Similarity in global accent promotes generalized learning of accent markers

Ann-Kathrin Grohe, Margarita Downing, Andrea Weber

University of Tübingen

ann-kathrin.grohe@uni-tuebingen.de, margarita.pelz@student.uni-tuebingen.de, andrea.weber@uni-tuebingen.de

Abstract

Is it easier to adapt to 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, English L1 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, i.e., 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 [8].

Index Terms: native accent, foreign accent, adaptation, accent similarity, accent strength, auditory word recognition

1. 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 [1], English L1 listeners overcame initial processing difficulties for sentences spoken by Spanish- and Chinese-accented speakers within one minute of listening to these speakers. In [2], listening to a 4-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 [3] for comparable results). Adaptation to native accents has been investigated, for example, in [4]. Students originally from Northern England adapted to Standard Southern British English within two years of their university studies in Southern England. Adaptation to native accents 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 [5]. Adaptation was, however, speaker-specific, i.e., it was restricted to the training speaker and did not generalize readily to new speakers.

Thus, adaptation to foreign- and native-accented speech occurs, but it is not clear yet if the underlying processes are the same. One possibility is that adaptation 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 adaptation ease. Support for the role of accent strength comes, for example, from [2] who found that priming effects were smaller for strongly-accented words than for weakly-accented words. On the other hand, [6] 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 English L1 speaker. This suggests more tolerant adaptation to foreign-accented speech than to native-accented speech. Further support for this notion was found in [7]. Dutch L1 listeners adapted to 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 adaptation process.

In all of the above studies, adaptation occurred through listening. However, producing an accent can also form the basis for adaptation. This was tested recently in [8]. In an exposure-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*s) with /t/ (e.g., *theft* became **teft*), or they read the same story aloud themselves with the instruction to substitute all *th*s with /t/. Neither German learners of English [9] nor the tested population of English L1 speakers typically replace *th* 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, English L1 participants showed no adaptation, 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 adaptation 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 a recent eye-tracking study [10] in which German L1 participants adapted to an accent marker (devoicing of word-initial voiced stops; e.g., *Balken* ‘beam’ became **Palken*) both after having listened to a German L1 speaker produce the accent and after having produced the accent themselves. If this pattern of results transfers to English L1 participants and to a lexical decision task with reaction times as dependent measure, then English L1 participants should adapt to an accent marker well enough to generalize effects to a new speaker using the materials in [8], but only if the materials are produced by English L1 speakers.

In summary, prior research has shown that adaptation to 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 adaptation.

This was investigated in the present study. English L1 participants performed the same production and listening training and subsequent word recognition task as in [8]. That is, they either listened to an English story in which all *th*s were produced 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 English rather than by German learners of English. Thus, the accent marker was the same as in [8], 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 [8] generalized accent learning across speakers because their own accent was similar enough to the pre-recorded speaker's accent. It was expected that English L1 participants in the present study would adapt to 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 was 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.

2. Experiment

2.1. Participants

Fifty-nine female students (18-26 years old) from the University of Maryland, all L1 American English speakers, were tested.

2.2. 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 with /θ/ in initial position were chosen as critical words (mean frequency: 163.6 occurrences per million according to the CELEX word form dictionary [11]. When /θ/ was replaced with /t/, the resulting word forms were nonwords (*theft* became **teft*). Twelve of the critical words were taken from the exposure 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 adaptation to the accent 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 words and half nonwords.

The selected material was the same as in [8], but the stimuli were recorded by English L1 speakers rather than by German learners of English as in [8]. 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 1. *Critical old words from the training story and critical new words.*

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

2.3. 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 18.3, 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 of [8], 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.

2.4. Results

Analyses were conducted with the software R (version 3.2.4, www.r-project.org). Endorsement rates and reaction times for accepted critical words (i.e., yes responses to **teft*) were analyzed with linear mixed effects models [12]. Reaction times, measured from word offset, between 85-1900 ms long were included in the analyses (5% outliers) and were log-normalized. For each analysis, we built an individual, best fitting model 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.

2.4.1. *Endorsement Rates*

Endorsement rates were on average 86% (listening group 86.1%; production group 91.6%; control group 81.5%). 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 (β=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 [8], focus will therefore be placed on reaction time analyses, which are particularly suitable to provide information about online processing.

2.4.2. *Reaction Times*

Reaction times for accepted critical tokens (i.e., yes responses to **teft*) 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 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. 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).

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 (60-1500 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. Statistical analyses showed a significant main effect of training group (χ^2=12.2, p<.04). 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%).

With the new, adjusted reaction times, the training*familiarity interaction was still significant (χ^2=10.9, p<.005). Training effects were observed for old, but not new items (see Figure 1). The listening group accepted old items faster than the control group (β=-0.17, *SE*=0.09, *t*=-1.9), and the production group was significantly faster than the control group (β=-0.18, *SE*=0.09, *t*=-2). Reaction times for new items were not affected by accent training (all *t*-values<|0.5|). Next, we analyzed our data together with the L1 data from [8]. 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 [8]. The new, large dataset consequently included the additional variable *speaker nativeness* (L1 vs. L2 English). The same model as above was run, replacing the two-way interaction with the three-way interaction of training*familiarity*speaker nativeness. Adding speaker nativeness significantly improved the model (χ^2=12.2, p<.04). This confirms that speaker nativeness was the critical factor that provoked the training effects.

