Phonetic-to-lexical mapping in listening to adult and child speech

[1]Thanh Lan Truong, [2]Ulrike Schild, [2]Claudia Friedrich, [1]Andrea Weber

[1]English Department, [2]Department of Psychology
University of Tübingen Germany
thanh-lan.truong@uni-tuebingen.de, ulrike.schild@uni-tuebingen.de, claudia.friedrich@uni-tuebingen.de,
andrea.weber@uni-tuebingen.de

ABSTRACT

The mapping of phonetic information to lexical representations in adult and child speech was examined using cross-modal priming. Native adult listeners were presented with German word fragments (e.g., *Para-* from *Parasit*, 'parasite') that mismatched in the second vowel with a visual target word (e.g., *Parodie*, 'parody'). Word fragments were spoken by a female adult speaker or a 7-year-old child. Overall, response times were faster following fragment primes spoken by the child. For word pairs with the vowels /u:/, /i:/, and /a:/, fragments spoken by the adult speaker primed target recognition with a directional asymmetry, while no priming was found for fragments spoken by the child speaker. This suggests that phonetic-to-lexical mapping is possibly sensitive to the age of the speaker. The results are interpreted as an effect of previous experience with the linguistic competence of child and adult speakers.

Keywords: Lexical activation, vowel mismatch, child speech, age of speaker, cross-modal priming.

1. INTRODUCTION

Phonetic-to-lexical mapping involves computing from a continuous speech signal the information that identifies matching words in the mental lexicon [e.g., 13, 29]. This process would be most economical if a mismatch between input and lexical representations leads to an immediate rejection of mismatching candidate words. And while it is certainly true that lexical activation is sensitive to all linguistic information, empirical evidence suggests that the process is not quite as parsimonious since activation of candidate words can be found despite a partial mismatch [e.g., 10, 19].

By and large, models of spoken-word recognition assume that the mapping process from phonetic input to lexical representations is not sensitive to social aspects, such as to which group of individuals the speaker of the input belongs to [26]. Recent research on foreign-accented speech suggests that this may not always be the case. Foreign-accented speakers typically deviate in their pronunciation from the norms of the target language [e.g., 20] and listeners can adjust their comprehension in line with the properties of the foreign-accented productions such that the same deviations are treated differently depending on the nativeness of the speaker [4, 8, 12]. It has been argued that experience with the source properties helps to adjust the comprehension process from the outset when encountering a (new) foreign-accented speaker.

Foreign-accented speakers are not the only speakers that recurrently deviate in their pronunciation from the norms. Children are "unreliable" speakers too, with a lower linguistic competence than native adult speakers [e.g., 21]. In the present study, we tested if the phonetic-to-lexical mapping process is sensitive to the age of the speaker. Listeners can recognize the approximate age of a speaker from his or her voice quite easily [e.g., 15], and age attributed to a speaker has previously been found to shift listeners' perception of vowels that are currently undergoing a chain shift in a language [7] and to influence listeners' interpretation of conceptual messages [24].

The aim of the present study was to investigate if adult native listeners map phonetic information to lexical representations differently in child speech and in adult speech. In a cross-modal fragment priming experiment, German listeners heard word onset fragments as primes (e.g., *Para-* from *Parasit*) before they had to decide if visually displayed target words (e.g., *Parodie*) were existing words of German or not. Prime and target words overlapped in onset but mismatched in the vowel of the second syllable (e.g., /a:/ in *Para-* and /o:/ in *Parodie*). Primes were either produced by a 7-year old child or by an adult speaker. If speaker age influences phonetic-to-lexical mapping, then the same mismatches in vowels were predicted to result in different priming effects, depending on speaker age.

2. METHOD

2.1. Participants

Thirty-one native listeners of German (21 female), all students at the University of Tübingen and between 18 and 30 years old (mean age = 23.5, SD = 3.4), participated in the experiment for monetary compensation.

2.2. Material

Fifty-six German word pairs from [9] were used as experimental items. The two words of a pair had the same stress pattern and overlapped segmentally in onset but mismatched in the vowel of the second syllable (e.g., *Parasit-Parodie*, 'parasite-parody). Across word pairs, the mismatching vowels differed in vowel height, backness, and roundedness, and represented the majority of German monophthongs [27]. A total of 16 different vowel mismatches were included. The onset fragments of one word of a pair (e.g., *Para-* from *Parasit*) always served as prime for the other word (e.g., *Parodie*). Both words of a pair functioned as fragment prime for the other word in different experimental lists (e.g., *Paro-* also served as prime for *Parasit*). Taking stress and vowel quality into consideration, onset fragments were never existing German words and only matched up with their carrier word in German. Since asymmetries in vowel perception (e.g., /oː/-/aː/ being less confusable than /aː/-/oː/) have been shown to affect lexical activation [e.g., 5, 6], direction of vowel mismatch was coded in the present experiment. For confusable mismatches, the mismatch might be opaque and not preclude (pre-)activation of the target word, while for dissimilar vowels the mismatch might preclude target activation.

