COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 41 (2017) 1106–1118

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12394

Relative Difficulty of Understanding Foreign Accents as a Marker of Proficiency

Shiri Lev-Ari,[a,b] Marieke van Heugten,[b,c] Sharon Peperkamp[b]

[a]*Max Planck Institute for Psycholinguistics*
[b]*Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL Research University*
[c]*Department of Psychology, University at Buffalo, State University of New York*

Received 21 September 2015; received in revised form 16 March 2016; accepted 17 March 2016

Abstract

Foreign-accented speech is generally harder to understand than native-accented speech. This difficulty is reduced for non-native listeners who share their first language with the non-native speaker. It is currently unclear, however, how non-native listeners deal with foreign-accented speech produced by speakers of a different language. We show that the process of (second) language acquisition is associated with an increase in the relative difficulty of processing foreign-accented speech. Therefore, experiencing greater relative difficulty with foreign-accented speech compared with native speech is a marker of language proficiency. These results contribute to our understanding of how phonological categories are acquired during second language learning.

Keywords: Foreign accent; L2 learners

1. Introduction

Our native language shapes the way we categorize the sound space (e.g., Kuhl et al., 2008; Werker & Tees, 1984). Therefore, learning a novel language typically requires acquiring a new manner of carving the sound space, independent of our first language—a task that most language learners find difficult. Native language categories often continue to interfere with the to-be-acquired sound patterns, leading to deviations in second language learners' pronunciations of words relative to the native standard. Such deviations make foreign-accented speech (henceforth, *accented speech*) more difficult to cope with than native speech. That is, despite listeners' abilities to adapt to accented speakers

Correspondence should be sent to Shiri Lev-Ari, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands. E-mail: shiri.lev-ari@mpi.nl

(e.g., Clarke & Garrett, 2004; although see Floccia, Butler, Goslin, & Ellis, 2009; Trude, Tremblay, & Brown-Schmidt, 2013 for counterexamples), accented speech is typically less intelligible and harder to understand than native speech (Munro & Derwing, 1995; van Wijngaarden, 2001). For non-native speakers, language processing is harder to begin with, at all linguistic levels. Therefore, one may expect non-native listeners, especially those who do not share the speaker's first language, to experience even more difficulty processing accented speech. However, contrary to these predictions, we show that early in second language acquisition, the added difficulty that accented speech imposes is relatively small, and that the emergence of an advantage for processing native over non-native speech is in fact a marker of gaining proficiency in the language.

So far, studies examining how non-native listeners process accented speech have focused on the Interlanguage Speech Intelligibility Benefit—the diminution of the relative difficulty of understanding accented speakers compared with native speakers when those accented speakers have the same language background as the listener (e.g., Bent & Bradlow, 2003; Hayes-Harb, Smith, Bent, & Bradlow, 2008; Imai, Walley, & Flege, 2005; Stibbard & Lee, 2006). For example, in an English word identification task, native English speakers reached a RAU[1] accuracy score of 109 with native speech, compared to only 77 with high-proficiency Mandarin-accented speech. Such a difference between native and Mandarin-accented speech did not appear when native Mandarin speakers were tested instead (a RAU of 56 vs. 64, respectively; Bent & Bradlow, 2003). The relative benefit is likely driven by two factors—similar categorization of the sound space by speaker and listeners, and, often, listeners' greater exposure to speakers of their own accent. Importantly, non-native listeners who do not share the first language of the non-native speaker are not more likely than native speakers to categorize the sound space similarly to the non-native speaker, or are they more likely than native speakers to have exposure to the non-native speaker's accent.

To our knowledge, to date, there has been no systematic study examining how non-native listeners process accented speech produced by non-native speakers of a different language than their own. Many assume that this group of listeners would struggle even more than native listeners, as, in general, adverse listening conditions are more detrimental for non-native than for native speakers' performance (see Lecumberri, Cooke, & Cutler, 2010; for a review). For example, Takata and Nábělek (1990) showed that native Japanese listeners who performed similarly to native English ones on an English word identification task in silence (99.3% vs. 97.4%, respectively) performed worse than the native listeners when the words were presented in noise (72.4% vs. 80.3%). This greater disruption is often explained as being due to the fact that language has redundant cues, allowing native (but not non-native) listeners to reconstruct lost information. That is, native listeners might be able to rely on grammatical, lexical, or phonotactic knowledge to narrow down the possibilities of what the distorted or missing information might be, whereas non-native listeners might lack the relevant linguistic knowledge or have fewer resources to devote to reconstruction.

