a training session. During training, single words were presented one after the other on a computer screen, and half of the words were read aloud by the participants and half of the words they listened to. After the training, a free recall task followed by a visual memory recognition task was administered. In the visual memory recognition task, participants decided with a button press if a word was presented during training or not. Experiments 9 and 10 specify the nature of a possible memory advantage for produced words in accented speech. Experiment 9 tested how long lasting the advantage is — to this end, participants performed the memory tests one week after the training session. Whether the memory advantage for produced words can also be observed when participants cannot hear their own productions was tested in Experiment 10. To test this, participants in the training phase in Experiment 10 wore headphones over which white noise was played to mask their own productions.

If familiarity with an accent facilitates memory, in line with Clopper and colleagues (2016), words with the familiar Swabian accent are expected to be recalled and recognized more easily than words with the unfamiliar northern German accent. If production has a further effect, self-produced words will be recalled more easily than words that have only been listened-to. Such a memory advantage for self-produced words in the present study would confirm that the advantage does not depend on imitation, but also applies without the double input. It is possible that the simultaneous influence of production and familiarity are additive effects, in which case there will be no interaction, i.e., the production advantage will be comparable for familiar and unfamiliar accents. If the memory advantage for self-produced words is directly prone to familiarity, the advantage will differ for the familiar Swabian accent and for the unfamiliar northern accent. The robustness of the memory advantages will furthermore be attested if the advantages can be observed with a week delay and in the absence of self-listening during production.

Experiment 8

Participants

Forty native speakers of German (33 female, 7 male), all students at the University of Tübingen (19–31 years old, mean age = 24.4, SD = 2.9), participated for monetary reimbursement. None suffered from any hearing disorders, and all had intact or corrected vision. All participants grew up with the Swabian accent and frequently used it in their own productions, as attested by self-ratings on their estimated frequency of daily usage (on a scale from 0 to 7, where 7 means "very often", on average 5.5, SD = 1.5, with their family; 4.1, SD = 1.9, with their friends; see Table 4-1). They were also familiar with Standard German, particularly in educational contexts such as school and university (daily usage of Swabian in university context was rated 2.51, SD = 1.53).

Table 4-1. *Participants' mean self-evaluations and standard deviations in using the Swabian accent in Experiments 8, 9, and 10.*

Frequency of listening to Swabian from 0 (never) to 7 (always)	*Frequency of speaking Swabian with family from 0 (never) to 7 (always)*	*Frequency of speaking Swabian with friends from 0 (never) to 7 (always)*	*Frequency of speaking Swabian in University from 0 (never) to 7 (always)*	*Performance of speaking Swabian from 0 (very bad) to 7 (excellent)*	*Performance understanding Swabian from 0 (very bad) to 7 (excellent)*
Experiment 8 (n=40)					
5.38 SD=1.58	5.53 SD=1.52	4.08 SD=1.86	2.51 SD=1.53	5.20 SD=1.45	6.18 SD=0.89
Experiment 9 (n=35)					
5.11 SD=1.69	5.91 SD=1.66	4.20 SD=1.79	2.14 SD=1.33	5.97 SD=1.18	6.11 SD=1.12
Experiment 10 (n= 40)					
5.43 SD=1.39	5.54 SD=1.88	4.28 SD=1.69	2.45 SD=1.32	5.33 SD=1.46	6.30 SD=0.84

Material

In total, 88 common German compound words (56 experimental words and 32 unrelated filler words; see Appendix C, Table C1) were used in the study. Half of the words were bisyllabic, and half were trisyllabic. The experimental words all contained a sibilant-alveolar stop sequence. In half of the words, the sequence corresponded to /st/ in standard pronunciation (*Zahnbür/st/e* '*tooth brush*'), and in half it corresponded to /ʃt/ in standard pronunciation (*Blumen/ʃt/rauß* '*bouquet*'). Compound words were used to ensure that the sibilant-stop sequence occurred in a comparable word position in both cases. The familiar Swabian accent was imposed on these words by replacing the /st/ in words like *Zahnbür/st/e* with /ʃt/. The unfamiliar northern accent was imposed by replacing the /ʃt/ in words like *Blumen/ʃt/rauß* with /st/.

Unrelated filler words did not include any sibilant-stop sequences and were matched for mean logged frequencies according to the CELEX word form dictionary (Baayen, Piepenbrock, & Gulikers, 1995) with the experimental words (mean accented: 0.43, SD = 0.44; mean fillers: 0.43, SD = 0.37). Moreover, they were matched in terms of number of letters of each word (mean accented = 9.6, SD = 1.4; mean fillers = 8.9, SD = 1.4), and the neighborhood density of each word was never greater than 1 according to the German CLEARPOND (Marian, Bartolotti, Chabal, & Shook, 2012).

Pre-Test

Before the actual experiment, a memory pre-test with all experimental tokens was run to preclude that the word forms themselves (due to word inherent differences) of one accent group, and not the accent itself, elicit better recall. A group of students (n = 19; 15 female; all native speakers of German, studying English at the University of Tübingen, mean age = 20.9, SD = 1.7) was asked to complete a free recall task on either the words used for the Swabian accent training (9 participants), or the words used for the northern German accent training (10 participants), each including 44 tokens (28 experimental words used for the accent and 16 fillers).

The compound words used for the experiment were presented auditorily in their canonical pronunciation. Word presentation was blocked (two blocks à 22 tokens), and a free recall task was administered after each block. Results showed a mean recall rate of 35.3 % of the Swabian accent tokens and 35.4 % of the northern accent tokens. A linear mixed effects model (see section *Data analysis* for methodological details) with the dependent variable *mean recall rate*, the fixed effect *accent group* (Swabian versus northern) and *test block position* (first versus second), as well as *participant* as random intercept revealed no significant difference in recall rate between the words used for the Swabian accent condition and those used for the northern German accent ($\chi^2 = 0$; $p = .99$). This supports the use of the suggested test words in the following experiments.

Design

Training. Fourteen of the experimental words were included in a training block with the familiar Swabian accent, and 14 of the experimental words were included in a training block with the unfamiliar northern accent; eight unrelated filler words were added to each training block. In each training block, half of the words were assigned to the self-production condition, and half were assigned to the listening condition. Participants were tested in pairs, and each pair was presented with both training blocks, counterbalanced for order, and given their own pseudo-randomized word order in each block (accounting for list order effects, cf. Forrin & MacLeod, 2016b), based on the following restrictions: accented experimental words never appeared more than three times in a row, and participants never had to produce more than two words in a row.

Memory Recognition test. Both training blocks were paired with a test list for memory recognition. The test list contained the 14 experimental words from the training (old words), 8 filler words from the training, and 22 filler words that did not occur during the training (of which 14 included a *st*-sequence). The order of the test list was pseudo-randomized for each participant pair.

Procedure

The experiment was programmed with *Presentation* (version 18.3, www.neurobs.com). It made use of a training-test paradigm that was applied to both experimental blocks (familiar versus unfamiliar accent block). During training, participants sat in pairs of two in the laboratory opposite one another with laptops in front of them. They were both provided with written instructions, and the experimenter additionally ensured that they understood how to produce the accent marker in question. The accent markers were not explicitly introduced as “Swabian” or “northern German”. During training, single words were presented one after the other on both computer screens simultaneously for 2,000 ms with an inter-trial interval of 1,000 ms.

Half of the words were read aloud by the first participant and listened to by the second; the other half were read aloud by the second participant and listened to by the first. A microphone (Samson Meteor) was placed in front of the participants so that both participants’ productions could be recorded during training. If a word was supposed to be produced aloud, it was printed in blue letters on the screen of the participant, if a word was supposed to be listened to it was printed in black letters. All words during training where shown in their correct orthographic transcription, and orthographic *st* sequences in the experimental words were always printed in bright pink letters (see Appendix C, Figure C1). For words that were supposed to be listened to, participants were instructed not to move their lips while listening.

After each block, first a free recall memory task and then a visual memory recognition task was administered. During free recall, participants had two minutes to recall and write down on a sheet of paper as many words from the training as possible. In the memory recognition task, single words from the training phase (old experimental words and fillers) and new words were printed in dark green on the screen for 2,000 ms each. Each word presentation was preceded by a fixation cross for 500 ms and followed by a 500 ms inter-trial interval. Participants had to decide with a button press if a word was old or new. Right-handed participants clicked on the right button for yes, while left-handed participants used the left button for yes responses. Participants were given as much time as they needed for their response. They were instructed to answer as accurately as possible. Hit rates were measured. The experiment concluded with a language background questionnaire that included self-reports on participants' Swabian accent use and knowledge. Overall, the experiment took about 25 minutes.

Data Analysis

Data were first inspected descriptively and then analyzed with linear mixed effects models (Baayen, Davidson, & Bates, 2008; Bates, Kliegl, Vasishth, & Baayen, 2015; Jaeger, 2008). Analysis of the free recall data was based on the number of correctly recalled words, regardless of their position during training. The analysis of memory recognition was based on hit rates, i.e., correctly detected old words, and false alarms, i.e., erroneous yes answers to new words. Based on this information, d-prime scores were calculated and inspected descriptively (see signal detection theory: Stanislaw & Todorov, 1999). D-primes were calculated using the AnotA function in R (R development core team, 2017) from the sensR package (Brockhoff & Christensen, 2010; Christensen & Brockhoff, 2016).

Linear mixed effects regression models were used for further statistical analysis. For each analysis an individual, best fitting model was built that included a particular choice of fixed and random factors. Random effect structure included random intercepts for *participants* and *items* as well as those random slopes that significantly improved model fit as tested by likelihood ratio tests (by means of the anova-function). Significance of factors was indicated by t-values > |2|. Corresponding p-values, as reported in the text below, were determined with likelihood ratio tests, using the anova-function. As fixed effects, *training* (self-production versus listening), *accent* (familiar accent versus unfamiliar accent), and *block position* (first training block versus second training block) were considered. Block position was included because participants might have anticipated the nature of the upcoming test phase prior to the second test block as a result of the test experience in the first block. Participants were not explicitly informed about the nature of the upcoming tasks before the training of the first block; they were only informed that there would be two training phases, each followed by “several memory tasks”. In the free recall task, *recall accuracy* was the dependent variable, and in the memory recognition task it was the *d-prime score*.

Results

During training, participants produced the instructed accent markers quite well in both blocks. The experimenter decided on the basis of perceptual judgements whether the respective sound change was produced in the experimental training tokens as communicated in the instructions. This was done already during the training phase and verified later by checking the recordings that were made during the training phase. Every instance where the substitution was not clearly perceivable was documented. On average, only 0.15 (SD = 0.4) out of 7 experimental Swabian trials (~2.1 %) and 0.3 (SD = 0.6) out of 7 critical northern trials (~4.3 %) were not pronounced as instructed.

Free recall. Overall, 19.0 % (SD = 0.4) of all training words (including unrelated fillers) and 20.4 % (SD = 0.4) of all accented training words were correctly recalled. Table 4-2 shows the proportion of correctly recalled words for self-produced and listened-to words for both familiar and unfamiliar accents. Descriptively, more self-produced words were recalled than listened-to words in the familiar accent condition (+5 % for self-produced words) as well as in the unfamiliar accent condition (+10 % for self-produced words). Furthermore, there was a numeric advantage for the familiar accent compared to the unfamiliar accent for listened-to words (+5.4 %) but no such advantage for self-produced words (+0.4 % for familiar accent).

Next, the number of correctly recalled words was analyzed with generalized linear mixed effects models. The final model included *training* as fixed factor, and *participant* and *item* as random intercepts. *Accent familiarity* was not included because it did not significantly improve the model. Significantly more self-produced words were recalled than words that had been listened to (χ^2 = 11.0; p < .001). Thus, a memory advantage for self-produced words over listened-to words was found in the free-recall task, but accent familiarity did not improve recall.

Table 4-2. *Mean proportions and standard deviations of correctly listed training words in the free recall task in Experiments 8 and 10.*

	Experiment 8 (n=40)		*Experiment 10 (n=40)*	
	Production training	Listening training	Production training	Listening training
familiar Swabian accent	.243 SD=0.430	.193 SD= 0.395	.261 SD= 0.440	.146 SD= 0.354
unfamiliar northern accent	.239 SD= 0.427	.139 SD= 0.346	.257 SD= 0.438	.164 SD= 0.371
both accents	.241 SD= 0.428	.166 SD= 0.372	0.259 SD=0.438	.155 SD= 0.363

Memory recognition. In the memory recognition task, the mean overall accuracy (i.e., the accuracy of accepting old words and rejecting new words) across all participants was 82.8 % (SD = 0.38). Experimental words and fillers from the study phase were on average remembered with 81.0 % (SD = 0.39) accuracy (experimental words: 84.6 %, SD = 0.36; fillers: 74.7 %, SD = 0.44). Table 4-3 and Figure 4-1 furthermore show that hit rates were higher for words that had been presented in the familiar Swabian accent during the study phase (86.6 %) than for words that had been presented in the unfamiliar northern accent (82.5 %). Also, hit rates were higher in both accent conditions for self-produced words (90.2 %) than for listened-to words (78.9 %). The d-prime results support this pattern, with higher d-prime scores suggesting greater memory discriminability.

Table 4-3. *Mean proportions of "old" responses (hits) in the memory recognition task for words from the production training and listening training in both accent conditions. Standard deviations and d-prime scores are listed below their corresponding means.*

	Experiment 8			*Experiment 9 (test after 1 week)*			*Experiment 10 (no self-listening)*		
	productio n training	listening training	both trainings	productio n training	listening training	both trainings	productio n training	listening training	both trainings
familiar Swabian accent	.92	.81	.87	.74	.62	.68	.90	.77	.84
	SD=0.28	SD=0.39	SD=0.34	SD=0.44	SD=0.49	SD=0.47	SD=0.30	SD=0.42	SD=0.37
	d'=2.42	d'=1.92	d'=2.15	d'=1.04	d'=0.71	d'=0.87	d'= 2.39	d'= 1.83	d'= 2.07
unfamiliar northern accent	.89	.76	.83	.71	.66	.69	.85	.75	.80
	SD=0.32	SD=0.43	SD=0.38	SD=0.45	SD=0.48	SD=0.46	SD=0.36	SD=0432	SD=0.40
	d'=2.23	d'=1.74	d'=1.95	d'=0.96	d'=0.80	d'=0.88	d'=2.12	d'=1.77	d'=1.93
both accents	.90	.79	.85	.73	.64	.69	.88	.76	.82
	SD=0.30	SD=0.41	SD=0.36	SD=.45	SD=.48	SD=.47	SD=0.33	SD=0.43	SD=0.39
	d'=2.32	d'=1.83	d'=2.04	d'=1.00	d'=0.75	d'=0.87	d'= 2.24	d'= 1.80	d'= 2.00
fillers (no accent)	.81	0.68	.75	.568	.42	.49	.81	.67	.74
	SD=0.39	SD=0.47	SD=0.44	SD=0.50	SD=0.49	SD=0.50	SD=0.39	SD=0.47	SD=0.44
	d'=1.90	d'=1.50	d'=1.69	d'=0.56	d'=0.19	d'=0.37	d'= 1.97	d'= 1.51	d'= 1.72

Further statistical analyses of d-primes were conducted with generalized linear mixed effects models. The final model included *training* and *accent* as fixed factors as well as *participant* and *item* as random intercepts. D-primes were significantly higher for self-produced words than for listened-to words (χ^2 = 23.3; p < .001). Moreover, words with the familiar Swabian accent were remembered more often than words with the unfamiliar northern accent (χ^2 = 4.1; p = .04). *Training* (self-production versus listening) and *accent familiarity* did not interact.

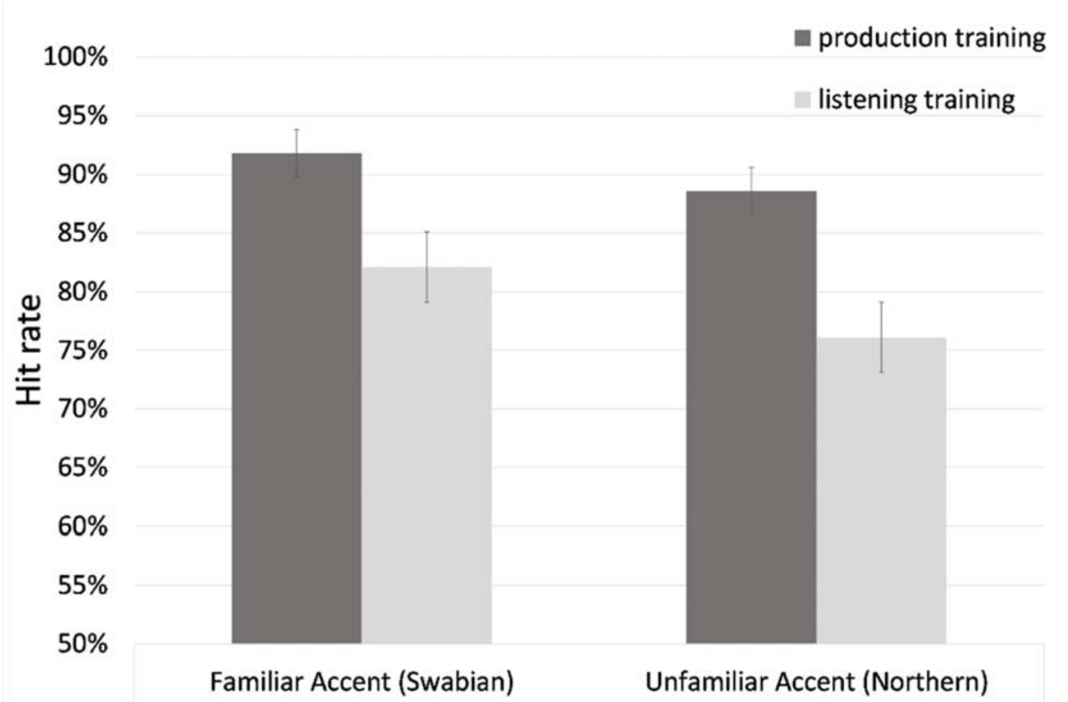

Figure 4-1. *Hit rates in the memory recognition task by accent and training condition for Experiment 8. Whiskers represent standard errors.*

Interim discussion

Experiment 8 provides further evidence for the production effect (MacLeod, 2011; MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010) in a task that does not involve imitation of a prompt. Words that were produced during training were recalled better at test than words that were listened to in both the free recall task and the memory recognition task. This production advantage occurred for words in the familiar and unfamiliar accents alike. The production effect is therefore not restricted to a familiar accent but is instead general in nature. Still, words in the familiar accent were remembered more readily in the recognition task than words in the unfamiliar accent. Thus, prior experience with an accent can facilitate memory performance. This advantage was not observed in the free recall task because this task elicits activation of only a fraction of all experimental words. Familiarity effects likely require more tokens (in contrast to the more robust production advantage that was also observed in the free recall task).