Eighty phonotactically legal nonword pairs were selected as filler items. In 22 pairs the two onsets overlapped but mismatched in the second vowel, in 22 pairs they were phonologically unrelated, and in 36 pairs they overlapped fully including the second vowel. The onset fragment of one nonword of a pair served as prime for the other nonword.

All words and nonwords were recorded by two female native speakers of Standard German who are currently living in Tübingen: a 34-year old adult and a 7-year old child. Recordings were made in a sound-attenuated room with a high-quality microphone and a sampling rate of 44 kHz. While the adult speaker read from orthographic transcriptions, the child was prompted with the adult recordings. Special care was taken that all items were produced as intended. Onset fragments were excised using Praat [3]. The durations of the onset fragments were on average longer in the child voice than in the adult voice (mean child voice = 604 ms; mean adult voice = 555 ms; $t = -2.7$, $p < 0.008$).

2.3. Procedure

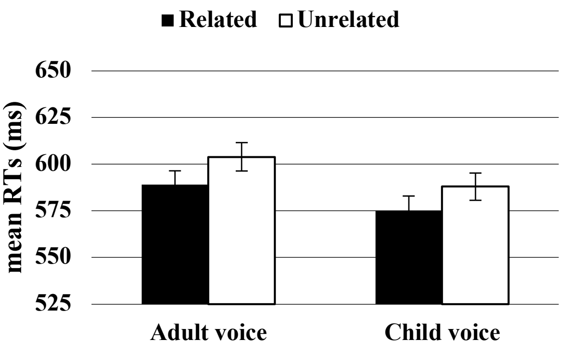
The experiment was controlled with Presentation (version 20.1, www.neurobs.com) and participants were tested individually. In each trial, an onset fragment was played as a prime over headphones, and at its offset a string of letters appeared on the screen. Participants had to indicate with a button press if the string of letters was an existing word of German or not.

In the related condition, the target word was preceded by the spoken onset fragment of its pair member (e.g., prime *Para-* from *Parasit*, target *Parodie*). Both pair members served as target and prime in a Latin-Square design (e.g., prime *Paro-* from *Parodie*, target *Parasit*). In the unrelated condition, the target word was preceded by the spoken onset fragment of a segmentally unrelated word (e.g., prime *Elo-* from *Eloquenz*, 'eloquence', target *Parodie*). All primes used in unrelated trials also served as primes in the related condition (e.g., prime *Elo-*, target *Element*, 'element'). Eight experimental lists with the 56 experimental items and the 80 filler items were created. Each experimental item appeared once in each list, counterbalanced for the role of target, the relatedness of the prime, and the speaker of the prime. The order of item presentation was pseudo-randomized. After the priming task was completed, participants filled in a short language background questionnaire.

3. RESULTS

Only trials with correct responses to target words were analysed. Participants answered on average 84.4 % correctly when the primes had been produced by the adult, and 84.0 % correctly when the primes had been produced by the child. Thus, neither the task nor the different speakers posed considerable difficulties and performance did not differ for the two speakers. Reaction times (RTs) faster than 250 ms and slower than 1200 ms were excluded (0.1 % of the data).

Figure 1: Mean RTs (in ms) following related and unrelated primes, presented in adult and child voice. The vertical bars represent standard errors.



R [16] and lme4 [2] were used to perform linear mixed effects analyses on log-normalized RTs. The full model included *relatedness* (related, unrelated) and *age of speaker voice* (adult, child), as well as *direction of vowel mismatch*, *lexical frequency of the target*, and *target word length* as fixed factors.

Participants and *items* were included as random factors with random slopes. A backward stepwise selection was applied when no model improvement was observed [1].

After stepwise selection, the final model showed a facilitatory effect of *relatedness* (β = -4.98, SE = 1.81, t = -2.75, p < .007), faster RTs for primes for the child *speaker* (β = -4.41, SE = 1.80, t = -2.44, p < .02), an effect of *lexical frequency* (β = -2.50, SE = 9.22, t = -2.27, p < .007), and marginally non-significant interactions between *direction* and *speaker age* (β = 4.71, SE = 2.54, t = 1.85, p < .07) between *direction* and *relatedness* (β = 6.06, SE = 2.56, t = 2.36, p < .02), and between *speaker age*, *direction*, and *lexical frequency* (β = -0.04, SE = 0.02, t = -1.78, p < .08) see Figure 1).