The difficulty associated with comprehending accented speech, however, is qualitatively different from the difficulty present in adverse listening conditions. Indeed, in the case of speech presented in noise, information in the signal is masked, whereas in the case of accented speech, it is mostly present but distorted. In other words, processing speech in noisy conditions requires reconstructing missing information, whereas processing accented speech requires adaptation to or correction of deviations. Importantly, deviations are defined in relation to typical native productions, but the representations that non-native listeners build might be imprecise or vague (at first). As a result, the productions of non-native speakers might not deviate from the listener's representation any more than those of native speakers. For example, neither French nor American English has the Dutch diphthong/ɛɪ/. Consequently, when a French listener hears an American-accented production of the Dutch /ɛɪ/ that is influenced by the English /aɪ/ (the American English category closest to Dutch /ɛɪ/ in phonetic space), she might not perceive it as deviant from a native Dutch production of /ɛɪ/. It is only when the French speaker gains proficiency and forms a more precise representation of the Dutch diphthong that she would start perceiving the American-accented version as deviant, and would therefore process the native Dutch version more easily and accurately than the American-influenced one. In other words, foreign accents may not affect language comprehension during the early stages of second language learning but might instead become progressively harder to cope with as listeners' language proficiency increases. The two contrasting accounts can be viewed in Fig. 1. An account according to which contending with accented speech is similar to contending with adverse listening conditions would predict that increased proficiency should lead to the minimization of the native speaker advantage (Fig. 1a). In contrast, our account predicts that at the beginning, the native speaker advantage would be small, but that it would increase as a function of proficiency level, as enhanced proficiency would mostly help performance with native speech (Fig. 1b). According to our account, it is only at a later stage, when proficiency is much higher, that increased proficiency would allow listeners to deal better with native speech (gray lines in Fig. 1b). Some evidence for our account comes from Bent and Bradlow's (2003) study mentioned

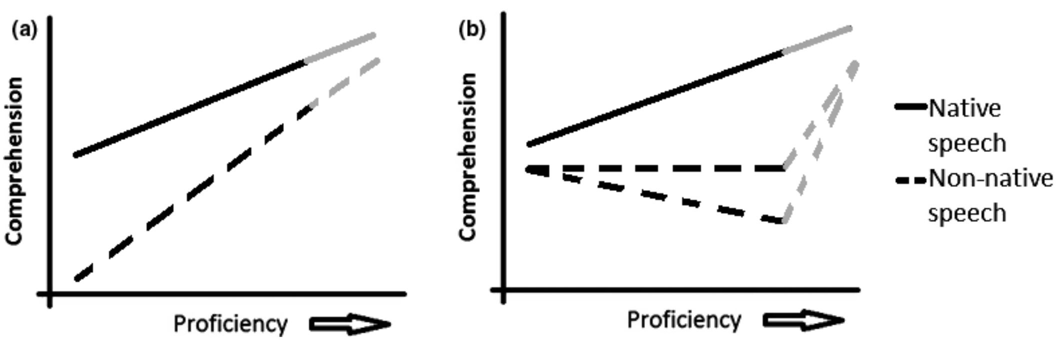

Fig. 1. Predicted performance with native (solid line) and non-native (dashed line) speech according to an account that assumes that accented speech is a type of adverse listening condition (a) and according to our account (b). The black lines illustrate the stage tested in our experiment. The gray lines illustrate what we predict at higher levels of proficiency.

above. In that study, a group of non-native listeners from a diverse set of language backgrounds other than Mandarin and English was also tested, and it was found that for these participants, English spoken by proficient Mandarin (and Korean) speakers is as intelligible as English spoken by native English speakers. The goal of this study is twofold. First, we investigate whether, in contrast to contending with adverse listening conditions, contending with accented speech may not be harder for non-native than for native listeners. Second, we explore whether the poorer performance with non-native compared with native speech is related to proficiency, such that the more proficient a speaker is, the more they would show an advantage for native over non-native speech.