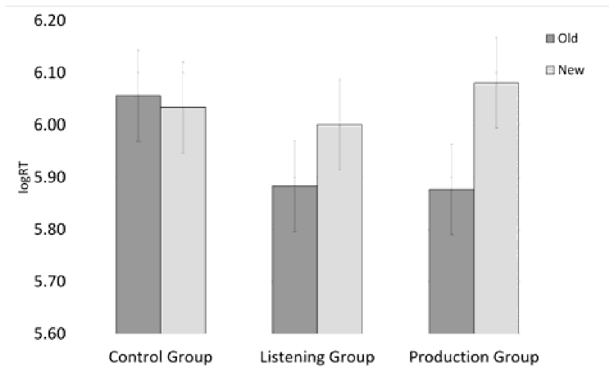

Figure 1. *Adjusted Reaction Times (with standard errors) to old and new items in the Listening Group, the Production Group, and the Control Group without training.*

3. 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 [8]. Joint analyses of the present data and the analogous L1 experiment in [8] confirmed that accent adaptation 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 adapted to 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 [8], 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 [8] 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 ([3], 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, i.e., whether an accent can be learned from one speaker and be applied to a second speaker with the same native language background. We found that generalization across speakers is possible as long as both speakers have similar accent properties. Adaptation across speakers was also found with L2 participants in [8], suggesting that accent adaptation is not necessarily speaker-specific as was suggested in [5]. This is in line with further studies [13–15] that suggest that accent adaptation 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., [16, 17]), we assume 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 [3] 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 [8], [3] 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 in [10]. 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 [8]? 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 adaptation with listening training is supported by [2]. Mild accents are more easily learned than stronger accents. Accent strength is further emphasized in the accent processing classification that was postulated in [1]. 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 adaptation. When learning through listening, stronger accents need more time or more intense training. In [8], 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.

4. 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.

5. References

[1] C. M. Clarke and M. F. Garrett, "Rapid adaptation to foreign-accented English," *Journal of the Acoustical Society of America*, vol. 116, no. 6, pp. 3647–3658, 2004.

[2] M. J. Witteman, A. Weber, and J. M. McQueen, "Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation," *Attention, Perception, & Psychophysics*, vol. 75, no. 3, pp. 537–556, 2013.

[3] M. J. Witteman, N. P. Bardhan, A. Weber, and J. M. McQueen, "Automaticity and stability of adaptation to a foreign-accented speaker," *Language and Speech*, vol. 58, no. 2, pp. 168–189, 2015.

[4] B. G. Evans and P. Iverson, "Plasticity in vowel perception and production: a study of accent change in young adults," *Journal of the Acoustical Society of America*, vol. 121, no. 6, pp. 3814–3826, 2007.

[5] A. M. Trude and S. Brown-Schmidt, "Talker-specific perceptual adaptation during online speech perception," *Language and Cognitive Processes*, vol. 27, no. 7-8, pp. 979–1001, 2012.

[6] F. Eisner, A. Melinger, and A. Weber, "Constraints on the transfer of perceptual learning in accented speech," *Frontiers in Psychology*, vol. 4, 2013.

[7] M. J. Witteman, A. Weber, and J. M. McQueen, "Tolerance for inconsistency in foreign-accented speech," *Psychonomic Bulletin & Review*, vol. 21, no. 2, pp. 512–519, 2014.

[8] A.-K. Grohe and A. Weber, "Learning to comprehend foreign-accented speech by means of production and listening training," *Language Learning*, in press.

[9] A. Hanulíková and A. Weber, "Sink positive: linguistic experience with th substitutions influences nonnative word recognition," *Attention, Perception, & Psychophysics*, vol. 74, no. 3, pp. 613–629, 2012.

[10] A.-K. Grohe and A. Weber, "The Penefit of salience: salient accented, but not unaccented words reveal accent adaptation effects," *Frontiers in Psychology*, 2016.

[11] R. H. Baayen, R. Piepenbrock, and L. Gulikers, The CELEX lexical database (Release 2). Philadelphia: University of Pennsylvania, 1995.

[12] R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language*, vol. 59, no. 4, pp. 390–412, 2008.

[13] A. R. Bradlow and T. Bent, "Perceptual adaptation to non-native speech," *Cognition*, vol. 106, no. 2, pp. 707–729, 2008.

[14] C. T. McLennan and J. González, "Examining talker effects in the perception of native- and foreign-accented speech," *Attention, Perception, & Psychophysics*, vol. 74, no. 5, pp. 824–830, 2012.

[15] S. Weil, "Foreign accented speech: encoding and generalization," *Journal of the Acoustical Society of America*, vol. 109, no. 5, p. 2473, 2001.

[16] J. L. McClelland and J. L. Elman, "The TRACE model of speech perception," *Cognitive Psychology*, vol. 18, no. 1, pp. 1–86, 1986.

[17] D. Norris, "Shortlist: A connectionist model of continuous speech recognition," *Cognition*, vol. 52, no. 3, pp. 189–234, 1994.