The interactions called for further analyses. Visual inspection suggested, that especially for the adult speaker, vowel mismatches in prime-target pairs often affected word recognition differently when the role of prime and target was reversed. For example, while *Para-* numerically facilitated recognition of *Parodie*, *Paro-* did not facilitate recognition of *Parasit*. It is well-attested that vowel discriminability can depend on the direction in which vowels are presented [e.g., 5, 17], and lexical activation has been shown to be affected by these perceptual asymmetries [e.g., 6, 10, 25].

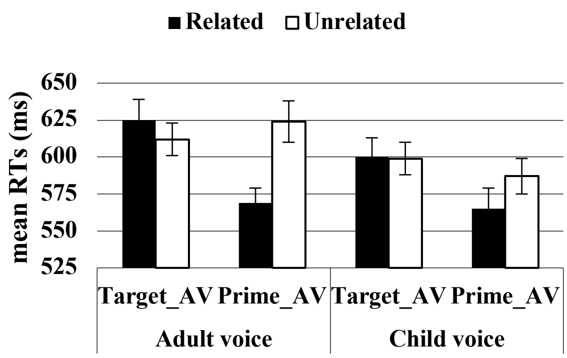
The prime-target pairs in the present study comprised 16 different vowel mismatches, and we found in the literature no theoretically-driven predictions about perceptual asymmetries for the complete set of mismatches. However, the Natural Referent Vowel (NRV) framework by Polka and Bohn [14], suggests that there is a universal default bias for the peripheral vowels /u:/, /i:/, and /a:/ which is especially relevant during language development [see also, 18]. While mature perceivers can adapt this initial bias to optimize access to language-specific vowel categories, a privileged fit of the peripheral vowels with human auditory abilities, ensures that the bias is also relevant for adult listeners and native contrasts.

Using the NRV framework for exploratory interpretation of the results, the German vowels /u:/, /i:/, and /a:/ are anchor vowels in the present study, and a change from an anchor vowel to a non-anchor vowel should be harder to detect than a change in the other direction. In other words, an anchor vowel in the prime should make the vowel mismatch in the target opaquer, while the vowel mismatch should be more transparent when the anchor vowel occurs in the target. In terms of phonetic-to-lexical mapping, the prediction would be that vowel mismatches that are opaque still prime target word recognition, while vowel mismatches that are transparent do not.

In 41 of our 56 target-prime pairs an anchor vowel was involved[1], and for this subset of items the new fixed factor *anchor* coded if the anchor vowel occurred in the prime or in the target. For the adult speaker, an interaction was found between *relatedness* and *anchor* (β = 0.06, SE = 0.03, t = 2.14, p < 0.04), and further analyses showed a facilitatory effect of *relatedness* when the anchor vowel was in the prime (β = -0.07, SE = 0.02, t = -3.25, p < .002), and no effect when the anchor vowel was in the target (β = -0.002, SE = 0.02, t = -0.09, p > .9). For the child speaker, only lexical frequency was significant (β = -3.78, SE = 1.58, t = -2.39, p < .02); *relatedness* did not interact with *anchor* (β = -3.82, SE = 3.04, t = -1.25, p > .2), and was neither significant when the anchor vowel was in the prime (β = -0.03, SE = 0.02, t = -1.49, p > .1) nor when it occurred in the target (β = -0.005, SE = 0.02, t = -0.25, p > .7; see Figure 2).

Note that when analyzing the subset of 41 target-prime pairs for both speakers together, the new factor *anchor* did interact significantly with *relatedness* (β = 0.06, SE = 0.03, t = 2.14, p < .04) but not with *speaker age* (β = 0.03, SE = 0.03, t = 1.21, p > .2), thus in fact not licensing the split for the two speakers. For the complete set of 56 target-prime pairs, interactions involving *direction* (rather than *anchor*), *speaker age,* and *relatedness* did license a split, and further analyses for both speakers showed the exact same pattern of results as was found for the subset with anchor-vowels. Since we found no literature on German vowel confusions that would allow theoretically-driven predictions for all 16 vowel mismatches of the complete set, presenting results based on just the target-prime pairs with anchor vowels seemed appropriate. The pattern of results is however backed up by the analysis of the complete set of target-prime pairs.