It is important to note that the increase in relative difficulty of understanding foreign accents that we are hypothesizing is relative in nature and involves a comparison with native speech. According to this view, listeners' absolute success at understanding the new language increases with proficiency, but it increases unevenly for native and accented speech. This was indeed the case in Bent and Bradlow (2003). The native listeners there exhibited a greater native speaker advantage, even though their performance was better than that of the non-native listeners on the non-native speakers. That is, the greater native advantage resulted from the fact that native listeners' superiority over the non-native listeners was much larger for the native than the non-native speakers. In general, depending on various factors, such as the sound space of the language and its relation to the listener's and speaker's native languages, performance with accented speech might deteriorate, remain constant, or even slightly increase, but crucially, it should not increase as much as performance with native speech.

2. Experiment

To test whether relative difficulty of understanding accented speech is a result of acquired proficiency in the language, we tested native French speakers' perception of native and accented Dutch words following a short or a longer Dutch learning session, inducing low and higher proficiency. In the short-learning condition, participants learned 12 words in a single round of approximately 10 min. In the long-learning condition, which lasted approximately an hour, participants learned 30 Dutch words over four learning rounds. Participants in the long-learning condition thus learned more Dutch words and heard each word more often than those in the short-learning condition. This resembles gaining second language proficiency in the real world, where more proficient speakers have a larger vocabulary than less proficient speakers, and have often also had greater exposure to each individual word. In both learning conditions, all learning input was recorded by native Dutch speakers. Following the learning rounds, participants were tested on their recognition of the words that they had learned. Crucially, these test words were produced by novel speakers, half of whom were native and half non-native. We predicted that participants would be more likely to recognize the learned words when they are produced by native as compared to non-native speakers of Dutch, but that this native speaker advantage would increase with proficiency, and would thus be larger for participants who were in the long-learning condition than for those who were in the short-learning condition.

2.1. Method

2.1.1. Participants

Ninety-nine native French speakers with no knowledge of either Dutch or German (a closely related language that is taught in French schools) participated in the experiment.[2] One participant fell asleep during the task and was excluded from our analyses, leaving 49 participants in each condition. Another participant quit the task halfway through the test stage, and hence provided only partial data.

2.1.2. Stimuli and procedure

The learning set included 30 monosyllabic words and phonotactically legal nonwords in Dutch. All words contained one of three Dutch vowels (/ɑ/, /eː/, /ɛɪ/) that are phonetically distinct from any French vowel.[3] Each word was represented by a picture, often reflecting the word's real meaning in Dutch.[4] Thirty additional words similar to the target words served as fillers in the test phase (see Appendix A for an overview of the target and filler words). Each target word was recorded by all 12 speakers: four native speakers for the learning phase (2 females, 2 males), four native speakers for the test phase (2 females, 2 males), and four non-native speakers for the test phase (2 females, 2 males). All speakers were recorded in a sound-attenuated booth. None of the non-native speakers was a native speaker of French. Their native languages were Brazilian Portuguese, Canadian English, Italian, and Romanian. The accented test words were selected by the second author, a native Dutch speaker, who ensured that the tokens were particularly accented, yet still identifiable as the intended word. Filler words were only recorded by the eight speakers used at test. In addition, the four speakers of the learning phase recorded sentences for the sentence task (see below).

To keep participants engaged, each round in the learning phase consisted of four different tasks in the order presented below. Participants in the short-learning condition (henceforth, *Short condition*) performed one short-learning round in which they were taught 12 words, followed by the test phase. Participants in the long-learning condition (henceforth, *Long condition*) performed four longer rounds, in which they were taught a total of 30 words, before completing the test. Each round included new words as well as all words taught in previous rounds. The first round of the Long condition consisted of the same set of 12 words that were used in the Short condition. In the remaining three rounds, 18 additional words were taught.