In summary, Experiment 8 shows that the production effect is robust for familiar and unfamiliar accents that both deviate from the standard form of a language. The results are in line with Clopper, Tamati, and Pierrehumbert (2016), who found a memory advantage for the familiar standard accent of American English over a less familiar northern accent. They seemingly contrast, however, with the results of Cho and Feldman (2013), who found that words spoken in an unfamiliar accent were remembered better than words in a familiar accent. A possible explanation for this difference lies in the fact that Cho and Feldman compared a foreign global accent that deviated on multiple cues from the target language and is therefore quite distinct with the standard variant of the target language. In the present study, both accents deviate from the standard to a similar degree. It is also likely that the imitation of the unfamiliar global Dutch accent by the American English participants in Cho and Feldman deviated somewhat from the Dutch imitation prompt, thereby highlighting a difference between the two productions that further enhanced the distinctiveness of the words.

In Experiment 9, the observed advantages for produced words and familiar accents found in Experiment 8 were tested for their stability. As in Experiment 8, participants had to produce or listen to words spoken in a familiar or an unfamiliar accent during the study phase; the memory testing was, however, delayed by one week.

Experiment 9

Prior research has shown that the memory advantage for produced words can be observed even a week after the study phase (Ozubko, Hourihan, & MacLeod, 2012).

Similar long-lasting effects were also shown for accent learning. For example, Witteman, Bardhan, Weber, and McQueen (2015) found that, with only a few minutes of listening training, single words with a Hebrew accent induced priming effects one week after the training. In Experiment 9, it was addressed if the memory advantage for produced and accented words is also stable across time; to this end, the memory test phase followed the study phase with a one-week delay.

Participants

Thirty-five native German students from the University of Tübingen participated for monetary reimbursement. The participants (21 female, 14 male) were 18–26 years old (mean age = 23.5; SD = 2.0), none of them suffered from any hearing disorders, and they all had intact or corrected vision. They grew up with the Swabian accent and frequently used it in their own production, as attested by self-ratings on their estimated daily usage (in the context of family: mean = 5.9, where 7 means "very often", SD = 1.7; in the context of friends: mean = 4.2, SD = 1.8). They were also familiar with Standard German, particularly in educational contexts such as school and university (daily usage of Swabian in university context was rated 2.14, SD = 1.33).

Material, design, and procedure

The same material and design as in Experiment 8 were used; however, the experiment was split into two sessions. As in Experiment 8, training was blocked for the familiar and unfamiliar accents, and in session one both training blocks were administered without an intervening memory test. There was a short break of 30 seconds between the two training blocks. Session two was scheduled one week later and included the memory test phase. Participants performed the same memory recognition task as in Experiment 8, once with the words from the first training block and once with the words from the second training block. The order was the same as during the training session. Again, there was a 30-second break between the two tasks. No free recall task was administered this time, because even without a delay recall rates in Experiment 8 were quite low (only around 20 %). The experiment concluded with the language background questionnaire from Experiment 8.

Results

The participants produced the instructed accents quite well. On average, only 0.1 (SD = 0.3) out of 7 experimental Swabian words (~1.3 %) and 0.3 (SD = 0.6) out of 7 experimental northern words (~3.8 %) were not pronounced as instructed. In the memory recognition task, the mean overall accuracy (i.e., the accuracy of accepting old and rejecting new words) across all participants was 63.4 %. Experimental words and fillers from the study phase were remembered with 61.5 % accuracy on average (experimental words: 68.5 %, SD = 0.46; fillers: 49.3 %, SD = 0.50).

Descriptively, hit rates for words in the familiar Swabian accent and for words in the unfamiliar northern accent were quite similar (68.4 % and 68.6 %, respectively), which was confirmed by comparable d-prime scores (see Table 4-3). Regarding training condition, mean hit rates were higher in both accent conditions for self-produced words (74.3 % for familiar accent and 71.4 % for unfamiliar accent) than for listened-to words (62.4 % for familiar accent and 65.7 % for unfamiliar accent) during training (see Table 4-3 and Figure 4-2).

The final model for the d-prime analysis included *training* as fixed effect, and *participant* and *item* as random intercepts. No other factors significantly improved the model (e.g., for accent familiarity $p = .6$). D-primes were significantly higher for self-produced words than for listened-to words ($\chi^2 = 6.8$; $p < .009$).

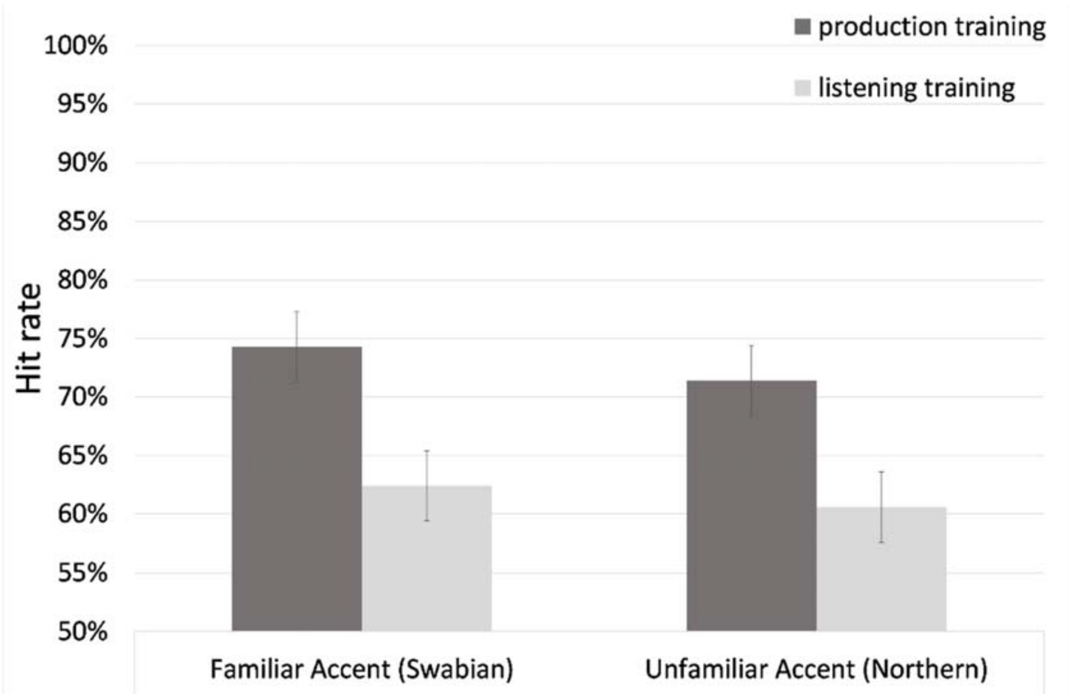

Figure 4-2. *Hit rates in the memory recognition task by accent and training condition for Experiment 9. Whiskers represent standard errors.*

Interim discussion

After one week, self-produced words in Experiment 9 were still more readily remembered than listened-to words. This finding expands upon prior findings on the long-lasting nature of the production effect (Ozubko et al., 2012). In Ozubko and colleagues' earlier study, a long-lasting memory advantage for self-produced words compared to words that were read silently was found. In the present study, a similar long-lasting memory advantage was found when comparing self-produced words with listened-to words. The advantage of accent familiarity found in Experiment 8, though, was no longer observed with a one-week delay. Words in the familiar Swabian accent were remembered just as well as words in the unfamiliar northern accent.

Thus, while the memory advantage for produced words was stable over time, the memory advantage for words in a familiar accent was seemingly not. The absence of an effect for familiarity could be a floor effect. Overall recognition accuracy decreased considerably between Experiments 8 and 9, making it more difficult to observe a difference between accent conditions. Note, however, that the production advantage also suffered from a decrease in overall accuracy, but was nevertheless stable over time. Alternatively, the results of Experiment 9 imply that there is a difference in strength and stability between the effect of production on memory and the effect of accent familiarity on memory.

If the strength of the production effect truly lies in the production itself or was thus far only endorsed by the chosen task was further investigated in Experiment 10. In Experiments 8 and 9, participants were visually presented with words that they either had to produce or listen to. With this design choice, the issue of double input in an imitation task with an auditory prompt was avoided. It can be argued, however, that the problem of double input nevertheless remains in production itself, since speakers hear themselves while they are producing speech. In Experiment 10, a memory advantage for produced words across accents was tested in a paradigm that precluded participants from hearing their own speech.

Experiment 10

The memory advantage for produced words in the previous experiments might be partially explained by the plurality of channels that producers use. The fact that during speech production, speakers hear themselves and also experience their articulatory movements at the same time results in a multi-modal input. During listening, on the other hand, listeners can only rely on the auditory channel. For speech without a specific accent, prior findings suggest that the plurality of channels does not play enough of a role to change the production effect. Mouthing silently has been shown to induce a memory benefit (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010), and when learning new words, producing the words and hearing oneself while doing so is still more efficient than listening twice to the to-be-learned words (Zamuner, Morin-Lessard, Strahm, & Page, 2016). Whether similar conclusions can be drawn for words in familiar and unfamiliar accents was tested in Experiment 10.

A distinct pattern for accented speech is conceivable if the experimental change modulates the degree of distinctiveness during self-production. The memory advantage for the familiar accent might stem primarily from listening or production. If listening is more important than production in terms of distinctiveness (after all, one does not often produce an unfamiliar accent), then in Experiment 10, the effect of accent familiarity might be smaller because the amount of listening in the production condition has been reduced. To control for this, participants wore headphones and were played white noise during self-production trials. Similar to mouthing, this precludes participants from hearing themselves while still allowing the articulators to be brought to the same positions as when speaking out loud. Playing white noise during production rather than mouthing silently was chosen because it allowed to check if the accents were produced correctly during the study phase, and the vocalizations were needed as listening input for the second participant.

Participants

Forty-five native German students from the University of Tübingen participated for monetary reimbursement. Five were excluded because they did not perform the sound changes during the training (n = 2), reported hearing problems (n = 2), or were not a speaker of Swabian according to the post-experimental questionnaire (n = 1). The remaining 40 participants (28 female, 12 male) were 19–30 years old (mean age = 24.2;

SD = 2.7), none of them suffered from any hearing disorders, and they all had intact or corrected vision.

They grew up with the Swabian accent and frequently used it in their own production, as attested by self-ratings on their estimated daily usage (in the context of family: mean = 5.5; SD = 1.9, where 7 means “very often”; in the context of friends: mean = 4.3; SD = 1.7). They were also familiar with Standard German, particularly in educational contexts such as school and university (daily usage of Swabian in university context was rated 2.45, SD = 1.32).

Material, design, and procedure

The material and design were identical to that of Experiment 8. However, in Experiment 10 participants wore headphones (Sennheiser HD 215 II) during the two training blocks. When participants were instructed to produce a word aloud, white noise was played over their headphones simultaneously. This ensured that air- and bone-conducted sound from the participants’ own voices was blocked, and they did not hear themselves during speaking. Furthermore, the recorded speech productions from the second participant were played directly over headphones during listening trials. During self-production trials, participants did not hear their own speech, but they did hear the second participant over headphones during listening trails.

Results

During the study phase, participants performed the instructed accents quite well. On average, only 0.1 (SD = 0.4) out of 7 critical Swabian trials (~1.4 %) and 0.3 (SD = 0.7) out of 7 critical northern trials (~3.9 %) were not pronounced as instructed.

Free recall. Overall, 19.8 % (SD = 0.4) of all training words (20.7 % of all accented training words; SD = 0.41) were correctly recalled. Table 4-2 shows that numerically, more self-produced words were recalled than listened-to words in both the familiar accent (+12 % for produced words) and in the unfamiliar accent (+9 % for produced words). The number of correctly recalled training words was analyzed with the same generalized linear mixed effects model as in Experiment 8.

The final model included *training* as fixed factor and *participant* and *item* as random intercepts. *Accent familiarity* was not included because it did not significantly improve the model. Significantly more self-produced words were recalled than listened-to words ($\chi^2 = 20.6$; $p < .001$). As in Experiment 8, a memory advantage for self-produced words over listened-to words was found in the free-recall task, but accent familiarity did not improve recall.

Memory recognition. The mean overall accuracy (i.e., the accuracy of accepting old words and rejecting new words) across all participants was 82.6 % (SD = 0.38). Experimental words and fillers from the study phase were remembered with 79.0 % (SD = 0.41) accuracy (experimental words: 82.0 %, SD = 0.39; fillers: 73.9 %, SD = 0.44) and the false alarm rate was 13.9 % (SD = 0.35). Descriptively, hit rates were higher for words in the familiar accent (83.7 %) than for words in the unfamiliar accent (80.2 %). The d-prime scores confirm this pattern. Table 4-3 and Figure 4-3 show that in terms of training condition, mean hit rates were higher for self-produced words than for listened-to words in both accent conditions.

The final statistical model included as fixed factors *training*, *test round*, and *accent*. Significantly higher d-primes were observed for self-produced words than for listened-to words ($\chi^2 = 17.7$; $p < .001$). The model also shows that participants recognized more words in the second test round ($\chi^2 = 4.4$; $p < .04$). *Accent familiarity* improved the model marginally ($\chi^2 = 2.3$; $p = .13$).

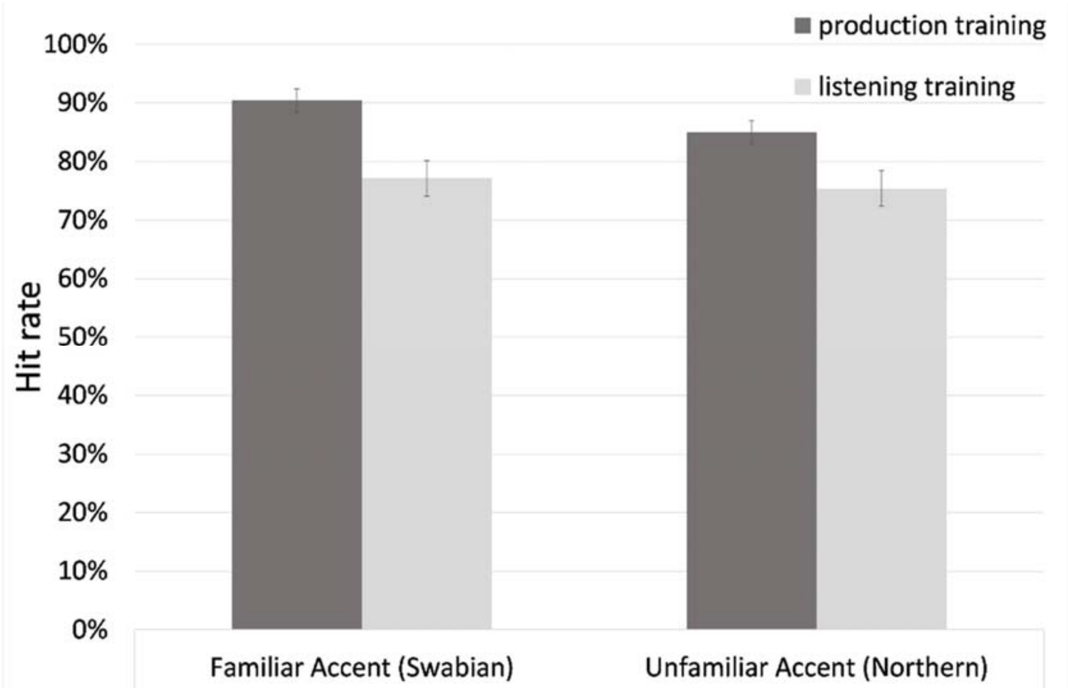

Figure 4-3. *Hit rates in the memory recognition task by accent and training condition for Experiment 10. Whiskers represent standard errors.*

Interim discussion

Even when participants could not hear themselves while they were speaking in Experiment 10, production improved word memory more than listening did in both the free recall task as well as the memory recognition task. This suggests that the motor movements of speech production are indeed driving the memory advantage for self-produced words.

The memory advantage for words in the familiar Swabian accent, however, was not as strongly pronounced in Experiment 10 as in Experiment 8, but the familiar Swabian words still tended to be recognized more easily than the unfamiliar northern words. This effect, however, does not show that for familiar accents, listening experience plays a stronger role, because within both training conditions, the familiarity advantage decreased. It is possible that the circumstances of the experimental session (no self-listening, hearing the second person only over headphones) allowed participants to rely less on the acoustic input but rather on the written words presented on the screen, resulting in a less strongly pronounced accent advantage.

Discussion

In three experiments, memory advantages for producing words with a familiar accent versus an unfamiliar accent were compared with the effects of listening to a second person produce the words. Experiment 8 provides evidence for a stable memory advantage for produced words with either accent as well as a memory advantage for words with the familiar accent in both learning conditions (self-production and listening). This finding is in line with the classical production effect (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010) and particularly with the finding that memory for self-produced words is better than memory for listened-to words that were produced by a second person in the same experimental session (MacLeod, 2011).

Furthermore, the production advantage was equally strong for words in the familiar Swabian accent as for words in the unfamiliar northern accent, but more words in the familiar accent were remembered than words in the unfamiliar accent overall. The fact that memory was better for words in the familiar accent is in line with recent comprehension studies in which lexical processing was facilitated by accent familiarity (e.g., Adank, Hagoort, and Bekkering, 2009; Floccia, Goslin, Girard, & Konopczynski, 2006; Sumner & Samuel, 2009).

Experiment 9 found that the production effect was still stable one week after the study phase, but the effect of accent familiarity was not. Experiment 10 again found a memory advantage for produced words even when self-listening during production was precluded. This is in line with results from a word transcription task where words produced by an average speaker were more accurately detected than words presented in the listener's own voice (Schuerman, Meyer, & McQueen, 2015). In Experiment 10, words in a familiar accent were also remembered better than words in an unfamiliar accent, albeit somewhat less strongly than in Experiment 8. Crucially, the effects of production and accent familiarity in Experiments 8 and 10 did not interact. Thus, two different, independent mechanisms are seemingly at work.

Production advantage

In the memory literature, the advantage of production is explained with the concept of distinctiveness, which has been integrated into a proceduralist account that assumes storage of concrete production traces (that represent the production act) together with lexical representations (MacLeod et al., 2010). The advantage of production training in the present study might be taken as support for the idea that the information that a word was produced is stored along with its lexical representation. The production advantage, moreover, does not seem to be of temporary nature because the memory advantage was still observed after one week in Experiment 9. The generally lower d-prime score in Experiment 9 (one-week delay) is in line with the common process of forgetting that has been postulated in the Ebbinghaus *forgetting curve* (Ebbinghaus, 1987). The proceduralist account furthermore assumes that the information itself that a word was produced facilitates recall.

Experiment 10, where a production advantage was found even when self-listening during production was precluded, can be taken as support for this assumption. Fawcett (2013) completed the proceduralist account with the concept of primary and secondary distinctiveness: both in within- and between-participant designs, encoding and retrieval processes play a role (due to secondary distinctiveness), but in a within-participants design, more retrieval processes are active (due to primary distinctiveness) and therefore the production advantage is greater within participants than between participants. In the present experiments, every participant both produced some words and listened to some words produced by a second participant. Regarding primary and secondary distinctiveness, both encoding and retrieval processes were active, but retrieval processes contributed more strongly. The information that an item was produced is stored and accessed at retrieval, which facilitated recognition of produced tokens.

