In: S. Fuchs, M. Grice, A. Hermes, L. Lancia, D. Mücke (eds.) *Proceedings of the 10th International Seminar on Speech Production*, 261-264.

Do people converge to the linguistic patterns of non-reliable speakers? Perceptual learning from non-native speakers

Shiri Lev-Ari[1,2], Sharon Peperkamp[2]

[1]*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*
[2] *Laboratoire de Sciences Cognitives et Psycholinguistique (DEC, EHESS, CNRS), Paris, France*
shiri.lev-ari@mpi.nl, sharon.peperkamp@ens.fr

Abstract

People's language is shaped by the input from the environment. The environment, however, offers a range of linguistic inputs that differ in their reliability. We test whether listeners accordingly weigh input from sources that differ in reliability differently. Using a perceptual learning paradigm, we show that listeners adjust their representations according to linguistic input provided by native but not by non-native speakers. This is despite the fact that listeners are able to learn the characteristics of the speech of both speakers. These results provide evidence for a disassociation between adaptation to the characteristic of specific speakers and adjustment of linguistic representations in general based on these learned characteristics. This study also has implications for theories of language change. In particular, it cast doubts on the hypothesis that a large proportion of non-native speakers in a community can bring about linguistic changes.

Keywords: phonological representations, perceptual learning, language change, language contact, non-native speakers

1. Introduction

Speakers' phonological representations are based on their accumulated experience, and specifically, the input they have received. For example, because speakers of different languages are exposed to different distributions of sounds, they will draw the boundary between similar sounds differently, and therefore interpret the same sound as belonging to a different category in accordance with their experience. For example, a French speaker will categorize as /p/ what an English speaker would interpret to be /b/ (Abramson & Lisker, 1970).

Furthermore, speakers' representations are malleable; they are constantly shaped by further input (e.g., Flege, 1995; Goldinger, 1998). Any novel production listeners are exposed to influences the way they will interpret future speech. For instance, listening to a speaker who produces stops with deviant Voice Onset Times influences the way listeners will interpret stops by the same speaker as well as by other speakers in the future (Kraljic & Samuel, 2007).

Several accounts for the manner in which input influences future perception exist. Some theories of speech perception postulate that listeners retain all tokens they have encountered, together with contextual information, such as speaker identity. When new tokens are encountered, the stored tokens are activated according to their perceptual and contextual similarity, and thus guide the new tokens' interpretation (e.g., Goldinger, 1998).

Other theories posit that listeners strip all the context-specific characteristic of the speech and only maintain its abstract representation. The properties of the abstract representation though depend on the common features of the encountered input (e.g., McClelland & Elman, 1986).

Importantly, listeners are not passive vessels that simply store all input, giving all tokens equal weight in all contexts. For example, listeners interpret the same token differently depending on who they believe the speaker to be. Thus, a certain token is more likely to be interpreted as /ʊ/ than /ʌ/ if the listeners believe the speaker is a woman rather than a man, because differences in vocal tract size influence the location of the boundary between these vowels (Johnson, Strand & D'Imperio, 1999). Similarly, speakers' perceived age or regional background can influence listeners' perception and interpretation of the sounds they hear (Koops, Gentry & Pantos, 2008; Niedzielski, 1999). These findings suggest that the weight given to stored tokens depends on their relevance for the situation.

An even more sophisticated aspect of the speech perception mechanism is its ability to modify existing representations according to input that is perceived to be reliable, but not according to input that is perceived to be unreliable or unrepresentative. For instance, listeners do not change their phonological representations in accordance with input provided by a speaker holding a pen in her mouth even though they change their representations after exposure to the exact same tokens when the speaker does not seem to have any obstruents in her mouth (Kraljic, Brennan & Samuel, 2008).

In this paper we investigate whether listeners similarly adjust their representation in accordance with input provided by reliable, but not by unreliable speakers, and in particular, whether listeners' representations are therefore influenced by the speech of native, but not by the speech of non-native speakers. Answering this question will help us better understand not only the mechanisms of perceptual learning, but also the relation between adapting to the characteristics of specific speakers and adjusting general representations. Relatedly, the results of this study can indicate whether a large proportion of non-native speakers in a community can lead to language change, as has been proposed.