Word in sentence identification: On each trial, participants heard a Dutch sentence followed after about 750 ms by a single word. Their task was to indicate whether the isolated word had appeared in the sentence. In the Long condition, the isolated words included all words that were to be learned during that round, as well as additional words that were never learned. In the Short condition, the isolated words were all ones that were never learned afterward. On half of the trials in both conditions, the isolated words had appeared in the sentence (thus requiring a "yes" response), and on the other half, they had not (thus requiring a "no" response). In total, participants in the Short condition completed 16 trials of this task, and participants in the Long condition completed 120 trials. Each of these 120 trials contained a different sentence followed by a nonrecurring isolated word. Sentences in rounds 2, 3, and 4 thus only contained new isolated words, not the words that participants had started to learn in the preceding blocks.

Word memorization: Each word was presented with its matched picture, with the picture appearing about a second before the word. Participants' task was to memorize the word–picture pairings. In the Short condition, participants only heard one production of the picture's label each time. In the Long condition, participants heard two productions each time the picture was presented; in each round, pictures of new words appeared twice and pictures of words taught in previous rounds appeared once. Participants in the Short condition memorized a total of 12 words, with each word repeating twice. Participants in the Long condition memorized a total of 30 words, with each word repeating between 4 and 10 times, depending on the round in which the word first appeared. In the first round, they memorized the same 12 words as participants in the Short condition; each additional round added six new words, in addition to containing repeated presentations of words from earlier rounds.

Picture matching: On each trial, participants saw four pictures of items on the screen whose names they had learned, and 500 ms later, they heard a word that matches one of these pictures. Their task was to click on the picture that matches the word. Participants received feedback. In the Short condition, each word repeated until participants had accurately responded to it twice. In the Long condition, each new word repeated until participants had responded to it correctly four times, and each word learned in previous rounds was repeated until participants had responded correctly to it once. Thus, in both the Short and the Long condition, participants did not advance to the next task until they had learned all word–picture mappings. In total, participants in the Short condition needed to provide 24 correct responses for 12 words, and those in the Long condition 174 correct responses for 30 words.

Word matching: On each trial, participants saw a picture and listened to one fixed-random sequence of words they had learned. Each word appeared 2,000 ms after the onset of the previous word. Participants' task was to press a key as soon as they heard the word that fit the presented picture. In the Short condition, pictures were presented once. In the Long condition, pictures of new words were presented twice, and pictures of words learned in previous learning rounds were presented once. If participants answered incorrectly or failed to press the key when the target word was named, they were given corrective feedback. In total, participants in the Short condition performed 12 trials with 12 words, and those in the Long condition 114 trials with 30 words.

Thus, in total, in the Short condition, each of the 12 words was repeated a minimum of 5 times for a total of 60 tokens. In the Long condition, each of these 12 words appeared a minimum of 23 times with a total of 276 tokens heard across the four blocks.

The test stimuli included all learned words and an equal number of fillers. Each word was repeated once by each of the eight speakers. Participants' task was to indicate as quickly as possible whether the presented word was the one they had been taught. Participants in the Short condition completed 192 test trials, whereas those in the Long condition completed 480 test trials.

2.2. Results

To test if participants in the Long condition exhibited a larger advantage for native over non-native pronunciations, we conducted a mixed-model analysis over participants' accuracy in the test phase, considering only responses to the 12 words that all participants had been taught (i.e., words from the first learning round). Our model included Participant and Word as random variables, and Speaker (Native, Non-native), Condition (Short, Long), and their interaction as fixed factors. The random structure included intercepts for Participant and Word as well as a slope for Speaker for the Participant variable, and slopes for Speaker and Condition for the Word variable. Results revealed an effect of Speaker and the predicted Speaker × Condition interaction. As Fig. 2 shows, the interaction reflects the fact that even though participants in both conditions were more accurate when listening to native than to non-native speakers (Short: $\beta = 0.69$, $SE = 0.13$, $p < .001$; Long: $\beta = 0.91$, $SE = 0.13$, $p < .001$), the advantage for native speakers was larger for those in the Long condition ($\beta = 0.21$, $SE = 0.10$, $p < .04$; see Appendix B for the full table of results). This indicates that increased proficiency in a novel language leads to more accurate representations, which consequently leads to better performance with native than with non-native speakers.