However, the present results can also be explained through abstractionist principles of the mental lexicon. These principles are juxtaposed with episodic accounts (Goldinger, 1998) where every concrete exemplar of a speech unit encountered by a listener is stored in the mental lexicon. In abstractionist models (e.g., McClelland & Elman, 1986; Norris, 1994; Norris & McQueen, 2008), abstract representations of a word's canonical representation build the lexicon itself.

An observed advantage of production could be explained as follows: Production increases a word's distinctiveness, which potentially amplifies encoding strength. In line with Craik and Tulving (1975), greater encoding strength implies better word memory. This increased word memory can be explained by a stronger resting activation level of the abstract lexical representation due to the amplified encoding strength. Based on pre-lexical processes, it can then be more easily activated in the subsequent comprehension process. This strengthening would rely on the strong learning benefits from self-production alone (because no self-listening is required for the advantageous effects of production) and is long lasting, as shown in Experiment 9.

It can be explained in a manner similar to how frequency effects have been explained within abstractionist models. McQueen, Cutler, and Norris (2006), for example, suggest that in an abstractionist account, processing might be probabilistic. In Shortlist B, a model describing word activation based on Bayesian principles, Norris and McQueen (2008) suggest that word activation depends both on a potential candidate's prior probabilities and on the current evidence in favor of them. These prior probabilities can be increased by means of production and the resulting distinctiveness and amplified encoding strength. Distinctiveness is further increased by the between-participants design, and this, in turn, increases the resting activation level boost. In comparison to the proceduralist account, the abstractionist principles imply greater economy, because not every single instance of a production experience is stored in the lexicon, but existing representations are strengthened.

The strength of the production advantage in the present study was not affected by accent familiarity, implying a quite general nature of the effect. In contrast, Cho and Feldman (2013) found a numerically smaller production effect for the global Dutch accent compared to the American English accent (cf. Table 3 in Cho and Feldman, 2013, p. 307 —— the difference between d-primes for produced words and listened-to words is generally greater within American English-accented words than within Dutch-accented words). This could have been caused by an acoustic discrepancy between the to-be-imitated Dutch sound prompt and the American English participants' actual imitations. The participants had to imitate a global Dutch accent with many different accent markers, and not, as in the present study, one specific accent marker.

Such an interpretation could also explain results by Baese-Berk and Samuel (2016). They found that imitating a newly learned L2 sound can inhibit learning. In their study, participants had to imitate sounds from an artificial sound continuum ranging from /s̠a/–/ʃa/, which can arguably be difficult for speakers to imitate precisely. A potential acoustic discrepancy between the sound prompt and the participants' productions may therefore have inhibited learning effects.

A discrimination study with Danish vowels (Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015) supports this interpretation. In that study, production accuracy was increased by concrete feedback on productions, which in turn resulted in better sound discrimination performance after production than after listening training. Moreover, Cho and Feldman (2016) found in a repetition task that the acoustic information of the to-be-repeated auditory prompt is not lost due to the overt repetition of a study participant. Therefore, discrepancy between the acoustic prompt and participants' productions can indeed negatively affect participants' memory, whereas congruency potentially facilitates memory.

Advantage of familiarity

The present experiments also suggest that words in a familiar accent are more likely to be remembered than words in an unfamiliar accent. This means that unlike the hypothesis formulated in the discussion section of Chapter 2A, during active recall acoustic details, i.e., the accent marker, are not ignored, but can affect memory. The advantage of the familiar accent probably results from long-term experience with that accent, and experimental changes in pre-lexical processing possibly then control the lexical representations' resting activation level (together with learning modality).

When familiarly accented words are presented, the pre-lexical rules describing that accent are activated and facilitate processing of words in the familiar accent. This increases the resting activation level of the respective lexical representation. For words in the unfamiliar accent, the listener (with no prior accent experience) has no pre-lexical rules available. Therefore, the resting activation level is not as strongly increased, resulting in a memory advantage for words in the familiar accent. In the present study, the familiarity advantage was not as strongly pronounced when participants did not hear themselves or heard the second participant over headphones, because of the unusual experimental circumstances. In that case, participants relied on the orthographically presented words to a greater extent, reducing accent familiarity effects. The observation that the accent familiarity effect was gone after one week reflects the reduction of the resting activation over time when no further accent exposure is provided.

The accent familiarity advantage was not found in free recall, but only in the memory recognition task because the former task elicited the activation of only a fraction of all experimental words. It was suggested that the small amount of recalled words accounts for this difference. It may, however, also reflect different psychological processes that the two tasks rely on. Recall and recognition are treated as distinct processes in declarative memory retrieval (e.g., Ben-Yakov, Dudai, & Mayford, 2015). Recall involves a generation phase followed by a recognition phase (e.g., Kintsch, 1970), whereas recognition is based on two different types of memory — familiarity and recollection (e.g., James, 1890).

Familiarity primarily refers to familiarity with the word form itself; however, it can also apply to the present investigations of accent familiarity. The involvement of familiarity memory in word recognition, but not in free recall, might account for why in the current study, accent familiarity effects were only present in the word recognition task. This was, however, not the motivation for focussing the present analysis on word recognition results. The memory recognition data were focused because the findings from relevant prior studies also rely on memory recognition data (e.g., Cho & Feldman, 2013; MacLeod et al., 2010; MacLeod, 2011).

Accent familiarity effects were also tested by Cho and Feldman (2013). Unlike the present results, they found that words with an unfamiliar accent were more easily recognized in the listening condition than words in the familiar accent. However, their familiar accent (American English) probably corresponds most closely to the unaccented filler words in the present study, both adhering to the standard variety of the target language. The unfamiliar accent condition (Dutch-accented English) in Cho and Feldman deviates from the standard pronunciation and is therefore comparable with the unfamiliar accent condition in the present study (northern accent), which also deviates from Standard German and is still unfamiliar to participants.

An appropriate comparison with Cho and Feldman's data might therefore be that of hit rates for unfamiliar northern-accented words and unaccented fillers in the present experiments. In the present study, the number of experimental words was considerably higher than the number of unaccented fillers (that were, however, controlled for frequency and length), rendering a direct comparison less reliable. Nevertheless, a post-hoc descriptive analysis of unfamiliar northern-accented words and unaccented fillers from all three experiments together suggests that for the unfamiliar northern accent, experimental words had a hit rate of 77.4 % (SD = 0.42) and unaccented fillers of 66.8 % (SD = 0.47)[14]. Thus, more words with an unfamiliar northern accent were correctly remembered than unaccented fillers. This trend aligns well with the findings of Cho and Feldman, that include that words with an unfamiliar global Dutch accent were recognized more easily than words produced by a standard speaker of American English.

On the other hand, Clopper, Tamati, and Pierrehumbert (2016) found an advantage for less strongly accented words (Midland accent) compared to more strongly accented words (northern accent). Their finding is supported by prior accent processing studies. For example, Bent and Bradlow (2003) found that for native English listeners, sentences produced by a native English speaker are more intelligible than those produced by native speakers of Chinese or Korean (evidence for greater

[14] The same pattern can be observed when hit rates for words with the familiar and unfamiliar accent together are compared with the filler hit rates: accented 78.8 % (SD = 0.41), fillers 66.7 % (SD = 0.47).

processing costs for accented speech was also provided in Gill, 1994 and Pickel and Staller, 2012).

However, in the present study, as well as in Cho and Feldman, the accented (and canonical) training words were always presented together with their orthographic transcript, which was a prerequisite for the production task. Presumably, this strategy eliminated major processing differences between accented and canonical words. The advantage of accented tokens then refers to distinctiveness. Accented speech implies a greater level of distinctiveness than canonical speech, which can, for example, increase the resting activation levels of lexical representations and therefore lead to a memory advantage. As such, it is assumed that distinctiveness is a universal concept that can apply to different factors and impact word memory independently.

Practically speaking, the present findings support the idea that reading aloud alternating with listening to a second person is a useful strategy in studying word lists, for example in the context of memorizing a shopping list or when studying a list of names. When considering accent as an additional learning strategy, the present findings suggest that using one’s own accent for studying is particularly helpful for remembering.

Conclusion

In conclusion, Chapter 4 shows that self-production and accent familiarity enhance word memory. The production advantage is long lasting and was observed even without self-listening, suggesting great robustness. The advantage of accent familiarity, however, was more flexible and vanished after one week. Accent familiarity was not as strongly pronounced without self-listening. The memory advantage of self-production is seemingly independent from the advantage of accent familiarity, suggesting that different processes are at play.

Chapter 5

Summary and conclusions

Abstract

This chapter summarizes the major findings and their interpretation of the present dissertation. It provides answers to the three major research questions asked in Chapter 1. More specifically, the findings of the preceding chapters are applied to accent learning, learning with production, and the role of salience. The conclusion emphasizes the key findings of the studies presented beforehand.

Global interpretations

The present dissertation asked whether and to what extent producing, compared to listening to, accented words can contribute to accented word learning and accent learning more generally. Learning effects of accented speech production were compared with learning effects resulting from listening to accented speech. This was specifically asked in the global Question 1. By this comparison, conclusions could be drawn regarding the nature of learning and how learning mechanisms induced by listening and production relate to one another, which was asked in Question 2.

Foreign accent learning was compared with native accent learning by presenting speech material recorded by L2 and L1 speakers. Moreover, the role of listeners' native language background was investigated. The speakers presented during training and test were always different in order to test speaker-general learning. A further goal of this dissertation was to characterize the learning effects in terms of the processing levels where they are observed. This was done with different experimental paradigms. Reaction time and eye-tracking tasks investigated the effects of learning on online processing, and memory tasks looked at the effects on memory recognition. The generality of learning with production was also tested by comparing learning with long-term familiar and unfamiliar accents. Further aspects that describe these learning effects refer to how long lasting they are and what the role of self-listening is. Finally, Question 3 scrutinized the role of salience in accent learning and learning with production and listening.

Altogether, the results described in Chapters 2A–4 demonstrate that an accent and specific accented words can be learned through listening and production training. Learning can be observed on different levels of processing, which demonstrates its generality. Accent learning generalizes across speakers, suggesting that it is based on abstract, pre-lexical rules. In the process of learning with production and listening, the similarity between speakers and listeners in terms of native language background seems to be more important than the difference between foreign and native accents. Under certain conditions, production training can help more than listening training. These conditions are being a second language speaker with more flexible categories than L1 speakers or experiencing mixed-modality training (including production and listening), which potentially increases distinctiveness and salience of produced tokens.

Still, learning accented words with production follows the same basic principles as learning with listening. The advantage of production is explained by a greater resting activation level of the lexical representations (and potentially pre-lexical rules) after production. Long-term accent familiarity does not affect the production advantage. The advantage of production lasts at least one week and still emerges when producers do not hear themselves during production. Salience during training can modify learning strength, and salience during test can determine whether learning effects are visible. The findings from the single experimental chapters that brought me to these conclusions are now summarized separately. Then, advanced considerations on the theoretical background of the current findings are presented.

In Chapter 2A (Experiments 1–3), the effects of production and listening training on the subsequent comprehension of foreign-accented speech were investigated in three experiments, each based on a training-test paradigm. During training, L1 English and German L2 participants either listened to a short story in which all /θ/s were replaced with /t/ (e.g., *theft* was pronounced by a German speaker as **teft*). They either read the story out loud with the *th*-substitutions or they had no accent training. During a subsequent test, participants made auditory lexical decisions on English words with *th*-substitutions — the accent from the training. The test words were recorded by a different speaker than the training material to test speaker-general learning effects.

L2 participants' reaction times to words from the training were significantly faster after having produced the story than after no training; having listened to the short story also resulted in faster reaction times, but less strongly so. The effects did not generalize to untrained words, however. For L1 participants, training did not facilitate accent processing significantly, and as such there was no difference between production and listening training in this group. Thus, only for L2 participants was the effect of accent production for learning superior to accent listening. Production training therefore affected comprehension processes, and it can even help comprehension more than listening training. However, these effects were limited to participants' second language.

A potential reason why the accent was not learned as effectively by L1 participants as by the L2 participants in Chapter 2A is the diverging native language background between the pre-recorded speaker (L1 German) and the participants (L1 English). This option was addressed in Chapter 2B (Experiment 4), which asked whether it is easier to learn an accent when the speaker has an accent similar to the listener's accent. The same training-test design as in Experiments 1–3 was used. During training, L1 English participants either listened to a story recorded by an L1 English speaker who replaced all /θ/s with /t/, or they produced the story with the *th*-substitutions themselves, or they had no training (control group).

Speaker-listener similarity was ensured by having L1 English speakers during training and test, meaning that both the pre-recorded speakers and the participants had the same native language background. Lexical decisions were significantly faster after production and listening training than without training. In contrast, the reaction time differences between the L1 control group and the L1 experimental groups were not significant in Chapter 2A (Experiment 3). Accent learning with production or listening was promoted by greater speaker-listener similarity as reflected by identical native language backgrounds of speakers and listeners.

Chapter 3A (Experiments 5–6) further investigated native accent learning for L1 participants, considering the factor *salience*. Two experiments tested the degree to which an accent is learned with production or listening training. A training-test paradigm was administered to native German participants utilizing an artificial German accent. During training in Experiment 5, participants either read single German words out loud and deliberately devoiced initial voiced stop consonants (e.g., *Balken 'beam'* pronounced as **Palken)* or they listened to pre-recorded words with the same accent. In a subsequent eye-tracking experiment, looks to auditorily presented target words with the accent were analyzed. Training and test words were recorded by two different L1 German speakers to test speaker-general learning.

Participants from both training conditions fixated accented target words more often than a control group without training. Training was identical in Experiment 6, but during test, canonical German words that overlapped in onset with the accented words from training were presented as target words (e.g., ***Pal**me 'palm tree'* overlapped in onset with the training word ***Pal**ken*) rather than accented words. This time, no training effect was observed; recognition of canonical word forms was not affected by having learned the accent (as shown in Experiment 5). The accented form was not activated when a potential competitor was presented auditorily. The accent was learned, but not strongly enough for accented forms to act as competitors.

Salience of the test words accounts for the fact that learning was only visible with accented words in Experiment 5 but not with canonical words in Experiment 6. Accent learning was only visible when the test tokens possessed sufficient salience. This was the case for the accented test tokens but not for the canonical test tokens. To become visible, an accent that is not very strongly learned requires test words with a sufficient degree of salience.

Chapter 3B (Experiment 7) focuses on eye-tracking as a research method. In Experiment 7, the same material and procedure as Experiment 5 were made use of but a different eye-tracking hardware, the *EyeTribe* tracker (https://theeyetribe.com/) was used. This mobile, low-cost eye-tracker has a maximum sampling rate of 60 Hz (records one data point every 16.67 milliseconds) and does not come with pre-installed software. In contrast, the *Eyelink 1000* (SR Research Ltd.) that was used for Experiments 5 and 6, can record eye movements with a sampling rate of up to 1000 Hz (one data point every single millisecond) and is equipped with pre-installed software.

Experiment 7 shows that the *EyeTribe* provides an acceptable pattern of results when compared with the data gathered with the *Eyelink 1000* in Experiment 5 — even with a sampling rate of 30 Hz (one data point every 33.3 milliseconds). This is good news in terms of eye-tracking hardware: low-cost, mobile technology with low sampling rates such as the *EyeTribe* suffice to obtain reasonable results in visual world experiments. However, use of the *EyeTribe* involves some drawbacks. For example, it does not come with pre-installed software and does not offer the option of drift correction during an experimental session. Moreover, a large amount of the collected data could not be included in the analysis because the eye-tracker did not always record fixation coordinates.

Chapter 4 (Experiments 8–10) also looked at native accents, this time focusing long-term accent familiarity. The effects of production and listening training as well as long-term accent familiarity on word memory were investigated in a training-test paradigm. This means that accent learning was not investigated, but it was instead tested, in which way an accent affects word memory. Single words were produced or listened to by L1 German participants (raised in southern Germany with the Swabian accent) in a training session. Training words either had a Swabian accent marker that was familiar to participants (/st/ pronounced as /ʃt/, **Zahnbür/ʃt/e ‘tooth brush’*) or an unfamiliar northern German accent marker (/ʃt/ pronounced as /st/, **Blumen/st/rauß* ‘*bouquet*’).

After training, participants decided with a button press whether each word was old or new in a visual memory recognition task for individual words from the training phase (old words) and new words. D-prime scores (a measure accounting for hit rates and false alarms) were significantly higher for self-produced words than words that were listened to in either accent condition (familiar Swabian or unfamiliar northern German). The same pattern was observed when participants did not hear themselves during production. When word memory was tested one week after the training, produced words were still more likely recognized.

Chapter 4 shows that accented words in participants' L1 can be learned with production more easily than with listening to a second person. This production advantage relies on production alone, is general across different accents, and long lasting. The production advantage relies on the training list design used in that study. Every participant had to produce some words and listen to other words during the training phase and was not restricted to one of either (as in Experiments 1–7). The specific processes behind accent learning, learning with production, and the role of salience in these processes are now further discussed and elaborated.

Accent learning

Foreign versus native accent learning

In Chapter 1, the processing of foreign versus native accents was discussed, and a definition of each was presented. A foreign accent was defined as an accent produced by non-native speakers, and a native accent as an accent produced by native speakers of a given language (Wells, 1982). Moreover, the difference between specific accent markers and global accent was exemplified. Specific accent markers result from the substitution of one specific sound with a second. The co-occurrence of multiple specific accent markers (presumably differing in strength) constitutes a global accent. An example of a global accent is German-accented English. All experiments reported in this dissertation used an accent that can be explicitly instructed to participants in the production training conditions. This is only possible with a specific accent marker. However, since naturally recorded (not computer-generated) speech was used, an additional global accent marked the speech signal. Most obviously, this was the case for the recorded L2 speakers in Chapter 2A (Experiments 1–3). The native German speakers were instructed to replace every dental fricative /θ/ with a /t/. However, without being aware of it, they added numerous minor accent markers resulting in the German accent in English on the global level (for possible specific accent markers, see Swan, 2001).

The nature of foreign and native accent processing has been implemented in two relevant hypotheses — the *Perceptual Distance Hypothesis* and the *Different Processes Hypothesis*. The *Perceptual Distance Hypothesis* suggests the same processes for foreign and native accents, and differences rely on the distance of the accent from non-accented, canonical speech. This distance can be paraphrased as *accent strength* that is referred to in Chapter 2B. The *Different Processes Hypothesis,* on the other hand, suggests that foreign accent processing is based on mechanisms that are different from those activated for native accent processing. Both hypotheses draw on empirical support from prior studies; however, the research community has not agreed upon either of the two options thus far. One reason might be that it is hard to determine the distance of a specific example of accented speech from non-accented speech, i.e., accent strength, without comprehensive, detailed acoustic information of the accented speech material.