1.1. Non-native speakers and language change

Non-native speakers' speech is influenced by the properties of their first language, and is therefore unrepresentative of the common native productions (Flege, 1987). The deviations from common values are what makes non-native speech accented, and usually harder to understand. With experience, listeners adapt to these deviations, and their understanding of novel accented speech improves (Baese-Berk & Bradlow, 2013). From the point of communication efficiency though, listeners should rely on these learned properties only when listening to speakers with the same characteristics. That is, input from non-native speakers should not influence the way the speech of native speakers is processed and interpreted, as its deviations are atypical for native speakers' speech.

Nevertheless, several language contact accounts postulate a substratum language change, that is, language change that is brought about by having many non-native speakers in the language community. They propose either that those non-native speakers introduce into the language features from their first language or that they fail to acquire existing features and thus lead to their demise. It is argued that via repeated interactions between native and non-native speakers, and therefore repeated accommodation of native speakers to these non-native speakers, the novel patterns spread in the community (Niedzielsi & Giles, 1996; Trudgill, 1986; but see Hinskens & Auers, 2005).

In the case of interaction between speakers of different languages or dialects, such accommodation is argued to be a trigger of dialect leveling, as the latter often follows regular interaction between speakers of different dialects (Trudgill, 1986). Others have similarly suggested that languages with a large number of non-native speakers become simplified with time, due to the learners' simplification of the language (McWhorter, 2007). Yet for non-native speakers to leave a mark on a language, native speakers must learn and propagate the foreign features that are used by the non-native speakers. As the literature on perceptual learning indicates, such a process might not be likely.

In this study, we test whether non-native speech can influence the representations of native listeners. Importantly, the feature that we manipulate and whose influence we examine is not explicitly perceived as deviant. Therefore, we do not test whether listeners will adopt features that they explicitly perceive as erroneous, as we know that listeners are reluctant to do that even when the speaker is native (Ivanova, Pickering, McLean, Costa & Branigan, 2012).

2. Study

We use a perceptual learning paradigm (Norris, McQueen & Cutler, 2003). In this type of paradigm, listeners are exposed to an ambiguous segment, such as a stop with a Voice Onset Time that is intermediate between that of /b/ and /p/, in a context that disambiguates it (e.g., *les six /bp?/ananes* 'the six bananas' vs. *le long /bp/ont* 'the long bridge') They are later tested on their perception of the two phonemes that the ambiguous segment fell in between with a phoneme categorization task. Generally, listeners exhibit learning of the speaker's productions by shifting their boundary between the two phonemes in accordance with the interpretation of the speaker's earlier productions. When the ambiguous feature is not one that simply reflects individual physiological characteristics, the listeners also generalize their learning, as indicated by exhibiting a boundary shift even when categorizing the speech of novel speakers (Kraljic & Samuel, 2007). We hypothesize that this generalization is blocked when during exposure, listeners are presented with the speech of a non-native speaker. To recapitulate, participants in our experiment listened to either a native or a non-native speaker whose either /p/s or /b/s were manipulated. Later, they were tested with the speech of either the same speaker or a novel native speaker. We hypothesized that all listeners should be able to learn the characteristic of the speech, and thus show an effect of perceptual learning when tested with the speaker again. In contrast only those who listen to the native speaker should generalize the learned altered VOTs to new speakers, indicating a general influence on their representations, and therefore only they should show an effect of perceptual learning with a novel speaker.

2.1. Method

Participants. One-hundred-fifty-nine native French speakers were randomly assigned to one of eight conditions. Five participants were excluded because analysis of their performance indicated that they do not rely on VOTs when categorizing stops, at least in the tested VOT range.

Stimuli. Exposure stage. Twenty French words containing /b/ and 20 French words containing /p/ were selected. All were common words depicting concrete objects (e.g., *bananes* 'bananas'). One male native French speaker and one male native Dutch speaker were recorded reading these words in short noun phrases that contained a determiner, the noun and an adjective (including numerals), e.g., *les six bananes*. None of the adjectives contained a stop. The VOTs of the two speakers in the target words did not differ (18.6 vs. 21.5 ms, $p>0.05$). We then created another version of each target word by cutting out the pre-voicing in the voiced stops and the positive voicing in the voiceless stops. Thus, all altered words had stops with a VOT of 0. An earlier pilot study confirmed that all productions sounded natural and that the modification was not detectable by naïve participants. The speakers also read 119 filler noun phrases that did not contain any stops.

Each target and filler word was then matched with two pictures. Both pictures depicted the object in the NP, but only one of each pair of pictures matched the adjective as well.