In this experiment, we manipulated proficiency by varying the amount of exposure and the number of words that participants learned. One may wonder, however, whether proficiency is also related to a greater native speaker advantage when it is defined in terms of better overall performance during the test phase. We therefore examined whether participants' overall accuracy correlated with the size of the native speaker advantage they exhibited. Specifically, we computed overall d' scores for each participant as a measure of overall accuracy (collapsed over the native and non-native speakers). We also

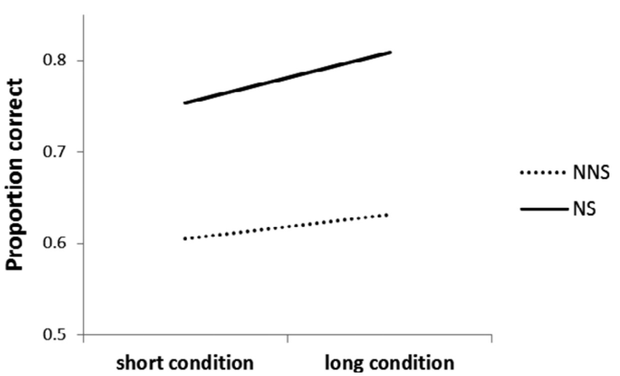

Fig. 2. Proportion of correct responses broken down by Speaker and Condition.

calculated, for each participant, the native speaker advantage as the difference between their d' prime scores with native and non-native speakers. As Fig. 3 shows, this correlation was significant when all participants were analyzed together ($r = .38$, $p < .05$). The correlation also held for participants in the Short condition only ($r = .37$, $p < .01$), whereas it was marginally significant for those in the Long condition ($r = .26$, $p < .08$).

The results of this experiment, then, show that higher proficiency, whether defined as greater exposure and larger vocabulary or as better performance during the test phase, is associated with a greater native speaker advantage.

3. General discussion

Processing accented speech is harder than processing native speech (Munro & Derwing, 1995; van Wijngaarden, 2001). Although such relative difficulty is in general rightfully considered an obstacle for communication between native and non-native speakers, we show that its emergence is, in fact, a marker of gaining proficiency in the second language. Specifically, exposing French participants to more Dutch words (both at the type and at the token level) led them to exhibit a greater difference between their accuracy on words produced by native and those produced by non-native Dutch speakers. Moreover, participants' overall accuracy positively correlated with the magnitude of the native speaker advantage.

We argue that the emergence of a native speaker advantage is due to the development of greater sensitivity to deviations from typical native productions. The lower one's proficiency is, the less likely one is to have precise representations of the phonological category, and therefore, the less deviant non-native productions are perceived. The more one develops proficiency and refines the phonological categories of the language, the harder it

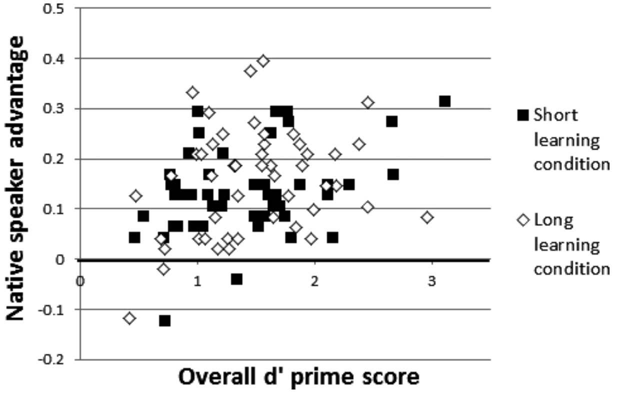

Fig. 3. The correlation between d' prime score and the native speaker advantage (average accuracy on items produced by Native speakers – average accuracy on items produced by Non-native speakers) for the two learning conditions. Each marker represents one participant.

becomes to understand non-native speech relative to native speech. Thus, even though more experience with and a larger vocabulary in a second language generally tends to improve word recognition, this improvement is larger when processing native than when processing accented speech. In our experiment, much like second language learning in real life, we varied both the amount of exposure and the number of words that were learned. This design, however, does not allow us to infer the relative contribution of each of these factors to the emergence of a native speaker advantage as a learner becomes proficient. Future research is needed to disassociate the two.