The present results touch on this problem, as demonstrated by the listening training groups. A direct comparison between foreign and native accent learning is possible between Experiments 3 and 4. In these experiments, the native accent is defined by one specific accent marker without additional global foreign accent markers because it was produced by two native speakers who were specifically instructed to produce the specific accent marker, but otherwise speak as neutrally as possible. The foreign accent is defined by one specific accent marker with additional non-specific global accent markers that were produced by two L2 speakers. Therefore, the foreign accent presumably has a greater accent strength[15]. All other experiments presented in the presented thesis tested native accent learning without directly comparing the results with foreign accent learning.

In Experiment 3, the American English participants listened to speech recorded by an L2 speaker (L1 German) and were tested on words recorded by a second L2 speaker with the same native language background (German). Trained listeners were only numerically faster than the listeners without training. In Experiment 4, the L1 American English participants listened to speech recorded by an L1 American English speaker and were tested on words recorded by a second L1 American English speaker. This time, the participants who completed the listening training were significantly faster than the participants without the training. Native listeners learned a weaker, native accent more easily than a stronger, foreign accent. It was harder for participants to learn the same specific accent marker when additional accent markers on the global level were included in the training and test material. This finding supports the basic idea that the same processes apply for native and foreign accent processing, as suggested in the Perceptual Distance Hypothesis.

This finding, however, does not imply that the *speaker*'s native language background *per se* determines the ease of accent learning. In Experiment 3, accent learning with foreign, i.e., more strongly accented, test material (German-accented

[15] Chapter 1 clearly states that native speakers can produce native accents that include segmental and suprasegmental accent markers. The recordings that were presented to the participants were not analyzed acoustically. However, based on L1 speakers' perceptual judgements of the present material and the instruction to the L1 speakers to speak as neutral as possible, I conclude that the pre-recorded L2 speakers did exhibit a stronger accent than the pre-recorded L1 English speakers.

English) was not observed by L1 English participants, but with the same test material, accent learning was observed by L1 German participants in Experiment 2. When the native language background was the same for the pre-recorded speakers and participants, accent learning was more successful. This assigns an important role to the *participant*'s native language background that likely relies on similarity between listener and speaker. Therefore, this section does not argue for one of the two hypotheses presented above, implying that not accent strength alone is the crucial factor in accent adaptation. It emphasizes the native language background similarity between a speaker and a listener. Nonetheless, accent strength should still be considered as a further relevant factor in accent adaptation, as argued in Chapter 2B. Stronger accents might reduce learning effects from listening training, but not from production training.

Models of spoken word recognition

Models of spoken word recognition can explain how an accent is learned with listening experience. Chapter 1 illustrated that there are two major groups of these models: abstractionist accounts (McClelland & Elman, 1986; Norris, 1994; Norris & McQueen, 2008) and episodic accounts (e.g., Goldinger, 1998). Whereas episodic accounts suggest the storage of every concrete exemplar of a speech unit encountered by a listener (including accent properties), in abstractionist models, abstract representations of a word's canonical representation build the lexicon. Variations of the canonical form, such as accents, can be accounted for by pre-lexical mapping rules. These rules are built on the basis of a few exemplars that are no longer stored. When, for example, an accented token is encountered after accent training, the learned rule is applied to the respective abstract entry in the lexicon. This explains why learning a specific variation can generalize across many different words (McQueen, Cutler, & Norris, 2006).

McQueen, Cutler, and Norris (2006) also suggest that in an abstractionist account, processing might be probabilistic. In Shortlist B (Norris & McQueen, 2008), an abstractionist model based on Bayesian principles, word activation depends both on a potential candidate's prior probabilities and the current evidence in favor of them. The more often a candidate is encountered, the higher the prior probability. This means that a candidate's so-called resting activation level is increased, which accounts for frequency effects (that is, the more often a word is encountered, the faster it is recognized). Additionally, there are hybrid models (e.g., McLennan, Luce, & Charles-Luce, 2003) that borrow ideas from both model groups. They can, for example, integrate exemplars and pre-lexical mapping rules into a single account.

The present findings overall support the existence of pre-lexical rules and thereby an abstractionist account. Most importantly, accent learning was observed across speakers. Participants were exposed to a different speaker during training and the test phase. A strictly episodic account cannot explain the results, because in such an account, learning only occurs for the same word produced by the same speaker. Abstractionist accounts, however, do predict generalization across voices. During the training phase, a pre-lexical rule that describes the accent (e.g., /θ/ is pronounced as /t/ in Experiments 1–4) is built, and this rule applies to all future input — with different voices — that might match the accent.

A further argument in favor of pre-lexical mapping rules comes from the findings from Experiment 5. Accent learning was observed for both accented words that were included in the training (old words) and for new words with the same accent. The critical accent rule (stop devoicing in word initial position) was learned by means of two sound examples: the bilabial (b/p) and the velar stop (g/k), which increased variability during training. The findings support the idea that the concrete exemplar of the accented word was not stored but rather an accent rule that is applied to new words with the trained accent.

In Experiments 1-4, however, learning effects were only observed for old words from the training and not for new words. Since voice general learning was still found in these studies, abstractionist models nevertheless account for these results. It is possible that more time and greater variability during training is required in these cases, which is supported by Experiment 5 and a different accent learning study by Bradlow and Bent (2008). Bradlow and Bent found that training with Chinese-accented English sentences produced by multiple speakers results in significantly better performance in a subsequent sentence transcription task by L1 English listeners than training of the same accent produced by a single speaker different from the test speaker.

The process of rule building might likely proceed in a stepwise fashion. The first step would be a partially abstract lexical representation that still involves accent, but no voice information. This allows generalization of accented word learning to new voices, but no generalization of accent learning to new words yet. Then, in the second step, based on a number of these lexical representations involving words with the same accent, pre-lexical accent rules are built. The first step is accomplished quite early after only little training, while the next step requires more intense training and a greater number of concrete exemplars on which pre-lexical rules can be built.

In Chapter 4, the effects of long-term accent familiarity on word memory, instead of short-term accent learning, were tested. Long-term accent familiarity resulting in pre-lexical accent rules facilitated word learning. When familiarly accented words were presented, the pre-lexical rules describing that accent were activated and facilitated processing of words in the familiar accent. This increased the resting activation level of the respective lexical representation. For the unfamiliar accent, the listener (with no prior accent experience) had no pre-lexical rules available. Therefore, for words in the unfamiliar accent, the resting activation level was not as strongly increased, resulting in a memory advantage for words in the familiar accent.

A different way to interpret the present results rejects the account of learning the specific accent and assumes that the representations in the lexicon are simply loosened. As a result, the listener is more tolerant towards variation in the speech signal. This would mean that learning that the German word *Balken 'beam'* is pronounced as **Palken* causes listeners to interpret both **Palken* and **Malken* equally well as *Balken*. Whether this is the case was not explicitly tested in the present experiments, but prior research supplies strong evidence for accent-specific learning and against general loosening of criteria.

For example, Maye, Aslin, and Tanenhaus (2008) found training effects for an accent that included front vowel lowering of certain vowels (e.g., *witch* pronounced as **/wɛtʃ/*). This pattern generalized to new words with front vowel lowering that were not included in the familiarization phase and a weaker effect was also observed for new words with a lowered back vowel. However, it did not generalize to words with a raised front vowel (e.g., *witch* pronounced as **/witʃ/*). Accent learning generalized to new words and new vowels, but it was direction-specific and did not simply loosen the criterion for what constitutes a member of a given sound category.

In summary, accent learning relies strongly on abstract rules and lexical representations. These rules are built during accent learning. In comparison to episodic accounts, the abstractionist principles imply greater economy, because not every single instance of a production experience is stored in the lexicon; existing representations are strengthened instead. I now turn to the second main aspect of my dissertation that considers learning with production.

Learning with production

The present experiments show that learning can be more effective with production than when compared to listening to a recording made by a second person (Chapter 2A) or a second person in the same experiemental session (Chapter 4). The memory advantage for produced words was still robust when speakers did not hear themselves during production (Experiment 10), and it lasted up to one week (Experiment 9). However, these experiments also show that the production advantage does not apply unconditionally. In a between-participants training design, where one participant was either only trained with production or only with listening, production promoted learning more than listening in participants' L2 (Experiment 2), but not in their L1 (Experiments 3–4). In L1, an advantage of production was found in a within-participants training design (Experiments 8–10) where one participant both produced and listened to single words during training. These findings generate two important factors that define the production advantage.

The first factor that defines the production advantage is the producer's native language background. In their second language, participants learn more easily with production than with listening only. The reason for this is the greater phonological and phonetic variability that speakers exhibit in their L2 compared with in their L1 (Baese-Berk & Morrill, 2015; Wade, Jongman, & Sereno, 2007). Greater variability provokes greater flexibility of the representations activated for production which are therefore prone to learning new variation in speech.

This idea is supported by prior findings on accent processing. Weber, Di Betta, and McQueen (2014) for instance, found effects for native Dutch and native Italian L2 listeners for primes with both a genuine (*trick* pronounced as **tri:k*) and an arbitrary Italian accent in English (*treat* pronounced as **trɪt*) in a cross-modal priming study priming, but L1 English listeners experienced priming only with the genuine accent. This potentially results from greater flexibility of L2 categories compared to L1 categories. In Chapter 2B of the present dissertation, the production advantage for L2 participants, but not L1 participants, in a within-participants modality training was further explained through accent strength. In this account, L2 listeners learned the accent less easily because the test material was recorded by an L2 speaker with a stronger accent that inhibited learning by the listening group. The L1 group in Experiment 4, on the other hand, was presented with test material recorded by a less strongly accented L1 speaker. Altogether, global accent strength might additionally bolster the production advantage in L2 participants, but the present evidence suggest that it is primarily induced by L2 producers' greater categorical flexibility.

The second factor concerns the training list design and has been discussed in the context of comprehensive studies on the so-called *production effect* (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). The production effect predicts a memory advantage of produced words over silently read or listened-to words when one study participant both produced and read words silently (or listened to pre-recorded speech) in the same experimental session. This is supported by the current experiments where, even in their L1, participants learned accented words more easily with production when the training included both production and listening (Experiments 8–10).

Hence, an important amount of weight is attached to the direct contrast between training modalities. However, in Chapter 2A, it was suggested that the production advantage might depend on the nature of the task, meaning that it primarily functions in memory tasks and less well in an online processing task. It was also suggested that in active recall, acoustic details are ignored and that the production advantage instead applies to abstract information. Analogously, the most robust production advantage was found in the memory tasks presented in Chapter 4 of this dissertation.

These memory tasks are, however, the only experiments in this thesis where training was manipulated within participants (one participant produces and listens during training) and not between participants (one participant either produces or listens). There is plausible evidence from various memory studies suggesting that learning is more robust within participants than between participants (presented in a meta-analysis by Fawcett, 2013). Moreover, a nonword learning study (Zamuner, Morin-Lessard, Strahm, & Page, 2016) in which single nonwords were learned either with repetition or with listening twice, found in the subsequent eye-tracking task that the repetition tokens were more easily recognized than the listening-only tokens[16]. This result strongly suggests that the training list design, and not the nature of the task, plays the critical role. Additionally, the second aspect included in the hypothesis from Chapter 2A, which suggests that acoustic details are ignored in the production advantage, is not supported, because there was an advantage for words with the familiar accent over words with the unfamiliar accent in the present Experiment 8.

In summary, producing accented words facilitates accent learning when appropriate conditions are fulfilled. Three relevant theories and hypotheses that focus on the role of production in learning are discussed in the following sections. The first is the distinctiveness account, which focuses on the learning process itself, the second discusses the nature of representations accessed for production and comprehension, and the third account, the similarity account, relates to recognition processes.

Step 1: Learning ––distinctiveness and production

The advantageous nature of production in learning had already been noted in early psychological theories. William James (1890) stated that activeness crucially facilitates

[16] Note that the study by Zamuner et al. (2016) involved a repetition paradigm (auditorily presented tokens had to be repeated aloud) during training, but in order to compensate for the double input in the repetition condition, the listening tokens were presented twice.

learning. The advantage of activeness was also emphasized by Zinchenko, Vygotsky, and Leont'ev (referred to by Wertsch, 1979) as well as Craik and Lockhart (1972). If production is considered a process that is more active than listening, then James' theory predicts facilitated learning with production, which is supported by the findings in the present Experiments 2, 8, 9, and 10. The advantage for produced words can also be explained through the embodiment hypothesis (for an overview: Glenberg, 2010). Listeners focus more on their own than on others' productions because their own productions are more embodied. Moreover, Experiment 10 showed that self-listening is not the crucial factor in the advantage of accented word learning with production. Therefore, the option that activeness or embodiment plays the critical role seems reasonable.

Recent memory literature explains the production advantage with the argument that distinctiveness provokes additional processing and, therefore, better memory (for a review: McDaniel & Geraci, 2006). For a word to be distinct, it must be different from the subsequent word, and there must be a direct contrast between the two. In the case of the production advantage, the direct contrast between produced tokens and non-produced tokens is guaranteed in a within-participants design. MacLeod and colleagues (2010) combined the assumptions of distinctiveness and the proceduralist account of Kolers (1973) to provide an explanation of the production advantage found in many memory studies. In this concept, a produced word has an additional source of discrimination compared to a word that was not produced during the study phase. In a memory recognition task, the information that the word was produced is accessed, and this information necessarily includes that the word was included in the study phase.

This account is based on the idea that the concrete learning experience plays a role in the nature of the resulting lexical representations. Therefore, this concept is most compatible with an episodic account of the lexicon (e.g., Goldinger, 1998) that assumes concrete, exemplary representations. The trace that includes the specific production information represents specific, indexical information. Within the scope of episodic theories, the lexical representation of a word can include the production trace resulting from production training. A strictly abstractionist account, on the other hand, prescribes that the abstract lexical representation can include no further information of speaker-specific properties or the learning modality either. In the end, the advantage of production can be explained within abstractionist models: production increases a word's distinctiveness, potentially amplifying encoding strength. The greater encoding strength increases the resting activation level of the abstract lexical representation and therefore its memory.

This conclusion is in line with Craik and Tulving (1975), who claimed that greater encoding strength implies better word memory. Based on pre-lexical processes, the word can then be more easily activated in the subsequent comprehension process. This strengthening relies on the strong learning benefits from self-production alone (recall that self-listening is not required to reap the advantageous effects of production, see Experiment 10) and is long lasting, as shown in Experiment 9. It is explained in a way similar to how frequency effects have been explained by abstractionist models. In line with the assumptions made in Shortlist B (Norris & McQueen, 2008), where word activation depends on both a potential candidate's prior probabilities as well as on the current evidence in favor of them, a candidate's prior probabilities are increased by means of production and the resulting distinctiveness and amplified encoding strength.

Distinctiveness can be even further increased by the between-participants design, and this, in turn, increases the amount that the resting activation level is boosted. The production advantage might also generalize to new words with the same accent, assuming that the production act similarly strengthens pre-lexical accent rules. If this is true, L2 participants who are trained with both production and listening on a specific accent should recognize produced words with the same accent more easiliy than listened-to words in a subsequent auditory lexical decision task, even when these words were not included in the training session.

Considering the concept of distinctiveness, having participants both produce and not produce (read silently or listen) in the same experimental list (within-participants design) was traditionally treated as a prerequisite for a memory advantage resulting from production (MacLeod et al., 2010). However, the L2 participants in the present Experiment 2 also had an advantage by production in a between-participants design[17]. This is a result of the greater flexibility of L2 categories resulting from production variability discussed above.

Step 2: Representations of production and comprehension

The questions surrounding the nature of and relation between representations accessed for production and comprehension have resulted in long-standing and ongoing discussions. Most well-established theories refer to the nature of representations in their assumptions regarding the connection between speech production and speech comprehension. For example, classical theories and models (Liberman, 1996; Liberman & Mattingly, 1985; Stevens, 1960) refer to the finding that speech perception can affect speech production, and this effect has been made use of in several experimental paradigms (e.g., the picture-word interference paradigm:

[17] There is also evidence that a weak production advantage can still be elicited in a between-participants design, even with L1 participants (Adank, Hagoort, & Bekkering, 2010; Fawcett, 2013). A within-participants design still produces a more general and stable production advantage. Fawcett (2013) explains this pattern with two different forms of distinctiveness: primary and secondary distinctiveness. Primary distinctiveness is reserved for effects that only apply within a participant, whereas secondary distinctiveness emerges both within and between participants. Secondary distinctiveness is attributed to both encoding and retrieval processes, while primary distinctiveness applies to retrieval processes only. Therefore, in both within- and between-participants designs, encoding and retrieval processes play a role (due to secondary distinctiveness), but within participants, more retrieval processes are active (due to primary distinctiveness). This explains why the production advantage is greatest within participants.

Glaser & Düngelhoff, 1984; Schriefers, Meyer, & Levelt, 1990; the syntactic priming paradigm: Bock, 1986). However, there is no consensus on how this connection is represented in the mental lexicon. In Chapter 1, various positions as suggested by Kittredge and Dell (2016) were introduced. The positions differ in two aspects: whether the representations are the same (inseparable) or different (separable) and whether there are connections between them if they are separable.

In Experiments 1–4, the effects of production versus listening training on comprehension were directly investigated. In these experiments, production training affected comprehension in a way that helped at least as much as listening training did; in Experiment 2, it helped even more. These observations exclude a separation of production and comprehension representations. The option that production and comprehension representations are different but connected also seems implausible, as it would reduce the economy of the mental lexicon and not adhere to the principles of an abstractionist account. Learning across modalities would always involve greater effort than learning within one modality because the information stored in one representation must first be passed to the second representation. This, again, cannot account for the finding that production training can help subsequent comprehension more than listening training.

The position of identical representations for production and comprehension is therefore adopted here. Within this position, differences between the effects of production training and listening training are possible. The advantage of production training over listening training relies on a greater resting activation level of mental representations after production training. Whether the resting activation level is increased by production depends on participants’ native language background and the training list design. With mixed modalities in the training list, the resting activation level strength is always greater after production than after listening. With a single modality in the training list, the native language background is decisive for the resting activation level strength. The resting activation strength is increased after production for L2 participants because their representations involve greater flexibility due to regular variability in their productions. This is not the case for L1 participants because they have more stable productions resulting in less flexible representations. Therefore, in this account, the same representations are activated for production and comprehension in both L1 and L2, but in L2, these representations are more sensitive to production experience than listening experience due to L2 speakers' greater production variability.

This position does not agree with the general tendency towards separate but connected representations for comprehension and production as reported by Meyer, Huettig, and Levelt (2016). It is instead in line with the conclusions by Chater, McCauley, and Christiansen (2016), which support identical representations.