Phoneme categorization. The same native French and native Dutch speakers, as well as an additional female French speaker were recorded saying the words *bulle* /byl/ and *pull* /pyl/. For the *pull* end of the continuum, we used a token of 29ms – the VOT value in the speech of the male native speaker. The VOT of the Dutch speaker was particularly short for this word (16ms), and would not have been sufficiently long to allow creation of a VOT continuum. We therefore artificially lengthened it by copying the part of the existing VOT and pasting it at the end to match the VOT duration of the other training speaker. The *bulle* end of each continuum was created by cutting out the positive VOT from the *pull* token and pasting pre-voicing from that speaker's production of *bulle*. We then created 10 additional steps in between the continuum's ends by cutting out medial sections of positive or negative voicing. The values of all tokens in the continua of the two training speakers (in ms) were: -90, -50, -15, -10, -5, 0, 5, 10, 15, 20, 25 and 29.

Pre-tests indicated that the perceptual transition point from voiced to voiceless stop was earlier for the two training speakers than for the generalization speaker, potentially due to the gender difference (Swartz, 1992). Therefore, the range of values in the continuum of the generalization speaker extended to higher VOT values and was: -90, -60, -10, -5, 0, 5, 10, 15, 20, 25, 30 and 59. Other than that, the continuum was prepared in the same manner as the continua of the training speakers. Note that in all continua, the two extreme tokens on each end of the continuum served as filler reference points. They were repeated fewer times and responses to them were not analyzed (see Procedure and Results).

Procedure. Participants came for two sessions, at least two days apart. In the first session, participants performed both the exposure task and the phoneme categorization task. In the second session, participants performed only the phoneme categorization taskproviding a baseline measure of participants' categorization. We obtained the baseline measure at the end rather than the beginning to avoid drawing participants' attention to the manipulated phonemes, and thus potentially to our manipulation of the audio files.

In each trial of the exposure stage, participants saw two different pictures of an object (e.g., a modern port and an old port), and listened to an auditory phrase that matched only one of the pictures (e.g., *le vieux port* 'the old port'). Their task was to select the picture that matched the phrase. In this stage, half of the participants listened to the native speaker and half listened to the non-native speaker. Half of the participants in each Speaker condition heard the natural tokens of /p/ and the altered tokens of /b/. The other half heard the natural tokens of /b/ and the altered tokens of /p/. There were 159 items in total: 20 with /b/, 20 with /p/, and 119 fillers. The first two items were filler items. The rest appeared in random order.

Following the exposure stage, participants performed the Phoneme Categorization task. Half of the participants in each speaker condition listened to the same speaker as in the exposure stage. The other half listened to the novel female French speaker. Participants were explicitly told whether the speaker they were listening to was the same one as before or a different one. The phoneme categorization task contained two blocks. In each block, the two shortest and two longest tokens were repeated four times each, and the other eight tokens repeated ten times each, totaling 96 trials per block. One trial with the extreme token on the *pull* end of the scale preceded the first block to provide a reference point and practice. In the second session, participants performed the exact same Phoneme Categorization task, but without the exposure phase.

2.2. Results

To examine whether non-native productions can influence listeners' linguistic representations, we needed to first identify the cases in which native productions influenced listeners' representations. We therefore first analyzed the responses of participants in the generalization condition who were exposed to manipulated native speech. A mixed model analysis with Participant as a random variable and Baseline Performance, Manipulated Phoneme (voiced, voiceless), Item and the interaction of the latter two as fixed effects revealed a main effect of Baseline (β=3.38, $p<0.0001$), indicating that participants' responses in the baseline session predicted their responses in the experimental session. The analysis also revealed effects of Manipulated Phoneme (β=1.36, $p<0.03$) and Item (β=-0.41, $p<0.0001$), but these were modulated by a Manipulated Phoneme x Item interaction (β=-0.29, $p<0.001$), indicating that the effect of perceptual learning was not present or of equal magnitude for the different items. We therefore carried out separate analyses on each item to determine in which items the effect of perceptual learning was manifested. Analyses for items 3, 4, 5, 6, 9 and 10 did not reveal any effect of Manipulated Phoneme (all p's>0.1), but only effects of Baseline Performance (all p's<0.04, except for item 9, p>0.1). For Item 7 there was a marginal effect of Manipulated Phoneme (β=-0.71, $p<0.06$), as well as an effect of Baseline Performance (β=2.23, $p<0.01$), and for Item 8, there was only an effect of Manipulated Phoneme (β=-1.02, $p<0.05$). These results indicate that the perceptual manipulation has most clearly influenced the perception of Item 8, where it was even strong enough to completely eliminate the effect of participants' baseline perception of this item.