It is interesting that even the participants in the Short condition did not perform equally well with native and non-native speakers. This finding is different from Bent and Bradlow's (2003) results, where non-native listeners did not show any native speaker advantage. The overall advantage for native speakers in our study might be at least partially due to the absolute greater similarity between the phonetic realizations of different native speakers than between those of native versus non-native speakers. That is, our native speaker advantage might be similar to the advantage that listeners experience when they process words from a familiar speaker compared to a novel one (e.g., Goldinger, 1998; Nygaard, Sommers, & Pisoni, 1994). This effect may have been reduced in Bent and Bradlow (2003), as their L2-English participants lived in the United States, where they had likely had exposure to both native and non-native speakers, whereas our participants were only exposed to native Dutch speakers during training and did not have prior Dutch knowledge. In addition, the early emergence of a native speaker advantage might be due to our focused training. While the exposure to each of the words that participants in the Short condition received was relatively low, all words contained one of only three vowels and some of the words formed minimal pairs differing in their vowels, thereby facilitating the acquisition of these target vowels. Furthermore, we provided participants with varied exposure: They were exposed to multiple tokens from each of the four speakers, and each token they heard was unique. In general, it is known that variability boosts learning (Bradlow & Bent, 2008; Lively, Logan, & Pisoni, 1993; Rost & McMurray, 2009); our participants might have therefore been more proficient than one would usually expect after such a short training. While it is reasonable to assume that the listeners in Bent and Bradlow's (2003) study were, overall, more experienced language users, differences in the stimuli (single words vs. sentences) and tasks (speeded word recognition vs. transcription) might have led to a greater native speaker advantage already at earlier stages of second language acquisition in our study.

It is interesting to compare our results to those in research on first language acquisition. There, the common finding is that, initially, infants experience difficulty recognizing familiar words in unfamiliar accents (Best, Tyler, Gooding, Orlando, & Quann, 2009; Van Heugten & Johnson, 2014). That is, in the absence of any prior accent exposure, infants are unable to comprehend accented speech until they near their second birthday. It thus appears that for infants, the greater linguistic proficiency and increases in vocabulary size that develop over time assist them in contending with accented speech (cf. Mulak, Best, Tyler, Kitamura, & Irwin, 2013; Van Heugten, Krieger, & Johnson, 2015). The common pattern of results in such studies is similar to the one we illustrate in Fig. 1a.

This raises questions regarding the source of the seemingly different trajectories in infants and second language learners.

One option is that, at closer inspection, the two developmental patterns are not at odds with one another. That is, the finding that infants' recognition of words produced by accented speakers enhances over time and with increasing proficiency does not necessarily imply that the *relative* difference between contending with native versus non-native speakers enhances as well. That is, the transition from an early inability to cope with accented speakers to later above-chance performance may only reflect a small improvement, one that may be relatively modest in comparison to the enhancement in the recognition of words produced by native speakers. It is also possible that infants, just like second language learners, at first show a relatively small native speaker advantage, which increases as they gain more proficiency. However, they might simply not have been tested early enough for this to be observed. In other words, even infants as young as 15 months of age (currently the youngest age group tested on the recognition of familiar words produced by accented speakers) may be the equivalent of the high-proficiency second language learners in this study. Note that we do not argue that gaining further proficiency will lead to a greater native speaker advantage at all levels of proficiency. Once proficiency with native speakers plateaus, listeners might have more knowledge and cognitive resources to devote to understanding accented speech, hence reducing the difference in performance between native and non-native speakers, as we illustrate in the gray lines in Fig. 1b.