Step 3: Retrieval ––a speaker-listener similarity based account

Learning effects become visible during subsequent lexical access. This happens during the subsequent comprehension phase, which is specified in the *integrated theory of language production and comprehension* (Pickering & Garrod, 2013). In what follows, this account is referred to as *the similarity account*. Pickering and Garrod claim that during comprehension listeners make predictions about the talker's utterance. Building predictions relies on forward models. A listener covertly imitates what the talker is saying. This strategy can promote the comprehension process. Predictions can rely on the *association route*, which is based on experience with comprehending others' utterances, or the *simulation route*, which draws on the comprehender's individual production experience. Similarity between the talker and comprehender determines which route is chosen. If the comprehenders are similar to the speaker, they rely on the *simulation route*; smaller similarity directs their trail towards the *association route*. Optionally, both routes can be chosen simultaneously. Criteria for similarity are, for example, native language background, dialect, or cultural aspects.

Earlier in this chapter, similarity in terms of native language background was identified as crucial factor in successful accent learning, independent from the production advantage. However, in these studies, similarity did not elicit a production training advantage; it instead strengthened learning effects based on both listening or production training. On the other hand, in Experiment 2 where a production advantage was found, the listeners were also quite similar to the test voices (L1 German, female, similar L2 proficiency in English, similar educational background).

For evaluating the similarity account, it must be emphasized that talker-listener similarity is a criterion that affects the retrieval process. Before retrieval is possible, different criteria, for example participants' native language background, affect the learning process, resulting in different resting activation levels. The present results suggest that these learning criteria are more important than the retrieval criterion of speaker-listener similarity. This is manifested by the present Experiments 4–7 where the L1 participants were quite similar to the pre-recorded speakers (same L1, female, from the same area), but still there was no advantage of production. The production advantage for L2 participants is explained by category flexibility. L1 categories are less flexible than L2 categories, meaning that learning is not promoted. In addition, in Experiments 8–10, there was a quite robust production advantage when the test words were presented visually and no speaker information was available during the test. This either means that speaker similarity only has a secondary effect or even that speaker similarity does not promote the production advantage, but that vice versa, speaker differences restrict the production advantage.

The present findings cannot, however, provide a direct argument in favor of the speaker-listener-similarity based assumptions. Importantly, the present experiments were not explicitly designed to test these assumptions. Whether similarity with a speaker makes listeners rely more on their individual production experience can, for example, be tested with a training-test design where the participant learns two concurrent accents –– one with production, and one with listening training. The speaker in a test session is either clearly similar or different (e.g., the same/different sex) from the participant and produces both accents. If similarity plays a role in the choice of production or listening experience, visible learning effects should differ between the similar and different test speaker conditions.

For testing this, a more specific definition of the term *similarity* must be found. This term implies that the manifestation of a given criterion lies on a continuum ranging from a positive pole (very similar) to a negative pole (not similar at all). The criteria in question can be *native language background* or *dialect*, but also, for instance, a *speaker's sex*, *age*, or *education*. The positive pole is clearly occupied by the identical manifestation of a criterion. For example, regarding the criterion *native language background*, an L1 German participant is identical to the L1 German recorded speaker. The negative pole, however, is not as easy to determine. For the L1 German recorded speaker, all participants that are non-native German speakers involve smaller similarity compared to the L1 German participant. Also, speakers of a Germanic language are more similar to the L1 German pole than speakers of an Italic language. However, there is no absolute, numeric measure of similarity that allows a direct comparison.

Moreover, the multitude of factors that contribute to similarity have not definitely been determined. There is no consensus on what exactly these factors are nor on the weighting of each factor. Therefore, a clear system must be established. Furthermore, either one of the two routes described earlier (*the association route* or *the simulation route*), or both can be chosen. This option complicates the process even more and makes predictions more difficult.

Taken together, this theory predicts under which conditions information that was learned with production or listening is accessed. The crucial factor in this process is similarity. The account refers to retrieval, not learning processes, and exclusively restricts the situations in which material learned with production is accessed. This implies that the information whether something is learned with production or with listening is available in the lexicon. In an abstractionist account with identical representations for production and comprehension, however, these assumptions cannot be integrated. Separate representations for production and comprehension do not exist, and the information whether a lexical token was produced or listened to cannot be stored. Following these considerations and the missing evidence in favor of the similarity account, this account is not included in the subsequent considerations.

Production theories brought together

In the preceding sections, three relevant approaches were recapitulated that treat the problem whether and how (accent) learning with production works. First, the distinctiveness account considers the learning stage that involves production and/or listening. It explains why learning with production can be more effective than learning without production. The second approach refers to the level of mental representations resulting from different training modalities. It considers the general nature as well as differences and similarities between these representations. Third, the similarity account concerns retrieval processes. It specifies a condition under which experience resulting from production is more likely to be accessed than experience resulting from listening. Therefore, each approach describes a different stage of the process. Bearing the preceding discussion and conclusions, two of them are, each slightly adapted, integrated into a single account.

Learning with production and listening is affected by distinctiveness and the resulting salience. Due to direct contrast between different modalities, the production advantage is robust in a mixed training modality design but less so in a single training modality design. A production advantage in a single modality design is only possible for L2 participants with flexible categories. This learning process results in abstract lexical representations and pre-lexical rules in the mental lexicon. Representations accessed for comprehension and for production are the same. Production increases the resting activation of lexical representations.

The findings on learning with production do not apply specifically for accent learning but instead to both accented and non-accented speech. This option is advocated for here, because first, in Experiments 8–10, the production advantage was observed no matter whether the training tokens had a familiar or unfamiliar accent. Second, it was observed for foreign-accented speech in Experiment 2. Third, prior studies (e.g., MacLeod, 2011) found the production advantage when accent did not play a role. The present experiments therefore supply a quite elaborated description of the production advantage. This phenomenon, however, offers the potential for further investigation. For example, advantage of greater voice variability during training on accent learning was suggested in prior research (Bradlow & Bent, 2008). During production, only one voice, producers' own voice, is perceived by the producers themselves. Therefore, it is worth investigating whether production training still results in greater learning effects than listening training when compared with listening training with multiple voices.

Accent learning with production and listening were explained in the preceding two chapters. This, however, is not the full story — one aspect was only touched *en passant* up to this point: *salience*. In which ways salience is involved in the processes discussed above is subject to the following section.

Salience

Accents, training modality, and salience

The concept of *salience* was first introduced in Chapter 1; it was defined as "the property of a linguistic item or feature that makes it in some way perceptually and cognitively prominent" (Kerswill & Williams, 2002, p. 81). In particular, two properties that might exhibit salience are relevant in this dissertation: *accent* and *learning modality* (production versus listening). However, it is hard to find an objective measure for the degree of salience. In terms of accent, the phonetic distance from the standard variant might operate as objective measure. The drawback is that this depends on a highly subjective criterion: the listeners' amount of prior experience with different accents and the canonical pronunciation in their L1.

For example, there is a difference between listeners who live in an environment where the canonical form of a specific language is typically spoken, as is the case for High German for L1 Germans (for example residing in the Hanover area in Germany; see Elmentaler, 2012) and listeners who are frequently exposed to the Swabian accent and who rarely encounter the canonical, non-accented form (those, for example, residing in rural areas of Baden-Württemberg[18]). A Swabian-accented form would be experienced as more salient by the High German environment group than the Swabian accent environment group. Therefore, it is important to consider the accent background of each study participant.

In terms of training modality, the measure cannot rely on participants' specific individual language environment. A participant's personal, everyday experience with production compared to listening to speech is difficult, if not impossible, to measure. In an experiment, experience might be controlled by means of a pre-determined

[18] Assuming that this population does not extensively consume national media, e.g., TV and radio and that they work in a dialect-dominant context.

training phase (specified below). A more indirect manner of determining the strength of salience is to measure its effects in a subsequent test. If something is perceptually or cognitively prominent, it can be more easily learned or it promotes learning of something else (as suggested by Kerswill and Williams, 2002). So, an accent can be salient and more easily learned, or the learning modality, for example *production*, can be salient and also facilitate accent learning.

The distinctiveness account, discussed earlier, assigns an important role to salience. In this account, produced words are more distinctive than non-produced words. This makes them salient, resulting in a memory advantage for these words. This account can also explain the salience exhibited by an accent. In sociolinguistic research, salience of an accented form has been claimed to result from the phonetic difference between an accented form and a non-accented form (Trudgill, 1986). With increasing difference, the dialect speaker becomes more aware of the difference. Greater phonetic difference can lead to greater distinctiveness of the form that deviates from the standard, and greater distinctiveness implies increased salience.

Objective and subjective criteria of salience

Measuring salience is a challenging task because of its subjective nature. With this problem in mind, Auer, Barden, and Grosskopf (1998) introduced *objective* and *subjective criteria* of salience. Subjective criteria increase the salience of a stimulus, due to, for example, regular practice, and the resulting cognitive pre-activation of a stimulus. Objective criteria refer to the properties of a stimulus that itself attracts attention because of distinct, physical characteristics. In Experiments 5 and 6, these assumptions were incorporated with the potential salience criteria *accent* and *learning modality*. According to the definition by Auer and colleagues, the training modality (production versus listening) was the subjective criterion and accent the objective criterion of salience.

In order to keep the objective criterion *accent* as objective as possible and in line with the arguments provided above, participants were tested with an artificial accent that has not been documented to exist in their L1 German. Salience was measured by means of learning effects. The objective criterion was helpful in this task because learning effects were visible with accented test words, whereas learning effects were not visible for test words without the accent. The subjective criterion, however, was not helpful, because production training did not help more than listening training did. This result indicates that training modality cannot exhibit any difference in salience when acting as subjective criterion.

The objective criterion *accent* did, however, exhibit learning effects. The ojective criterion is defined by the direct contrast. Whether training modality can act equally well as an objective criterion was tested in Experiments 8–10, where participants produced and listened during one single training session — and the manipulation was effective. Produced words were more likely to be recalled than listened-to words. This links to the assertion that training list design is an important prerequisite of the production advantage. Bringing these concepts together, the mixed list training modality design involves training modality as an objective criterion, whereas training with only one modality by participant involves training modality as subjective criterion.

Both training modality and accent triggered salience effects, but only when acting as objective criteria, indicating an advantage of objective over subjective salience criteria. It is possible that the training was simply not intense enough for training modality to act as subjective criterion in the present experiments. Subjective criteria potentially require long-term experience of several months or years to be effective. In short-term training studies, effects might only be possible with objective criteria. Therefore, the following discussion focuses on objective criteria of salience.

Salience during learning and recognition

Taken together, salience has its effects both during the learning and the recognition phase. During learning, greater salience can elicit more strongly-integrated accent rules or lexical representations. The distinctiveness account argues for the effect of salience in the learning phase because it implies that more distinct tokens are more salient and can therefore be learned more easily. In the recognition phase, situated in the test phases of the present experiments, test tokens with greater salience make learning effects more likely to become visible. Chapter 3A provides a representative example for salience effects during recognition. This chapter argues that accented test tokens are more salient than non-accented test tokens; therefore accent learning effects were visible with accented test tokens (Experiment 5) but not with non-accented test tokens (Experiment 6).

Chapter 4 (Experiments 8–10) suggests similar conclusions. In Experiment 8, the unfamiliar northern German accent was presumably more salient for the Swabian participants because it was new to them and stood out. However, the accented forms were presented explicitly only in the training phase and not in the test phase (where test words were printed visually on the computer screen). This resulted in an advantage for the Swabian accent that was familiar to the Swabian participants compared to the unfamiliar northern German accent. If, however, salience is also effective during the learning phase, where both the northern and the Swabian accent were presented auditorily, the northern German accent should have been more strongly learned because it is expected to be more salient to the Swabian participants. This did not occur, though, as a result of the way that accents were presented during training. They were presented in two different training blocks, therefore precluding the direct contrast between the two different accents and prohibiting (or at least reducing, if considering Fawcett, 2013) objective salience effects.

These considerations highlight the nature of salience. The production advantage relies on distinctiveness, resulting in a greater level of salience for produced tokens than listened-to tokens. Moreover, accent learning can be modified by salience, potentially resulting in advantages for more-salient over less-salient word forms. Accented forms might be more salient than canonical forms, and unfamiliar accents might be more salient than familiar accents. Salience has effects in both the test phase and the accent learning stage. Salience of to-be-learned tokens can facilitate learning, and salience of test tokens can determine whether learning effects are visible or not.

Conclusion

This dissertation provides evidence for speaker-general foreign and native accent learning. Speaker-listener similarity in terms of native language background plays a role in this process. It suggests that accented word learning and accent learning with production and with listening more generally rely on the same processes and abstract representations and rules, but that production can strengthen mental representations (and potentially rules) and thereby facilitate the learning process. This is easier with a mixed-modality list design and, due to the direct contrast and the resulting greater distinctiveness and salience as well as for L2 participants due to their flexible categories resulting from production variability. The production advantage was observed for differentially familiar accents and on different processing levels; it does not rely on self-listening and is long lasting.

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(2), 520–529. doi:10.1037/a0013552

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, *21*(12), 1903–1909. doi:10.1177/0956797610389192

Adank, P., & McQueen, J. M. (2007). The effect of an unfamiliar regional accent on spoken word comprehension. In J. Trouvain, & W. J. Barry (Eds*.), Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1925-1928). Dudweiler: Pirrot.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. doi:10.1006/jmla.1997.2558

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. . In J. M. Henderson & F. Ferreira (Eds.), *Interfacing language, vision, and action* (pp. 347–386). New York: Psychology Press.

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

Auer, P., Barden, B., & Grosskopf, B. (1998). Subjective and Objective Parameters Determining 'Salience' in Long-term Dialect Accommodation. *Journal of Sociolinguistics, 2*(2), 163–187. doi:10.1111/1467-9481.00039

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. doi:10.1016/j.tics.2012.06.010

Azaz, M. (2017). Metalinguistic Knowledge of Salient vs. Unsalient Features: Evidence From the Arabic Construct State. *Foreign Language Annals, 50*(1), 214-236. doi:10.1111/flan.12248

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (Release 2). Linguistic Data Consortium, Philadelphia.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of America*, *133*(3), EL174-EL180. doi:10.1121/1.4789864

Baese-Berk, M. M., & Morrill, T. H. (2015). Speaking rate consistency in native and non-native speakers of English. *Journal of the Acoustical Society of America*, *138*(3), EL223-EL228. doi:10.1121/1.4929622

Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language, 89*, 23–36. doi:10.1016/j.jml.2015.10.008

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. doi:10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2016). *lme4: Linear mixed-effects models using Eigen and S4_*. R package versions 1.1-7 to 1.1-12, retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Beach, E. F., Burnham, D. K., & Kitamura, C. (2001). Bilingualism and the relationship between perception and production: Greek/English bilinguals and Thai bilabial stops. *International Journal of Bilingualism*, *5*(2), 221–235. doi:10.1177/13670069010050020501

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., & Sawusch, J. R. (1979). Some relationships between speech production and perception. *Phonetica*, *36*(6), 373–383. doi:10.1159/000259974

Ben-Yakov, A., Dudai, Y., & Mayford, M. R. (2015). Memory Retrieval in Mice and Men. *Cold Spring Harbor perspectives in biology, 7*(12). doi:10.1101/cshperspect.a021790

Bent, T., Atagi, E., Akbik, A., & Bonifield, E. (2016). Classification of regional dialects, international dialects, and nonnative accents. *Journal of Phonetics, 58*, 104–117. doi:10.1016/j.wocn.2016.08.004

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

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.

Bock, J. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. doi:10.1016/0010-0285(86)90004-6

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. doi:10.1016/j.cognition.2007.04.005

Brockhoff, P. B., & Christensen, R. H. B. (2010). Thurstonian models for sensory discrimination tests as generalized linear models. *Food Quality and Preference, 21*(3), 330–338. doi:10.1016/j.foodqual.2009.04.003

Chater, N., McCauley, S. M., & Christiansen, M. H. (2016). Language as skill: Intertwining comprehension and production. *Journal of Memory and Language*, *89*, 244–254. doi:10.1016/j.jml.2015.11.004

Cho, K. W., & Feldman, L. B. (2013). Production and accent affect memory. *The Mental Lexicon*, *8*(3), 295–319. doi:10.1075/ml.8.3.02cho

Cho, K. W., & Feldman, L. B. (2016). When repeating aloud enhances episodic memory for spoken words: Interactions between production- and perception-derived variability. *Journal of Cognitive Psychology*, *28*(6), 1–11. doi:10.1080/20445911.2016.1182173

Christensen, R. H. B., & Brockhoff, P. B. (2016). *sensR: An R-package for Thurstonian modelling of discrete sensory data.* R package version 1.4-7, retrieved from http://www.cran.r-project.org/package=sensR/.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658. doi:10.1121/1.1815131

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*, *118*(3), 1661–1676. doi:10.1121/1.2000774

Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics, 39*(2), 237–245, doi: 10.1016/j.wocn.2011.02.006.

Clopper, C. G., Tamati, T. N., & Pierrehumbert, J. B. (2016). Variation in the strength of lexical encoding across dialects*. Journal of Phonetics, 58*, 87–103. doi:10.1016/j.wocn.2016.06.002

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. doi:10.1016/S0022-5371(72)80001-X

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294. doi:10.1037/0096-3445.104.3.268

Crystal, D. (2007). *How language works*. London: Penguin UK.

Crystal, D. (2012). *English as a global language*. Cambridge: Cambridge University Press.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382. doi:10.1177/0267658312443651

Cutler, A. (1995). Spoken word recognition and production. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 97–136). New York: Academic Press.

Cutler, A. (2012). Native listening: The flexibility dimension. *Dutch Journal of Applied Linguistics*, *1*(2), 169–187. doi:10.1075/dujal.1.2.02cut

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America*, *116*(6), 3668–3678. doi:10.1121/1.1810292

Dalmaijer, E. (2014). Is the low-cost EyeTribe eye tracker any good for research? *PeerJ PrePrints.* Advance online publication. doi:10.7287/PEERJ.PREPRINTS.585V1

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *369*(20120394). doi:10.1098/rstb.2012.0394

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review, 8*(1), 155–161. doi:10.3758/BF03196152

Dussias, P. E., Valdés Kroff, J. R., & Gerfen, C. (2013). Visual world eye-tracking. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 93-126). New York: Routledge.

Ebbinghaus, H. (1987). *Translation of Memory: A Contribution to Experimental Psychology*. New York: Dover Publications.

Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology*, *4,* 1-9. doi:10.3389/fpsyg.2013.00148

Elmentaler, M. (2012). In Hannover wird das beste Hochdeutsch gesprochen. In L. Anderwald (Ed.), *Kieler Forschungen zur Sprachwissenschaft. Sprachmythen. Fiktion oder Wirklichkeit?* (pp. 101-115). Frankfurt am Main: Peter Lang AG.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: a study of accent change in young adults. *Journal of the Acoustical Society of America*, *121*(6), 3814–3826. doi:10.1121/1.2722209

Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: a meta-analysis. *Acta Psychologica, 142*(1), 1–5. doi:10.1016/j.actpsy.2012.10.001

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, *93*(3), 1589–1608. doi:10.1121/1.406818

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470. doi:10.1006/jpho.1997.0052

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*(5), 2973–2987. doi:10.1121/1.428116

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

Flege, J. E., & Schmidt, A. M. (1995). Native speakers of spanish show rate-dependent processing of english stop consonants. *Phonetica*, *52*(2), 90–111. doi:10.1159/000262062

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research*, *38*(4), 379–412. doi:10.1007/s10936-008-9097-8

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology. Human Perception and Performance*, *32*(5), 1276–1293. doi:10.1037/0096-1523.32.5.1276

Forrin, N. D., & MacLeod, C. M. (2016a). Auditory presentation at test does not diminish the production effect in recognition. *Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale, 70*(2), 116–124. doi:10.1037/cep0000092

Forrin, N. D., & MacLeod, C. M. (2016b). Order information is used to guide recall of long lists: Further evidence for the item-order account. *Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale, 70*(2), 125–138. doi:10.1037/cep0000088

Geraci, L., & Manzano, I. (2010). Distinctive items are salient during encoding: Delayed judgements of learning predict the isolation effect. *Quarterly Journal of Experimental Psychology (2006)*, *63*(1), 50–64. doi:10.1080/17470210902790161

Gill, M. M. (1994). Accent and stereotypes: Their effect on perceptions of teachers and lecture comprehension. *Journal of Applied Communication Research, 22*(4), 348–361. doi:10.1080/00909889409365409

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 640–654. doi:10.1037/0096-1523.10.5.640

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 586–596. doi:10.1002/wcs.55

Goldinger, S. D. (1996). Auditory lexical decision. *Language and Cognitive Processes*, *11*(6), 559–568. doi:10.1080/016909696386944

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279. doi:10.1037/0033-295X.105.2.251

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102. doi:10.1016/j.bandl.2012.04.017

Grohe, A.-K., Downing, M., & Weber, A. (2016). Similarity in global accent promotes generalized learning of accent markers. In Australasian Speech Science and Technology Association (Ed.), *16th Speech Science and Technology Conference* (p. 61-64).

Grohe, A.-K., & Weber, A. (2016a). Learning to comprehend foreign-accented speech by means of production and listening training. *Language Learning, 66*(S2), 187-209. doi:10.1111/lang.12174

Grohe, A.-K., & Weber, A. (2016b). The *Penefit* of salience: salient accented, but not unaccented words reveal accent adaptation effects. *Frontiers in Psychology*, *7*(864). doi:10.3389/fpsyg.2016.00864

Gut, U. (2012). The LeaP corpus. A multilingual corpus of spoken learner German and learner English. In: T. Schmidt, & K. Wörner (Eds.), *Multilingual corpora and multilingual corpus analysis* (pp. 3–23). Amsterdam: John Benjamins.

Hanulíková, A., & Weber, A. (2012). Sink positive: linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception, & Psychophysics*, *74*(3), 613–629. doi:10.3758/s13414-011-0259-7

Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: an investigation of English /r/-/l/ perception and production by adult Japanese speakers. In M. Nacano (Ed.), *Interspeech Workshop on Second Language studies: Acquisition, Learning, Education and Technology.* Tokyo: Waseda University.

Hayes-Harb, R., Nicol, J. L., & Barker, J. (2010). Learning the phonological forms of new words: effects of orthographic and auditory input. *Language and Speech*, *53*(3), 367–381. doi:10.1177/0023830910371460

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality. In S. N. Spencer (Ed.), *Proceedings ETRA 2012. Eye Tracking Research and Applications Symposium: Santa Barbara, CA, March 28-30, 2012* (p. 45). New York: Association for Computing Machinery. doi:10.1145/2168556.2168563

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482. doi:10.1016/j.jml.2007.02.001

Huettig, F., & McQueen, J. M. (2008). Retrieval and use of components of lexical knowledge depend on situational demands. *Paper presented at the AMLaP 2008 conference in Cambridge, UK*.

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. doi:10.1016/j.jml.2007.11.007

James, W. (1890). *The principles of psychology*. New York: Holt.

Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruents in German*. *Studies in functional and structural linguistics* (Vol. 44). Amsterdam, Philadelphia: J. Benjamins.

Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent* (Doctoral dissertation, University of Stuttgart).

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *Journal of the Acoustical Society of America*, *138*(2), 817–832. doi:10.1121/1.4926561

Keating, G. D. (2013). Eye-tracking with text. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 69-92). New York: Routledge.

Kerswill, P. & Williams, A. (2002). "Salience" as an explanatory factor in language change: evidence from dialect levelling in urban England. In M. C. Jones & E. Esch (Eds.), *Language change: the interplay of internal, external and extra-linguistic factors* (pp. 81–110). Berlin, New York: Walter de Gruyter.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of Human Memory* (pp. 331–370). New York: Academic Press.

Kittredge, A. K., & Dell, G. S. (2016). Learning to speak by listening: Transfer of phonotactics from perception to production. *Journal of Memory and Language*, *89*, 8–22. doi:10.1016/j.jml.2015.08.001

Kleiner, S., & Knöbl, R. (2015). *Duden, das Aussprachewörterbuch* (6., überarbeitete und aktualisierte Auflage). *Der Duden in zwölf Bänden: Vol. 6*. Berlin: Dudenverlag.

Kohler, K. (1999). Illustrations of the IPA: German. In *Handbook of the international phonetic association* (pp. 86–89). Cambridge: Cambridge University Press.

Kolers, P. A. (1973). Remembering operations. *Memory & Cognition, 1*(3), 347–355. doi:10.3758/BF03198119

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1–15. doi:10.1016/j.jml.2006.07.010

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests in Linear Mixed Effects Models.* Retrieved from http://CRAN.R-project.org/package=lmerTest

Labov, W. (1998). The three dialects of English. In M. D. Linn (Ed.), *Handbook of dialects and language variation* (pp. 39–81). San Diego: Academic Press.

Ladefoged, P. (1999). Illustrations of the IPA: American English. In *Handbook of the international phonetic association* (pp. 41–44). Cambridge: Cambridge University Press.

Landaulab (2017). Time precision in PsychoPy. Retrieved from http://www.landaulab.com/experimentation-tips.html

Larraza, S., Samuel, A. G., & Onederra, M. L. (2016a). Listening to Accented Speech in a Second Language: First Language and Age of Acquisition Effects. *Journal of experimental psychology. Learning, memory, and cognition.* Advance online publication. doi:10.1037/xlm0000252

Larraza, S., Samuel, A. G., & Onederra, M. L. (2016b). Where do dialectal effects on speech processing come from? Evidence from a cross-dialect investigation. *Quarterly Journal of Experimental Psychology* *70*(1), 92-108. doi:10.1080/17470218.2015.1124896

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. doi:10.3758/s13428-011-0146-0

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1–38.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. *ACL-MIT Press series in natural-language processing*. Cambridge, MA: MIT Press.

Liberman, A. M. (1957). Some Results of Research on Speech Perception. *Journal of the Acoustical Society of America*, *29*(1), 117-123. doi:10.1121/1.1908635

Liberman, A. M. (1996). *Speech: a special code*. Cambridge, MA: MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. doi:10.1037/h0020279

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. doi:10.1016/0010-0277(85)90021-6

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187–196. doi:10.1016/S1364-6613(00)01471-6

Lombardi, L. (2003). Second language data and constraints on manner: explaining substitutions for the English interdentals. *Second Language Research*, *19*(3), 225–250. doi:10.1177/026765830301900304

MacLeod, B. (2012). *The Effect of Perceptual Salience on Phonetic Accommodation in Cross-Dialectal Conversation in Spanish* (Doctoral Dissertation, University of Toronto).

MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review*, *18*(6), 1197–1202. doi:10.3758/s13423-011-0168-8

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671-685. doi:10.1037/a0018785

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE, 7*(8), e43230. doi:10.1371/journal.pone.0043230

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language and Hearing Research*, *50*(4), 940-967. doi:10.1044/1092-4388(2007/067)

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543–562. doi:10.1080/03640210802035357

McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. In *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 1619–1624).

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. doi:10.1016/0010-0285(86)90015-0

McDaniel, M. A., & Geraci, L. (2006). Encoding and retrieval processes in distinctiveness effects: Toward an integrative framework. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and Memory* (pp. 65–88). Oxford: Oxford University Press.

McLennan, C. T., & González, J. (2012). Examining talker effects in the perception of native- and foreign-accented speech. *Attention, Perception, & Psychophysics*, *74*(5), 824–830. doi:10.3758/s13414-012-0315-y

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 539–553. doi:10.1037/0278-7393.29.4.539

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126. doi:10.1207/s15516709cog0000_79

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, *60*(5), 661–671. doi:10.1080/17470210601183890

Meyer, A. S., Huettig, F., & Levelt, W. J. (2016). Same, different, or closely related: What is the relationship between language production and comprehension? *Journal of Memory and Language, 89*, 1–7. doi:10.1016/j.jml.2016.03.002

Moosmüller, S., & Ringen, C. (2004). Voice and aspiration in Austrian German plosives. *Folia Linguistica*, *38*(1-2). doi:10.1515/flin.2004.38.1-2.43

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge: Cambridge University Press.

Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition, 17,* 17–34.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *49*(S1), 285–310. doi:10.1111/0023-8333.49.s1.8

Nguyen, N., Dufour, S., & Brunellière, A. (2012). Does imitation facilitate word recognition in a non-native regional accent? *Frontiers in Psychology*, *3.* doi:10.3389/fpsyg.2012.00480

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234. doi:10.1016/0010-0277(94)90043-4

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357–395. doi:10.1037/0033-295X.115.2.357

Ooms, K., Dupont, L., Lapon, L., & Popelka, S. (2015). Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental set-ups. *Journal of eye movement research*, *8*(1), 1-24. doi:10.16910/jemr.8.1.5

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: the production effect endures and improves memory for text. *Memory, 20*(7), 717–727. doi:10.1080/09658211.2012.699070

Peirce, J. W. (2008). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in Neuroinformatics*, *2*. doi:10.3389/neuro.11.010.2008

Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011)* (pp. 161–164). Rundle Mall: Causal Productions.

Pickel, K. L., & Staller, J. B. (2012). A perpetrator's accent impairs witnesses' memory for physical appearance. *Law and Human Behavior, 36*(2), 140-150. doi:10.1007/s10979-011-9263-7

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. doi:10.1017/S0140525X12001495

Popelka, S., Stachon, Z., Sasinka, C., & Dolezalova, J. (2016). EyeTribe Tracker Data Accuracy Evaluation and Its Interconnection with Hypothesis Software for Cartographic Purposes. *Computational intelligence and neuroscience*, *2016*(9172506). doi:10.1155/2016/9172506

R development core team. (2017). *R: a language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing, Retrieved from https://www.r-project.org/, versions 3.2.2 to 3.3.3.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2000). Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions. In M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (2nd ed., pp. 539–552). Cambridge, MA: MIT Press.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192. doi:10.1146/annurev.neuro.27.070203.144230

Robb, M. P., Maclagan, M. A., & Chen, Y. (2004). Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics & Phonetics, 18*(1), 1–15, doi: 10.1080/0269920031000105336

Roediger, H. L. 3., Gallo, D. A., & Geraci, L. (2002). Processing approaches to cognition: the impetus from the levels-of-processing framework. *Memory, 10*(5-6), 319–332. doi:10.1080/09658210224000144

Schmidt, A. M., & Flege, J. E. (1995). Effects of speaking rate changes on native and nonnative speech production. *Phonetica*, *52*(1), 41–54. doi:10.1159/000262028

Schirmunski, V. (1928). Die schwäbischen Mundarten in Transkaukasien und Südukraine. *Teuthonista*. (H. 1), 38–60.

Schirmunski, V. (1962). *Deutsche Mundartkund.: Vergleichende Laut-und Formenlehre der deutschen Mundarten.* Berlin: Akademie-Verlag.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*(1), 86–102. doi: 10.1016/0749-596X(90)90011-N

Schuerman, W. L., Meyer, A., & McQueen, J. M. (2015). Do We Perceive Others Better than Ourselves? A Perceptual Benefit for Noise-Vocoded Speech Produced by an Average Speaker. *PLoS ONE, 10*(7), e0129731. doi:10.1371/journal.pone.0129731

Sereno, J., Lammers, L., & Jongman, A. (2016). The relative contribution of segments and intonation to the perception of foreign-accented speech. *Applied Psycholinguistics*, *37*(02), 303–322. doi:10.1017/S0142716414000575

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America*, *125*(5), 3306-3316. doi:10.1121/1.3101452

Siebs, T., Boor, H. de, Moser, H., & Winkler, C. (1969). *Siebs deutsche Aussprache: Reine und gemäßigte Hochlautung mit Ausspracheworterbuch*. Berlin: de Gruyter.

Siegel, J. (2010). *Second dialect acquisition*. Cambridge: Cambridge University Press.

Spiekermann, H. (2008). *Sprache in Baden-Württemberg: Merkmale des regionalen Standards*: Walter de Gruyter.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. doi:10.3758/BF03207704

Stetson, R. H. (1951). *Motor phonetics*. Amsterdam: North-Holland Publishing Company.

Stevens, K. N. (1960). Toward a Model for Speech Recognition. *Journal of the Acoustical Society of America*, *32*(1), 47-55. doi:10.1121/1.1907874

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. doi:10.1016/j.tics.2009.06.003

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2013). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology, 4,* 1015. doi:10.3389/fpsyg.2013.01015

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*(4), 487–501. doi:10.1016/j.jml.2009.01.001

Swan, M. & B. Smith (2001). *Learner English: A Teacher's Guide to Interference and other Problems*. Cambridge: Cambridge University Press, 37-40.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(01), 1–30. doi:10.1017/S0272263106060013

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes, 27*(7-8), 979–1001. doi:10.1080/01690965.2011.597153

Trude, A. M., Tremblay, A., & Brown-Schmidt, S. (2013). Limitations on adaptation to foreign accents. *Journal of Memory and Language*, *69*(3), 349–367. doi:10.1016/j.jml.2013.05.002

Trudgill, P. (1986). *Dialects in contact*. Oxford: Blackwell.

Tzeng, C. Y., Alexander, J. E. D., Sidaras, S. K., & Nygaard, L. C. (2016). The role of training structure in perceptual learning of accented speech. *Journal of Experimental Psychology. Human Perception and Performance*, *42*(11), 1793–1805. doi:10.1037/xhp0000260

van Els, T., & de Bot, K. (1987). The role of intonation in foreign accent. *The Modern Language Journal*, *71*(2), 147–155. doi: 10.1111/j.1540-4781.1987.tb01597.x

Vogt, F. E. (1977). *Schwäbisch in Laut und Schrift: eine ergründende und ergötzliche Sprachlehre*. Stuttgart: Steinkopf.

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica*, *64*(2-3), 122–144. doi:10.1159/000107913

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989–994. doi:10.1016/S0028-3932(02)00316-0

Weber, A., di Betta, A. M., & McQueen, J. M. (2014). Treack or trit: adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of Phonetics*, *46*, 34–51. doi:10.1016/j.wocn.2014.05.002

Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics*, *39*(4), 479–491. doi:10.1016/j.wocn.2010.12.004

Weber, A., Melinger, A., & Lara Tapia, L. (2007). The mapping of phonetic information to lexical presentations in Spanish: Evidence from eye movements. In J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1941-1944). Dudweiler: Pirrot.

Weil, S. (2001). Foreign accented speech: encoding and generalization. *Journal of the Acoustical Society of America*, *109*(5), 2473. doi:10.1121/1.4744779

Wells, J. C. (1982). *Accents of English: Vol I: An Introduction*. Cambridge: Cambridge University Press.

Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development,7* (1), 49–63.

Wertsch. (1979). *The concept of activity in Soviet psychology*. New York: M.E. Sharpe, Inc.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665–679. doi:10.1016/j.jml.2011.12.010

Witteman, M. J., Bardhan, N. P., Weber, A., & McQueen, J. M. (2015). Automaticity and stability of adaptation to a foreign-accented speaker. *Language and Speech*, *58*(2), 168–189. doi:10.1177/0023830914528102

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*(3), 537–556. doi:10.3758/s13414-012-0404-y

Witteman, M. J., Weber, A., & McQueen, J. M. (2014). Tolerance for inconsistency in foreign-accented speech. *Psychonomic Bulletin & Review*, *21*(2), 512–519. doi:10.3758/s13423-013-0519-8

Zamuner, T. S., Morin-Lessard, E., Strahm, S., & Page, M. P. (2016). Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language, 89*, 55–67. doi:10.1016/j.jml.2015.10.003

Zusammenfassung

Die vorliegende Dissertation untersucht in welchem Ausmaß Hören verglichen mit Produzieren von spezifischen Wörtern mit einem Akzent zum Lernen dieser Wörter mit Akzent und zu generalisiertem Akzentlernen beitragen kann. Außerdem wurde die Rolle von Salienz im Prozess des Akzentlernens hinterfragt. In zehn Experimenten, die in fünf Kapiteln beschrieben werden, wurde diese Fragestellung untersucht.

Kapitel 2A (Experimente 1–3) untersucht die Effekte von Training mit Produktion oder Hören auf das Verstehen gesprochener Sprache mit einem Fremdsprachenakzent. In drei Experimenten wurde ein Training-Test Paradigma umgesetzt. Während des Trainings hörten amerikanisch-englische Muttersprachler (L1 Probanden) und deutsche Muttersprachler (L2 Probanden) eine englische Kurzgeschichte, in der jedes /θ/ mit einem /t/ ersetzt wurde (z.B. wurde *theft* als **teft* ausgesprochen), eine zweite Gruppe laß die Geschichte selbst laut mit den *th*-Ersetzungen vor und eine dritte Grupe hatte kein Akzenttraining. In der darauf folgenden lexikalischen Entscheidungsaufgabe entschieden die Probanden, ob es sich bei auditiv präsentierten Wörtern mit den *th*-Ersetzungen um ein Wort des Englischen handelte oder nicht. Das Trainings- und Testmaterial wurde mit zwei verschiedenen Sprecherinnen aufgenommen, um Generalisierungseffekte über Sprecherinnen hinweg zu testen. L2 Probanden reagierten schneller auf Wörter, die im Training vorkamen, wenn sie die Wörter selbst produziert haben, als ohne Training. Die Reaktionszeiten der Hörgruppe beschleunigten sich auch, allerdings war dieser Effekt weniger stark. Neue Wörter, die nicht im Training vorkamen, waren von diesem Effekt nicht betroffen. Für die L1 Probanden hat das Training die Akzentverarbeitung nicht vereinfacht und somit gab es in dieser Gruppe auch keinen Unterschied zwischen Produktions- und Hörtraining. Folglich bestand ein Vorteil des Produktionstrainings verglichen mit dem Hörtraining nur für die L2 Probanden. Produktionstraining beeinflusst Verstehensprozesse und es kann Akzentlernen sogar mehr beeinflussen als Hörtraining. Diese Effekte beschränken sich aber auf die Zweitsprache eines Sprechers.