We next examined whether the performance of participants who were exposed to the non-native speech also showed an influence of the manipulated phonemes on their representations. Analyses of performance on each item separately, as well as on all items together, showed that participants were not influenced by the manipulated phoneme for any of the items (all p's>0.1). Furthermore, we ran a joint mixed model analysis of participants in both the Native and Non-native speaker conditions over responses for Item 8 to see if participants' responses in the two Speaker condition are indeed different. The analysis revealed an effect of Baseline Performance (β=3.2, $p<0.01$) and the predicted Speaker x Manipulated Phoneme interaction (β=-1.14, $p<0.05$). To conclude, the results show that only the speech input that was provided by the native speaker influenced listeners' phonological representations.

One may wonder though whether non-native speech failed to influence listeners' representations because it is harder to learn rather than because listeners block the influence of less representative speech. To examine whether that is the case, we tested whether participants who were tested with the same speaker succeeded in learning the speech of the non-native speaker. Separate analyses for each item indicated that listeners' performance showed an effect of Manipulated Phoneme for item 5 (β=-0.49, $p<0.04$), and marginal effects of Manipulated Phoneme for items 7 (β=-0.41, $p<0.06$) and 8 (β=-0.58, $p<0.06$). Importantly, jointly analyzing the responses of participants who listened to the native and non-native speakers in the Same Speaker condition did not reveal an interaction of Speaker and Manipulated Phoneme for any of the items separately nor for all items together (all p's>0.1). This indicates that the magnitude of the perceptual learning effect was similar in the Native and Non-native speaker conditions. Lastly, to ensure that the performance in the Same Speaker and Generalization condition are indeed different, we ran an analysis of responses to Item 8 by participants in all conditions. This analysis revealed an effect of Baseline Performance (β=2.15, $p<0.0001$), an interaction of Speaker with Manipulated Phoneme (β=-1.02, $p<0.04$), a marginal interaction of Generalization and Manipulated Phoneme (β=-0.77, $p<0.09$), and the predicted 3-way interaction of Speaker, Manipulated Phoneme and Generalization (β=1.61, $p<0.02$). Together, the results show that even though listeners are able to learn the speech of both native and non-native speakers, they do not adjust their representation in accordance with it, and therefore only generalize native input.

3. Discussion

Individuals learn language from their environment: They learn the characteristics of the language of the speakers they hear and adjust their representations accordingly. Yet not all input that listeners encounter is equally reliable. For example, individuals should rely less on speech input provided by someone who has just been to the dentist or who is holding a pen in her mouth. Similarly, individuals should rely less on the speech of speakers who deviate much from the majority of their community, such as non-native speakers, people with speech impediments, or children. Our results indicate that listeners indeed take the reliability of the speaker into account when learning their speech.

Using a perceptual learning paradigm, we manipulated the VOT of bilabial stops, a feature that differs across languages, and that non-native speakers therefore sometimes produce with a foreign accent. The manipulation was subtle, and due to the semantic context, imperceptible to listeners. Previous research has shown that exposure to such altered speech influences listeners' later interpretation of speech by that speaker, as well as by other speakers (Kraljic & Samuel, 2007; Norris et al., 2003). In other words, such exposure can alter listeners' representations of voiced and voiceless stops in general. Indeed, the results of our study replicated this effect. Crucially, the results of our study show that such influence on representation only occurs when the input is provided by native speakers. On the one hand, all listeners were able to learn the characteristics of the speaker's speech, in line with previous research showing quick accommodation to foreign accent (Bradlow & Bent, 2003). Yet despite this ability to learn the speaker's speech characteristics, only those who listened to the native speaker adjusted their representation in general according to the newly learned characteristic, and used it when interpreting the speech of a new speaker.

One may wonder why listeners showed generalized perceptual learning on a single item only, that of 15 ms of VOT. One potential reason might be that even though listeners generalize the characteristics they learn to other speakers, the influence of these characteristics might be smaller than when listening to the same speaker, especially when there is a need to map the characteristics to speech with different properties due to, for example, a change in gender. It is also worth noting that even though the effect of generalized perceptual learning was only significant for one item, the participants' performance in three additional items (i.e., items 5, 6 & 7, with VOT values of 0, 5 & 10ms) was in the predicted direction, but failed to reach significance.