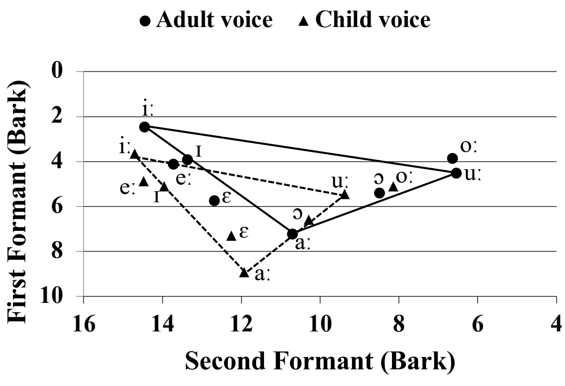
Figure 2: Mean RTs (in ms) for the 41 target pairs with the anchor vowels /u:/, /i:/, or /a:/, when the anchor vowel occurred in the prime (Prime_AV) and when it occurred in the target (Target_AV). The vertical bars represent standard errors.



The results for the adult voice are in line with the predictions we derived from the NRV framework [14]: when the vowel mismatch between prime and target was opaque, the onset fragment of the prime facilitated recognition of the target word (e.g., *Para-*

primed *Parodie)*; when the vowel mismatch was transparent, there was no priming (e.g., *Paro-* did not prime *Parasit*). However, for the child voice, onset fragments never primed target recognition. Possibly, the vowel space of the child speaker was warped, and vowel categories were not distributed as clearly as for the adult speaker. Figure 3 shows averages for the first two formants at the midpoint of the mismatching vowel in the onset fragments (e.g., [aː] in *Para-* and [oː] in *Paro-*), separately for the adult voice and the child voice. For each voice, a total of 82 vowels were measured (41 target-prime pairs X 2 members of each pair). Note, that the number of measurements for each vowel varies considerably, since vowel type was not controlled in the experiment (e.g., 25 vowels [aː] for each speaker and only 3 vowels [ɔ]). As can be expected, the formant values for the child voice were higher, but the overall patterning of vowels in the vowel space seems quite comparable across speakers, certainly with respect to the anchor vowels /uː/, /iː/, and /aː/ [see also, 11]. Also note, that overall recognition rates were equally accurate for the two speakers. An alternative explanation for the different patterns is based on listeners' previous experience with the linguistic competence of adult and child speakers. Young children are known to deviate regularly from target norms in their pronunciation [21], and listeners could take this experience into consideration and hesitate to rely on, for example, vowel information in their interpretation.

Figure 3: Average mid-vowel F1/F2 values (Bark) for all vowels in the subset of 41 target-word pairs with anchor vowels, by the adult speaker and by the child speaker.



4. CONCLUSIONS

In a cross-modal priming experiment, we investigated the influence of age of the speaker on phonetic-to-lexical mapping. In a first analysis, no interaction between facilitation for related primes and age of the speaker was observed. However, marginally non-significant interactions involving age of the speaker, relatedness of the prime, and direction of the vowel mismatch warranted further exploratory analyses. The NRV framework [14] was used to motivate post-hoc predictions for the directionality of vowel mismatches in subsequent exploratory analyses (accounting for a subset of 72.3 % of the items). When taking directionality into account according to the NRV framework, different priming patterns for the adult voice and the child voice were found. For the adult voice, onset fragments that mismatched in the vowel primed target word recognition when the anchor vowels /uː/, /iː/, or /aː/ occurred in the onset fragments (e.g., *Para-* primed *Parodie*) but not when they occurred in the target (e.g., *Paro-* did not prime *Parasit*). No priming was found for the child voice, neither when the anchor vowels occurred in the fragment primes nor when they occurred in the target.

A comparison of the F1 and F2 values of the vowels produced by the two speakers, made it unlikely that the influence of speaker age was due to less accurate productions of the child speaker. It rather seems likely that the phonetic-to-lexical mapping itself was sensitive to the age of speaker. One plausible reason for this could be previous experience with children's speech that often deviates from canonical pronunciations. This experience could set expectations and influence the comprehension process whenever we encounter a (new) child speaker.

The consequence of experience was such that onset fragments with vowel mismatches of the child speaker never facilitated target word recognition. Thus, vowel information in the child voice was never deemed a reliable indicator for the lexical mapping process. Just as well, it could have been that experience leads to all vowel mismatches being accepted as matches for the lexical mapping. In research on foreign-accented speech, previous experience has indeed been found to make deviations in pronunciation acceptable matches for canonical pronunciations [e.g., 8, 23, 29]. Note however that in most of these studies, experience with specific accents and/or single accent markers have been tested, whereas in the present study the vowel mismatches comprised a whole range of contrasts. Also, it is possible that children vary more between and within speakers in pronunciation than speakers of a specific foreign accent tend to do. Thus, it might be much harder to adapt to anything specific in children's mispronunciations.