Die Unterschiede in der Muttersprache zwischen den aufgenommenen Testsprechern (L1 Deutsch) und den Probanden (L1 Amerikanisches Englisch) könnte erklären, warum die L1 Probanden in Kapitel 2A den Akzent nicht gelernt haben. Diese Hypothese wurde in Kapitel 2B (Experiment 4) getestet. Die Frage ist, ob Akzentlernen einfacher ist, wenn der Sprecher einen ähnlichen Akzent wie der Hörer hat. Das gleiche Training-Test Paradigma wie in den Experimenten 1–3 wurde umgesetzt, aber die Aufnahmen für Training und Test wurden von amerikanisch-englischen Muttersprachlerinnen gemacht. Somit hatten die Probandinnen und die Sprecherinnen denselben L1-Hintergrund. Reaktionszeiten in der lexikalischen Entscheidungsaufgabe waren für beide Trainingsgruppen (Produktion oder Hören) signifikant schneller als die der Kontrollgruppe ohne Training. Im Unterschied dazu zeigten L1 Probanden in Kaptel 2A keinen Trainingseffekt. Das bedeutet, dass Akzentlernen mit Produktion oder Hören durch die Ähnlichkeit zwischen Sprechern und Hörern beeinflusst wird. Diese Ähnlichkeit ist dann gegeben, wenn Sprecher und Hörer den gleichen muttersprachlichen Hintergrund haben.

Kapitel 3A (Experimente 5–6) untersucht das Akzentlernen von L1 Probanden genauer. Die beiden Experimente testen, wie stark ein Akzent mit Produktions- und Hörtraining gelernt wird. In einem Training-Test Paradigma wurden deutsche Muttersprachler mit einem künstlichen deutschen Akzent getetstet. In Experiment 5 laß die Produktionsgruppe während des Trainings einzelne deutsche Wörter laut vor und ersetzte dabei den stimmhaften Plosivlaut am Wortanfang mit seinem stimmlosen Pendant (z.B. wurde *Balken* ausgesprochen als **Palken*). Die Hörgruppe hörte aufgenommene Wörter mit demselben Akzent. Danach wurden in einer Blickbewegungsmessungsaufgabe die Blicke zu auditiv präsentierten Zielwörtern analysiert. Trainings- und Testwörter wurden von zwei verschiedenen deutschen L1-Sprechern aufgenommen. Beide Trainingsgruppen fixierten die Zielwörter mit dem Akzent öfter als eine Kontrollgruppe ohne Training. In Experiment 6 wurde das gleiche Training durchgeführt, aber während des Tests wurden statt Wörtern mit Akzent kanonische deutsche Wörter als Zielwörter präsentiert, die am Wortanfang mit den Wörtern mit Akzent überlappten (z.B. ***Pal**me* überlappt mit dem Akzentwort **Pal**ken*). Dieses Mal gab es keinen Trainingseffekt; Das Akzentlernen, das in Experiment 5 gefunden wurde, beeinflusste nicht das Erkennen der kanonischen Wörter. Das bedeutet, dass der Akzent gelernt wurde, aber nicht so stark, dass Akzentwörter als Kompetitoren fungieren konnten. Die Akzentform wurde nicht aktiviert, wenn ein potentieller kanonischer Kompetitor auditiv präsentiert wurde. Dass Lerneffekte nur für Wörter mit Akzent in Experiment 5, aber nicht für kanonische Wörter in Experiment 6 sichtbar waren, wurde außerdem mit *Salienz* erklärt. Akzentlerneffekte waren nur sichtbar, wenn die Testwörter salient genug waren. Das war der Fall für die Testwörter mit Akzent in Experiment 5, aber nicht für die kanonischen Testwörter in Experiment 6. Salienz der Testwörter war in diesen Experimenten so wichtig, weil der Akzent nicht stark genug gelernt wurde, als dass auch weniger saliente kanonische Testwörter die Akzent-Wortform hätten aktivieren können.

Kapitel 3B (Experiment 7) behandelt den methodischen Gesichtspunkt von Blickbewegungsmessungen. In Experiment 7 wurde Experiment 5 mit einer anderen Hardware, dem *EyeTribe* (https://theeyetribe.com/), repliziert. Dieser mobile, günstigere Blickbewegungmesser hat eine maximale Abtastrate von 60 Hz (ein Datenpunkt je 16.67 Millisekunden) und wird ohne Software geliefert. Im Gegensatz dazu kann der *Eyelink 1000* (SR Research Ltd.), der in den Experimenten 5 und 6 verwendet wurde, Augenbewegungen mit einer Abtastrate von bis zu 1,000 Hz (ein

Datenpunkt pro Millisekunde) aufnehmen und wird mit Software geliefert. Experiment 7 zeigt, dass der *EyeTribe* akzeptable Ergebnisse liefert, die selbst bei einer Abtastrate von nur 30 Hz (ein Datenpunkt je 33.3 Millisekunden) mit den zuvor gesammelten *Eyelink*-Ergebnissen vergleichbar sind. Das bedeutet, dass günstige, mobile Technologien mit niedrigen Abtastraten reichen, um verlässliche Ergebnisse in *Visual World* Experimenten zu erhalten. Allerdings geht der *EyeTribe* auch mit einigen Nachteilen einher. Zum Beispiel wird das Gerät nicht mit einer installierten Software geliefert und es bietet nicht die Option der Abweichungs-Korrigierung während einer experimentellen Sitzung. Außerdem konnten bei vorliegender Testung große Datenmengen nicht analysiert werden, weil das Gerät nicht immer die Fixationskoordinaten aufgenommen hat.

Kapitel 4 (Experimente 8–10) untersucht auch L1-Akzente, konzentriert sich aber auf längerfristige Akzenterfahrungen. In einem Training-Test Paradigma wurden die Effekte von Produktions- und Hörtraining und längerfristiger Akzenterfahrung auf das Erinnerungsvermögen von Wörtern mit Akzent untersucht. Deutsche Muttersprachler, die mit schwäbischem Dialekt in Süddeutschland aufgewachsen sind, produzierten oder hörten einzelne Wörter in einer Trainingsphase. Trainingswörter beinhalteten entweder einen schwäbischen Akzentmarker, der den Probanden bekannt war (/st/ als /ʃt/ ausgesprochen, wie in **Zahnbür/ʃt/e*) oder einen ihnen unbekannten norddeutschen Akzentmarker (/ʃt/ als /st/ ausgesprochen, wie in **Blumen/st/rauß*). Nach dem Training übten die Probanden eine visuelle Gedächtnisaufgabe aus, in der sie für einzelne Wörter aus dem Training und neue Wörter entschieden, ob sie bekannt oder neu waren. D-Primes (ein Maß, das korrekt erkannte alte Wörter und fälschlerweise als alt akzeptierte neue Wörter berücksichtigt) waren signifikant höher (was auf bessere Erinnerungsleistung schließen lässt) für produzierte Wörter als für gehörte Wörter, und zwar in beiden Akzentbedingungen (bekannter schwäbischer Akzent und unbekannter norddeutscher Akzent). Das gleiche Muster wurde gefunden, wenn die Probanden sich während des Produzierens nicht selbst hören konnten. Wenn die Testphase erst eine Woche nach dem Training durchgeführt wurde, wurden produzierte Wörter immer noch mit höherer Wahrscheinlichkeit wieder erkannt als gehörte Wörter. Kapitel 4 zeigt, dass L1 Wörter mit Akzent mit Produzieren einfacher gelernt werden als mit Hören. Dieser Vorteil lässt sich auf das Produzieren alleine zurückführen, trifft auf verschiedene Akzente zu und ist langanhaltend. Der Effekt könnte außerdem auf die Gestaltung der Trainingslisten zurückzuführen sein. Jeder Proband produzierte einige Wörter und hörte andere Wörter während des Trainings, war also nicht auf eine der beiden Trainingsmodalitäten beschränkt (wie in den vorherigen Kapiteln).

Insgesamt zeigen die Experimente dieser Dissertation, dass Akzente mit Hören und Produzieren gelernt werden können. Lernen wurde auf verschiedenen Verarbeitungsebenen beobachtet, was auf die Allgemeinheit dieser Effekte schließen lässt. Lerneffekte über Sprecherinnen hinweg zeigen, dass die Akzentinformation nicht dauerhaft auf lexikalischer Ebene gespeichert wird, sondern auf abstrakten, prälexikalischen Regeln basiert. In den vorliegenden Experimenten gab es nicht immer eine Generalisierung des Produktionsvorteils hin zu neuen Wörtern, zum Beispiel in Experiment 2. Wahrscheinlich braucht Generalisierung und dadurch Regelbildung in diesen Fällen längeres und intensiveres Training. Außerdem kann die Sichtbarkeit von Lernprozessen durch Salienz modifiziert werden.

Wenn gewisse Voraussetzungen erfüllt sind, hilft Produktionstraining mehr als Hörtraining. Diese Voraussetzungen bestehen darin, ein L2 Sprecher zu sein oder das Akzenttraining mit verschiedenen Modalitäten (Produzieren *und* Hören) durchzuführen. Verschiedene Trainingsmodalitäten erhöhen den direkten Kontrast zwischen einzelnen Lern-Tokens und erhöhen somit die Salienz der produzierten Wörter. Der Vorteil durch Produktion wird daher mit Salienz erklärt. Akzentlernen mit Produzieren basiert auf denselben Prinzipien wie Lernen mit Hören. Das setzt identische lexikalische Repräsentationen für Produzieren und Verstehen voraus. Der Unterschied zwischen Lernen mit Produktion und Hören resultiert aus der Stärke des Lernens. Produzieren stärkt durch seine höhere Salienz die Aktivierung einer lexikalischen Repräsentation und das erleichtert die künftige Abrufbarkeit des Wortes. Langfristige Akzenterfahrung beeinflusst den Produktionsvorteil nicht. Der Produktionsvorteil hält mindestens eine Woche und ist immer noch zu beobachten, wenn der Sprecher sich während des Produzierens nicht selbst hört.

Curriculum Vitae

Ann-Kathrin Grohe was born in Heilbronn, Germany, on January 2nd, 1988. She completed her Bachelor's degree in French studies and Psychology at the University of Freiburg, Germany, in 2010. During the winter term 2009/10, she took linguistic classes at the Université de Toulouse II as an Erasmus student. From September 2010 on, Ann-Kathrin studied linguistics at the Université de Montréal, Canada, as an exchange student for two semesters (funded by the Baden-Württemberg Stiftung), and then commenced her Master studies in Speech and Language Processing at the University of Konstanz, Germany. She completed this study path in 2013 with the thesis *Familiarity effects of voice and intonation on lexical processing*, supervised by Prof. Bettina Braun. After a six-month internship at the Volkswagen Foundation in Hannover, she started her PhD studies with Prof. Andrea Weber at the English Department of the University of Tübingen. During her PhD, she completed several research stays at the University of Maryland, United States, funded by the Maryland-Tübingen cooperation (*Language structures in German and English*). In late 2016, she spent six weeks at the MARCS Institute, Western Sydney University, Australia, to foster collaborations between Anne Cutler's group and the SpeechNet BaWü, a network of speech researchers affiliated to the southern German Universities of Tübingen, Konstanz, Stuttgart, and Freiburg. This collaboration was funded by the German Academic Exchange Service (DAAD), *Online processing of prosodic structure across languages*.

Appendix

Appendix A *(refers to Chapter 2A and 2B)*

Table A1: *Screenshots of the background story shown to the production groups in Experiments 1, 2, and 3. Every orthographic instance of "th" was highlighted yellow. The listening group saw the same story, but the "th"s were not highlighted.*

1/6

King Thrushbeard

A king named George had a daughter whose youth and beauty shadowed a thousand others in the kingdom. But at the same time – and this is true – she was so proud and arrogant that no candidate from throughout the kingdom was good enough for her. She rejected one after the other, ridiculing them as well.

[ENTER]

2/6

One Thursday, the king sponsored a great feast, and invited from far and near all healthy men being worth to be considered as the daughter's husband. They came from the North, East, South and West. They were all placed in a row according to their rank and standing. Then the king's daughter was led through the ranks, but she objected to something about each one. One was too fat: "The wine barrel," she said. Another was too tall: "Thin and tall, no good at all." The third was too short: "Short and thick is never quick." The fourth was too pale: "As pale as death." The fifth had no thumb: "Missing finger, not a winner!" The sixth was not very straight: "Green wood, dried behind the stove."

[ENTER]

3/6

And thus she had some objection to each one, but she ridiculed especially one good king who stood at the very top of the row. He had a fine face, but his chin had grown a little crooked. "Look!" she cried out, laughing, "He has a chin like a thrush's beak." And from that time he was called "Thrushbeard". Now the old king, seeing that his daughter did nothing but making fun of all the candidates who were gathered there, became very angry, and he swore that she should have for her husband the very first beggar to come to his door.

[ENTER]

4/6

A month later a minstrel came and sang beneath the window. When King George heard him he said, "Let him come up, don't throw him out. The risk is low that he commits a theft." So the minstrel, in his dirty, ragged clothes, came in and sang before the king and his daughter. He was hungry and thirsty. Therefore, when he was finished he begged for some food and water. George said, "My servants will prepare a meal for you. You can fill your mouth with it. But also, I liked your song so much that I will give you my daughter for a wife – for the rest of your life!"

[ENTER]

5/6

The king's daughter took fright, "I prefer another man, I need a therapist", she said. But George said, "I have sworn to give you to the very first beggar, and I think I will keep it." Her protests did not help and the minstrel accepted the offer. He was very thankful. The priest was called in, and she had to marry the minstrel at once. After that had happened King George said, "It is not proper for you, a beggar's wife, to stay in my palace any longer. All you can do now is to go away with your husband."

[ENTER]

6/6

The beggar led her out by the hand, and she had to leave with him. Walking on foot along dirty paths, they came to a large forest, and she asked,

"Who owns these beautiful trees?"

"They belong to King Thrushbeard. If you had taken him, it would be yours."

"Oh, I am a miserable thing; if only I'd taken the Thrushbeard King."

[the end - ENTER]

Table A2: *Critical words in Experiments 1, 2, and 3.*

Original word form	t-replacement
thankful	tankful
Thanksgiving	Tanksgiving
theft	teft
thematic	thematic
theory	teory
therapist	therapist
thesis	tesis
thickness	thickness
thief	tief
thing	ting
think	tink
thinner	tinner
third	tird
thirsty	tirsty
thirty	tirty
threaten	treaten
threshold	threshold
thriller	triller
throat	troat
throughout	troughout
throw	trow
thumb	tumb
thunder	tunder
Thursday	Tursday

Table A3.1: *Endorsement rates in Experiment 1, L2 learners of English (L1 German)*

	Control Group		Listening Group		Production Group	
all critical items	53.5%		61.7%		46.2%	
split by familiarity	26.5% old	27% new	29.9% old	31.9% new	21.5% old	24.7% new

Table A3.2: *Endorsement rates in Experiment 2, L2 learners of English (L1 German)*

	Control Group		Listening Group		Production Group	
all critical items	84%		79.8%		73%	
split by familiarity	41.9% old	42.1% new	42.1% old	37.7% new	35.8% old	37.2% new

Table A3.3: *Endorsement rates of accented words in Experiment 3, L1 American English*

	Control Group		Listening Group		Production Group	
all critical items	76.4%		74.5%		74.8%	
split by familiarity	38.0 % old	38.4% new	40.3% old	34.2% new	39.0% old	35.8% new

Table A3.4: *Endorsement rates of word fillers in Experiment 3, L1 American English*

	Control Group		Listening		Production	
all word fillers	94.8%		95.9%		93.6%	
split by familiarity	47.6% old	47.2% new	48.3% old	47.6% new	48.6% old	45.1% new

Table A3.5: *Endorsement rates of accented words in Experiment 4, L1 American English*

	Control Group		Listening		Production	
all critical items	81.6%		85.5%		91.6%	
split by familiarity	41.8% old	39.8% new	43.4% old	42.1% new	46.6% old	45.0% new

Table A4.1: *Reaction times from Experiment 1 (as reported, refers to Figure 2-1).*

```
model1=lmer(logRT~Training*Familiarity+wavdur+(1+Familiarity|Subject)+(1|Item),
                                                              data=rtexp1)
                    Number of obs: 890, groups: Subject, 72; Item, 24
                         Intercept: Training=Control, Familiarity=Old
```

Fixed effects	*Estimate*	*Std. Error*	*df*	*t-value*	*p-value*
Intercept	7.2173471	0.1711023	42.59	42.181	< 2e-16 ***
TrainingListen	-0.2658813	0.0939318	58.83	-2.831	0.006348 **
TrainingProd	-0.3173454	0.0971117	66.03	-3.268	0.001722 **
FamiliarityNew	-0.1377198	0.0843006	40.92	-1.634	0.109999
Wavdur	-0.0011606	0.0002643	33.95	-4.391	0.000104 ***
TrainingListen:FamiliarityNew	0.2047448	0.0812812	54.53	2.519	0.014733 *
TrainingProd:FamiliarityNew	0.2006684	0.0868229	62.80	2.311	0.024114 *

Table A4.1thief: *Reaction times from Experiment 1 (without thief).*

```
model1b=lmer(logRT~Training*Familiarity+wavdur+(1+Familiarity|Subject)+(1|Item),
                                    data=rtexp1, subset = Item != "thief")
                    Number of obs: 834, groups:  Subject, 71; Item, 23
                         Intercept: Training=Control, Familiarity=Old
```

Fixed effects	*Estimate*	*Std. Error*	*df*	*t-value*	*p-value*
Intercept	7.3169950	0.1667453	38.93	43.881	< 2e-16 ***
TrainingListen	-0.2611617	0.0950826	57.35	-2.747	0.00803 **
TrainingProd	-0.3259868	0.0981153	64.75	-3.322	0.00147 **
FamiliarityNew	-0.0880568	0.0831201	40.84	-1.059	0.29565
Wavdur	-0.0013282	0.0002572	29.98	-5.164	1.47e-05 ***
TrainingListen:FamiliarityNew	0.2079850	0.0847884	50.69	2.453	0.01765 *
TrainingProd:FamiliarityNew	0.1797483	0.0904202	58.38	1.988	0.05152 .