Our study, then, shows that listeners block input of non-native speakers form influencing their representations. Future research should examine whether there are native speakers that are also considered to be a less fit model to learn from, and whose input is therefore not weighted as heavily as others'. For example, children, or even adults of a different generation might be considered to be less good models. Another question that remains open is how listeners decide how to classify speakers. Do listeners rely on top-down information, such as knowledge that they listen to a non-native speaker, or are they influenced by bottom-up cues, such as the general deviations of the speech input from other speakers? In other words, it could be that participants in our study implicitly blocked any influence of the speech of the non-native speaker because of their knowledge that he was a non-native speaker. Alternatively, it could be that the participants encoded the deviation of the non-native speaker from average speakers in the community on multiple features other than VOT, and thus implicitly tagged him as an outlier and blocked any influence of his speech. The latter account suggests then that even native speakers might not all be learned from to the same degree. In particular, listeners may learn less from native speakers the further their speech is from the average speaker or from the speakers that are perceived to be good models.

The results of this study also have implications for language change. Many communities in today's world include a significant proportion of non-native speakers. According to some theories, such composition should lead to language change. Yet for substratum language change to be a likely result, non-native speech must be able to influence native representations. The results of our study suggest that this is not the case.

One objection might be that substratum language change could also come about not by accommodation of native speakers to non-native speakers but by transfer of the features from non-native parents to their children. While our study does not speak to that, some research suggests that this might not be the case either. Children of non-native speaking parents are usually exposed to speech by other community members as well, and research on dialect variation has shown that the ultimate speech patterns of children of other-dialect speakers often end up resembling those of the community rather than the parents (e.g. Floccia, Delle Luche, Durrant, Butler & Goslin, 2012). More research is needed, though, to examine whether children of non-native parents have representations that differ from those of children of native speakers.

At the same time, there is ample evidence that language contact can lead to language change. Note that we do not argue that language contact cannot trigger language change. Rather, we argue that such contact-driven language change is not brought about by accommodation to non-native speakers. One alternative route for language change might be the influence of a second language on the first one when native speakers learn the languages of other language communities.

The main conclusion of this study, though, is that while listeners' representations are influenced by the speech they hear, not all speech input is equally able to modify listeners' representation, and in particular, non-native speech does not influence native representations.

4. References

Abramson, A.S. & Lisker, L. (1970). Discriminability along the voicing continuum: cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences*, pp. 569–573.

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

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

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of phonetics*, *15*, 1, 47-65.

Flege, J.E. (1995) Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, pp. 233-277.

Floccia, C., Delle Luche, C., Durrant, S., Butler, J. & Goslin, J. (2012) Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition 124,1*, 95-100.

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

Hinskens, F.L.M.P. & P. Auer. (2005) The role of interpersonal accommodation in a theory of language change. In: Auer, P. & F. Hinskens & P. Kerswill (ed.). *Dialect change. The convergence and divergence of dialects in contemporary societies.* Cambridge UK : Cambridge University Press, 2005, pp. 335-357.

Ivanova, I., Pickering, M.J., McLean, J.F., Costa, A. & Branigan, H.P. (2012) How do people produce ungrammatical sentences? *Journal of Memory and Language, 67,* 355-370.

Johnson, K. (1997) Speech perception without speaker normalization: An exemplar model. In Johnson & Mullennix (eds.) *Talker variability in speech processing*. San Diego: Academic Press, pp. 145-165.

Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas, *University of Pennsylvania Working Papers in Linguistics: Selected papers from NWAV 36, 14,* 91–101.

Kraljic, T. & Samuel, A.G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56, 1-15*.

Kraljic, T., Brennan, S.E., & Samuel, A.G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition, 107, 1,* 51-81.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*, 1, 1-86.

McWhorter, John H. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford: Oxford University Press, 2007.

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

Niedzielski, N. & Giles, H. (1996) Linguisticaccommodation.In H. Goebl, P. Nelde, Z. Starý,& W. Wolck, (eds.), *Contact linguistics: An international handbookof contem- porary research*, pp.332–342. Berlin: Walter de Gruyter

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47, 2,* 204-238.

Swartz, B. L. (1992) Gender difference in voice onset time. *Perceptual and Motor Skills, 75,* 983-992.

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