In conclusion, we present some evidence for the phonetic-to-lexical mapping process being possibly sensitive to the age of the speaker. To our knowledge, this is the first time that such an influence has been shown exploratorily for child speech, and we interpreted it as an effect of previous experience with the linguistic competence of child and adult speakers.

7. REFERENCES

[1] Bates, D., Kliegl, R., Vasishth, S., Baayen, R. H. 2015. Parsimonious mixed models.

[2] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. lme4: Linear mixed-effects models using Eigen and S4: R Package http://CRAN.R-project.org/package=lme4

[3] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.40 http://www.praat.org/

[4] Bosker, H. R., Quené, H., Sanders, T., de Jong, N. H. 2014. Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language 75*, 104-116.

[5] Cutler, A., Weber A., Smits, R., Cooper, N. 2004. Patterns of English phoneme confusions by native and non-native listeners. *J. Acoust. Soc. Am*. 116(6), 3668-3678.

[6] Cutler, A., Weber, A., Otake, T. 2006. Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics* 34(2), 269-284.

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

[8] Eisner, F., Melinger, A., Weber, A. 2013. Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology* 4, 148.

[9] Friedrich, C. K., Felder, V. Lahiri, A., Eulitz, C. 2013. Activation of words with phonological overlap. *Frontiers in Psychology* 4, 556.

[10] Friedrich, C. K., Lahiri, A., Eulitz, C. 2008. Neurophysiological evidence for underspecified representations: asymmetries with word initial variations. *J. Exp. Psychol. Hum. Percept. Perform*. 34, 1545–1559.

[11] Lee, S. Potamianos, A., Narayanan, S. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J. Acoust. Soc. Am*. 105(3), 1455-1468.

[12] Lev-Ari, S. 2015. Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. *Frontiers in Psychology* 5, 1546.

[13] McQueen, J. M., Norris, D., Cutler, A. 1994. Competition in spoken word recognition: Spotting words in other words. *J. Exp. Psychol.: Learn., Mem. Cogn.* 20(3), 621.

[14] Polka, L., Bohn, O. S. 2011. Natural referent vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics* 39(4), 467-478.

[15] Ptacek, P. H., Sander, E. K. 1966. Age recognition from voice. *Journal of Speech, Language, and Hearing Research* 9(2), 273-277.

[16] R Core Team. 2013. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for statistical computing http://www.R-project.org/

[17] Repp, B. H., Crowder, R. G. 1990. Stimulus order effects in vowel discrimination. *J. Acoust. Soc. Am*. 88(5), 2080-2090.

[18] Schwartz, J. L., Abry, C., Boë, L. J., Ménard, L., Vallée, N. 2005. Asymmetries in vowel perception, in the context of the Dispersion–Focalisation Theory *Speech Communication* 45(4), 425-434.

[19] Soto-Faraco, S., Sebastián-Gallés, S., Cutler, A. 2001. Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language* 45, 412-432.

[20] Steinlen, A. K. 2005. *The influence of consonants on native and non-native vowel production: A cross-linguistic study* Tübingen: Gunter Narr.

[21] Stoel-Gammon, C. L., Menn, L. 2005. Phonological development: Learning sounds and sound patterns. In: Gleason, J. B., Ratner, N. B. (eds), *The development of language*. Boston: Pearson, 52-85.

[22] Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., Nishi, K. 2007. Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *J. Acoust. Soc. Am*. 122(2), 111-129.

[23] Trude, A. M., Tremblay, A., Brown-Schmidt, S. 2013. Limitations on adaptation to foreign accents. *Journal of Memory and Language 69*(3), 349-367.

[24] van den Brink, D., van Berkum, J. J., Bastiaansen, M. C., Tesink, C. M., Kos, M., Buitelaar, J. K., Hagoort, P. 2010. Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social Cognitive and Affective Neuroscience* 7(2), 173-183.

[25] Weber, A., Cutler, A. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50(1), 1-25.

[26] Weber, A., Scharenborg, O. 2012. Models of spoken-word recognition. *WIREs Cognitive Science* 3, 387-401.

[27] Wiese, R. 2000. *The phonology of German* Oxford: Oxford University Press.

[28] Wittemann, 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), 537556.

[29] Zwitserlood, P. 1989. The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* 32(1), 25-64.

---

[1] Also German /a/ was considered as anchor vowel /a:/, as the two phonemes are considered to differ only in duration in German [22].