Table A4.2: *Reaction times from Experiment 2 (as reported, refers to Figure 2-2).*

```
model2=lmer(logRT~Training*Familiarity+Voice+wavdur+(1|Subject)+(1|Item),
                                                              data=rtexp2)
                    Number of obs: 1332, groups: Subject, 73; Item, 24
             Intercept: Training=Control, Familiarity=Old, Voice=VoiceA
```

Fixed Effects	*Estimate*	*Std. Error*	*df*	*t-value*	*p-value*
Intercept	7.146	0.1895	70.4	37.719	< 2e-16 ***
TrainingListen	-0.1033	0.07776	91.0	-1.329	0.18716
TrainingProd	-0.1821	0.07989	98.5	-2.280	0.02479 *
FamiliarityNew	-0.02312	0.09584	30.3	-0.241	0.81102
VoiceB	-0.1033	0.05888	66.6	-1.754	0.08406 .
Wavdur	-0.001403	0.0002946	68.3	-4.763	0.0000104 ***
TrainingListen:FamiliarityNew	0.2650	0.06329	1241	4.187	0.0000303 ***
TrainingProd:FamiliarityNew	0.2128	0.06566	1246	3.242	0.00122 **

Table A4.2thief: *Reaction times from Experiment 2 (without thief).*

```
model2b = lmer(logRT~Training*Familiarity+Voice+wavdur+ (1|Subject)+(1|Item),
                                       data=rtexp2, subset = Item != "thief")
                         Number of obs: 1265, groups:  Subject, 73; Item, 23
                              Intercept: Training=Control, Familiarity=Old
```

Fixed Effects	*Estimate*	*Std. Error*	*df*	*t-value*	*p-value*
Intercept	7.200e+00	1.926e-01	6.930e+01	37.374	< 2e-16 ***
TrainingListen	-1.021e-01	7.919e-02	8.760e+01	-1.289	0.200657
TrainingProd	-1.752e-01	8.139e-02	9.440e+01	-2.153	0.033880 *
FamiliarityNew	2.933e-03	9.796e-02	2.960e+01	0.030	0.976311
VoiceB	-9.881e-02	6.055e-02	6.630e+01	-1.632	0.107432
Wavdur	-1.501e-03	3.012e-04	6.670e+01	-4.983	4.71e-06 ***
TrainingListen:FamiliarityNew	2.594e-01	6.508e-02	1.176e+03	3.986	7.14e-05 ***
TrainingProd:FamiliarityNew	2.251e-01	6.748e-02	1.180e+03	3.336	0.000876 ***

Table A4.3: *Reaction times from Experiment 2 and 3 in one analysis (refers to Figure 2-3).*

```
model3=lmer(logRT~Training*Familiarity*native+wavdur+(1|Subject)+(1|Item),
                                                          data=rtexp2_3)
                        Number of obs: 2371, groups: Subject, 133; Item, 24
                 Intercept: Training=Control, Familiarity=New, Native=GER
```

Fixed Effects	*Estimate*	*Std. Error*	*df*	*t-value*	*p-value*
Intercept	6.817	0.1843	104.4	36.991	< 2e-16 ***
TrainingListen	0.1610	0.07907	172.1	2.036	0.043238 *
TrainingProd	0.03221	0.07984	174.5	0.403	0.626031
FamiliarityOld	0.04804	0.09758	30.3	0.492	0.389679
nativeENG	0.07175	0.08319	172.4	0.862	0.000156 ***
wavdur	-0.001005	0.0002573	118	-3.907	0.000156 ***
TrainingListen:FamiliarityOld	-0.2610	0.06205	2218	-4.207	0.000027***
TrainingProd:FamiliarityOld	-0.2146	0.06440	2228	-3.333	0.000874 ***
TrainingListen:nativeENG	-0.2587	0.1187	179.4	-2.179	0.030604 *
TrainingProd:nativeENG	-0.1203	0.1189	178.5	-1.012	0.313092
FamiliarityOld:nativeENG	-0.1573	0.06671	2218	-2.358	0.018448 *
TrainingListen:FamiliarityOld:nativeENG	0.2477	0.09509	2220	2.605	0.009252 **
TrainingProd:FamiliarityOld:nativeENG	0.2755	0.09641	2224	2.858	0.004303 **

Table A4.4: *Original reaction times from Experiment 4 (refers to Figure 2-4).*

```
model4=lmer(logRT~Training*Familiarity+scale(wavdur)+scale(ItemOrder)+
(1+scale(ItemOrder)|Subject)+(1+scale(ItemOrder)+Voice|Item) ,data=rtexp4)
Number of obs: 1160, groups: Subject, 59; Item, 24
Intercept: Training=Control, Familiarity=Old
```

Fixed Effects	*Estimate*	*Std. Error*	*df*	*t-value*	*p-value*
Intercept	6.05390	0.08843	59.7	68.458	< 2e-16 ***
TrainingListen	-0.30504	0.09051	75.9	-3.370	0.001184 **
TrainingProd	-0.15547	0.09010	74.5	-1.725	0.088598 .
FamiliarityNew	-0.02285	0.09960	30.7	-0.229	0.820030
scale(wavdur)	-0.10484	0.04030	35.8	-2.602	0.013410 *
scale(ItemOrder)	-0.05747	0.02283	24.9	-2.517	0.018651 *
TrainingListen:FamiliarityNew	0.14056	0.07011	1012.2	2.005	0.045260 *
TrainingProd:FamiliarityNew	0.22792	0.06902	1008	3.302	0.000992 ***

Table A4.5: *Reaction times to filler words from Experiment 4 (refers to Figure 2-5).*

```
model5=lmer(logRT~Training+Familiarity+scale(wavdur)+(1|Subject)+(1|Item),
data=rtexp4_filler)
Number of obs: 3318, groups: Subject, 59; Item, 60
Intercept: Training=Production
```

Fixed Effects	*Estimate*	*Std. Error*	*df*	*t-value*	*p-value*
Intercept	6.512	0.126	400.7	51.684	< 2e-16 ***
TrainingControl	-0.005644	0.08091	55.9	-0.070	0.9446
TrainingListen	-0.2123	0.08097	56	-2.622	0.0112 *
scale(wavdur)	-0.001344	0.0001862	434.4	-7.218	2.37e-12 ***

Table A4.6: *Adjusted Reaction times from Experiment 4 (refers to Figure 2-6).*

```
model6=lmer(logRT_adapted~Training*Familiarity+scale(wavdur)+scale(ItemOrder)
+(1|Subject)+(1|Item),data=rtexp4)
Number of obs: 1160, groups: Subject, 59; Item, 24
Intercept: Training=Control, Familiarity=Old
```

Fixed Effects	*Estimate*	*Std. Error*	*df*	*t-value*	*p-value*
Intercept	6.05661	0.08817	59.7	68.694	< 2e-16 ***
TrainingListen	-0.17323	0.09024	75.9	-1.920	0.058663 .
TrainingProd	-0.17998	0.08983	74.5	-2.003	0.048765 *
FamiliarityNew	-0.02284	0.09930	30.7	-0.230	0.819626
scale(wavdur)	-0.10438	0.04018	35.8	-2.598	0.013536 *
scale(ItemOrder)	-0.05734	0.02277	24.9	-2.518	0.018598 *
TrainingListen:FamiliarityNew	0.14023	0.06990	1012.2	2.006	0.045113 *
TrainingProd:FamiliarityNew	0.22733	0.06881	1008	3.304	0.000988 ***

Table A4.7: *Reaction times from Experiment 3 and 4 in one analysis.*

```
model7=lmer(logRT_adapted~Training*Familiarity*Speaker_Nativeness+scale(wavdur)
                                 +scale(ItemOrder) +(1|Subject)+(1|Item),data=rtexp3_4)
                                     Number of obs: 2205, groups: Subject, 119; Item, 25
                  Intercept: Training=Control, Familiarity=Old, Speaker_Nativenes=GER
```

Fixed Effects	*Estimate*	*Std. Error*	*df*	*t-value*	*p-value*
Intercept	6.16173	0.08942	62.3	68.908	< 2e-16 ***
TrainingListen	-0.13454	0.08878	154.3	-1.515	0.13170
TrainingProd	-0.05032	0.08922	157.1	-0.564	0.57357
FamiliarityNew	0.1209	0.10111	26.3	1.196	0.24248
Speaker_NativenessENG	-0.0762	0.09617	141.8	-0.792	0.4295
scale(wavdur)	-0.13331	0.03189	61.7	-4.180	9.35e-05 ***
scale(ItemOrder)	-0.05405	0.01597	27.7	-3.385	0.00215 **
TrainingListen:Familiaritynew	0.06164	0.07020	1944.5	0.878	0.38005
TrainingProd:Familiaritynew	-0.02898	0.07004	1950.4	-0.414	0.67909
TrainingListen:speakerENG	-0.05434	0.12546	150.2	-0.433	0.66555
TrainingProd:speakerENG	-0.13103	0.12554	150.5	-1.044	0.29827
FamiliarityNew:speakerENG	-0.15073	0.08144	71.7	-1.851	0.06833 .
TrainingListen:FamiliarityNew:speakerENG	0.07909	0.09728	1933	0.813	0.41631
TrainingProd:FamiliarityNew:speakerENG	0.26083	0.09651	1936.2	2.703	0.00694 **

Appendix B *(refers to Chapter 3A and 3B)*

Table B1: *Critical target-competitor pairs for Experiments 5 and 6. Voiced items were used as targets in Experiment 5 (initial plosive was devoiced, i.e., /b/→/p/ and /g/→/k/), and voiceless items were competitors. In Experiment 6, voiceless items were targets, and voiced items were competitors.*

	voiced			voiceless		
	German Token	IPA transcript	English translation	German Token	IPA transcript	English translation
bilabial	Butter	ˈbʊtɐ	butter	Putzer	ˈpʊtsɐ	cleaner
	Bistum	ˈbɪstu:m	diocese	Piste	ˈpɪstə	ski slope
	Beize	ˈbaitsə	marinade/stain	Peitsche	ˈpaitʃə	whip
	Beifall	ˈbaifal	acclaim	Peiler	ˈpailɐ	detector
	Baron	baˈro:n	baron	Paris	paˈri:s	Paris
	Balken	ˈbalkn̩	beam	Palme	ˈpalmə	palm tree
	Becher	ˈbɛçɐ	beaker/mug	Pächter	ˈpɛçtɐ	tenant
	Benzin	bɛnˈtsi:n	gas	Pension	pɛnˈzi̯o:n	guest house/pension
	Bilanz	biˈlants	balance	Pilot	piˈlo:t	pilot
	Ballett	baˈlɛt	ballet	Palast	paˈlast	palace
	Banner	ˈbanɐ	banner	Panne	ˈpanə	breakdown
	Bazille	baˈʦɪlə	bacillus	Pazifik	paˈtsi:fɪk	Pacific
	Bettler	ˈbɛtlɐ	beggar	Petzer	ˈpɛtsɐ	telltale
	Bauwerk	bauvɛrk	building	Pause	ˈpauzə	break
velar	Gorilla	goˈrɪla	gorilla	Korea	koˈre:a	Korea
	Gulasch	ˈgʊlaʃ	goulash	Kuli	ˈkʊli	ballpoint pen
	Galerie	galəˈri:	gallery	Kalorie	kaloˈri:	calorie
	Gasthaus	ˈgasthaus	guest house	Kasten	ˈkastn̩	box
	Gürtel	ˈgʏrtl̩	belt	Kürzung	ˈkʏrtsʊŋ	abridgement
	Gitter	ˈgɪtɐ	grid/fencing	Kittel	ˈkɪtl̩	tunic
	Ganove	gaˈno:və	crook	Kanone	kaˈno:nə	cannon/rod
	Gammler	ˈgamlɐ	loafer	Kammer	ˈkamɐ	small room/chamber
	Germane	gɛrˈma:nə	Teuton	Keramik	kɛˈra:mɪk	ceramics
	Geltung	ˈgɛltʊŋ	validity/prestige	Kälte	ˈkɛltə	cold
	Garant	gaˈrant	guarantor	Karat	kaˈra:t	carat
	Garage	gaˈra:ʒə	garage	Karaffe	kaˈrafə	carafe
	Gassenjunge	ˈgasənˈjʊŋə	street urchin	Kassenzettel	ˈkasənˈtsɛtl̩	(sales) receipt
	Gartenzaun	ˈgartn̩tsaun	garden fence	Kartenspiel	ˈkartn̩ʃpi:l	game of cards

Appendix C *(refers to Chapter 4)*

Table C1: *Critical familiar (Swabian) and unfamiliar (northern German) accent words and filler words used in Experiments 8, 9, and 10.*

	Familiar/Swabian Accent (st → ʃt)			Unfamiliar/northern German Accent (ʃt → st)		
	German Token	IPA transcript	English translation	German Token	IPA transcript	English translation
critical (2-syllabic)	Brustkorb	/brʊstkɔrp/	chest	Altstadt	/altʃtat/	old city
	Christbaum	/krɪstbaʊm/	Christmas tree	Grabstein	/graːpʃtaɪn/	tombstone
	Dienstplan	/di:nstplaːn/	service plan	Bleistift	/blaɪʃtɪft/	pencil
	Festmahl	/fɛstmaːl/	banquet	Kraftstoff	/kraftʃtɔf/	fuel
	Gasthaus	/gasthaʊs/	restaurant	Bahnsteig	/baːnʃtaɪk/	train platform
	Herbsttag	/hɛrpstaːk/	autumn day	Fahrstuhl	/faːrʃtuːl/	elevator
	Kunstwerk	/kʊnstvɛrk/	artwork	Baumstamm	/baʊmʃtam/	tree trunk
	Nistplatz	/nɪstplat͡s/	nesting site	Kuhstall	/kuːʃtal/	cow barn
	Postamt	/pɔstˀamt/	post office	Kleinstaat	/klaɪnʃtaːt/	microstate
	Rasthof	/ʀasthoːf/	service area	Erbstück	/ɛʁpʃtʏk/	heirloom
	Restmüll	/ʀɛstmʏl/	trash	Impfstoff	/ɪmpfʃtɔf/	vaccine
	Testfahrt	/tɛstfaːɐrt/	trial run	Diebstahl	/diːpʃtaːl/	theft
	Trostpreis	/troːstpraɪs/	booby prize	Laufsteg	/laʊfʃteːk/	catwalk
	Wurstbrot	/vʊʁstbʀoːt/	sliced sausage (German) sandwich	Handstand	/hantʃtant/	handstand
critical (3-syllabic)	Angsthase	/aŋsthaːzə/	scardey cat, coward	Badestrand	/baːdəʃtʀant/	bathing beach
	Bierkasten	/biːrkastən/	beer case	Baustelle	/baʊʃtɛlə/	construction site
	Geisterbahn	/gaɪstɐbaːn/	haunted house	Besenstiel	/beːzənʃtiːl/	broomstick
	Hausmeister	/haʊsmaɪstɐ/	building caretaker	Blumenstrauß	/bluːmənʃtʀaʊ̯s/	bouquet
	Hustensaft	/huːstənzaft/	cough sirup	Edelstein	/eːdəlʃtaɪn/	gemstone
	Meisterschaft	/maɪstɐʃaft/	championship	Hungerstreik	/hʊŋɐʃtraɪk/	hunger strike
	Mistgabel	/mɪstgaːbl/	pitchfork	Klapperstorch	/klapɐʃtɔʁç/	stork
	Obstmesser	/oːpstmɛsɐ/	fruit knife	Lenkstange	/lɛŋkʃtaŋə/	handlebars
	Osterei	/oːstɐˀaɪ/	Easter egg	Liegestuhl	/liːgəʃtuːl/	deck chair
	Preisliste	/praɪslɪstə/	price list	Lippenstift	/lɪpənʃtɪft/	lipstick
	Sandwüste	/zantvyːstə/	sand desert	Türsteher	/tyːrʃteːɐ/	bouncer
	Schaufenster	/ʃaʊfɛnstɐ/	shop window	Wanderstock	/vandɐʃtɔk/	hiking pole
	Waldmeister	/valtmaɪstɐ/	woodruff (plant)	Wasserstrahl	/vasɐʃtraːl/	water jet

Table C1 (continued)

	Familiar/Swabian Accent (st → ʃt)			Unfamiliar/northern German Accent (ʃt → st)		
	German Token	IPA transcript	English translation	German Token	IPA transcript	English translation
filler (2-syllabic)	Bergbahn	/bɛrkbaːn/	mountain railway	Ohrring	/oːrɪŋ/	earring
	Einhorn	/aɪnhɔrn/	unicorn	Nachthemd	/naxthɛmt/	nightdress
	Fahrrad	/faːɐ̯raːt/	bicycle	Bargeld	/baːrgɛlt/	cash
	Fernrohr	/fɛrnroːɐ̯/	telescope	Rollmops	/ʁɔlmɔps/	pickled herring
	Glatteis	/glatˀaɪs/	black ice	Frühling	/fryːlɪŋ/	spring
	Goldfisch	/gɔltfɪʃ/	goldfish	Pfarrhaus	/pfaʁhaʊs/	rectory
	Trickfilm	/trɪkfɪlm/	animated cartoon	Dampfbad	/dampfbaːt/	steam bath
	Volkslied	/fɔlksliːt/	folksong	Weissbrot	/vaɪsbroːt/	white bread
filler (3-syllabic)	Hintertür	/hɪntetyːr/	backdoor	Herdplatte	/heːrtplatə/	hotplate
	Kofferraum	/kɔferaʊm/	car trunk	Buchhalter	/buːxhalte/	accountant
	Minigolf	/minɪgɒlf/	mini golf	Baumwolle	/baʊmvɔlə/	cotton
	Oberarm	/oːbeˀaːrm/	upper arm	Tintenfisch	/tɪntnfɪʃ/	octopus
	Rosenkohl	/roːznkoːl/	Brussels sprout	Segelboot	/zeːglboːt/	sailboat
	Sommerkleid	/zɔmeklaɪt/	summer dress	Blutprobe	/bluːtproːbə/	blood sample
	Taschenbuch	/taʃnbuːx/	paperback book	Winterschlaf	/vɪnteʃlaːf/	hibernation
	Turnhalle	/tʊʁnhalə/	gymnasium	Wochenmarkt	/vɔxnmaʁkt/	weekly market

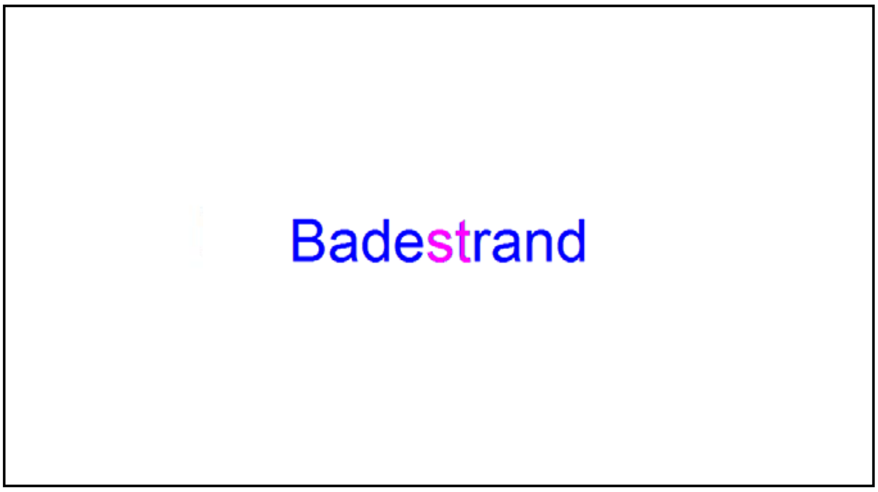

Figure C1: *Example screen from the training phase in Experiments 8, 9, and 10. For the critical word "Badestrand" in the production condition (unfamiliar northern accent) participants received the instruction to pronounce "st" as /st/.*

List of figures

List of tables