Alternatively, infants might indeed be different from second language learners. For example, second language learners use first language categories as reference, at least initially. This reliance on first language categories might influence what listeners perceive as the norm versus deviations from it, sometimes leading them to be more accepting of accented productions. Although infants might also have yet to develop fine-tuned categories, all their input is native accented and, consequently, they are not misled by any secondary language categories. Future research is required to tease apart these two possibilities.

In sum, this study shows that the relative difficulty of understanding non-native speech, although being a challenge for communication, is a marker of developing proficiency in a second language, as it indicates that one has calibrated the phonological categories of that language. The native speaker advantage might therefore be analogous to other linguistic, cognitive, and social processes that improve our performance with ingroup-relevant stimuli at the expense of processing ingroup-irrelevant stimuli.

Acknowledgments

This research was supported by Fondation Pierre-Gilles de Gennes, Fondation Fyssen, and by the Agence Nationale de la Recherche under grants ANR-13-APPR-0012**,** ANR-10-LABX-0087 IEC, and ANR-10-IDEX-0001-02 PSL.

Notes

1. Rationalized arcsine units are transformed proportions commonly used in speech perception research. This transformation solves some of the problems involved in using proportions in linear statistical models. Higher RAU scores indicate higher accuracy.
2. Most of our participants likely had some knowledge of English, as it is commonly taught in school. This might have affected their performance in our experiment (especially considering that one of the test speakers was a native Canadian English speaker). However, any such influence should be similar for the Short and Long condition, as there is no reason why knowledge of English would have differed between participants in the two conditions.
3. Although the French vowel inventory historically contained the vowel /ɑ/, this vowel has gradually merged with French /a/ since the 1950s (see Berns, 2015, and references cited therein).
4. Due to constraints that all words had to (a) have one of the three target Dutch vowels; (b) not be cognates of French or English words; and (c) be presented by pictures of easily recognizable objects, our final word set included—besides pairings of real Dutch words with their images—pairings of phonotactically legal nonwords with arbitrary images, and pairings of real Dutch words with images depicting unrelated objects.

References

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

Berns, J. (2015). Merging low vowels in metropolitan French. *Journal of French Language Studies*, *25*(3), 317–338.

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, *20*(5), 539–542.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*, 707–729.

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

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.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105* (2), 251–279.

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics*, *36*(4), 664–679.

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

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11), 864–886.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*, 3 Pt 1, 1242–1255.

Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., & Irwin, J. R. (2013). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent. *Child Development*, *84*, 2064–2078.

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

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349.

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

Takata, Y., & Náběelek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *The Journal of the Acoustical Society of America*, *88*(2), 663–666.

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

Van Heugten, M., & Johnson, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General*, *143*(1), 340–350.

Van Heugten, M., Krieger, D. R., & Johnson, E. K. (2015). The developmental trajectory of toddlers' comprehension of unfamiliar regional accents. *Language Learning and Development*, *11*(1), 41–65.

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*, 49–63.

van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication.*, *35*, 103–113.

Appendix A: List of stimuli

Learned words:

/beːk/, /dɑft/, /fleːs/, /flɛɪs/, /kɑf/, /kleːt/, /lɛɪm/, /seːp/, /stɛɪns/, /strɑnt/, /trɑp/, /xrɛɪp/.

Fillers used in test:

/bɔk/, /dʏft/, /flis/, /flus/, /kɛf/, /klɑp/, /klɑt/, /sɛɪst/, /soːp/, /xɑt/, /xlɛɪft/, /xrip/.

Additional words learned in the Long condition:

/bɑk/, /bɛɪt/, /deːn/, /eːnt/, /feːx/, /fɛɪx/, /flɑs/, /keːns/, /kɛɪf/, /klɛɪn/, /kreːft/, /kwɑst/, /lɑn/, /leːks/, /prɛɪ/, /swɑm/, /tɑɳ/, /wɛɪk/.

Appendix B: Table of results

	β	*SE*	*z*	*p*-value
(Intercept)	0.43	0.18	2.36	<.02
Speaker (Native)	0.69	0.13	5.21	<.001
Condition (Long)	0.11	0.17	0.66	.51
Speaker × Condition	0.21	0.10	2.09	<.04