THE EFFECT OF INFERENCES ON THE PERCEPTUAL CATEGORIZATION OF BERLIN GERMAN FRICATIVES

Stefanie Jannedy, Melanie Weirich & Jana Brunner

Center for General Linguistics (ZAS), Berlin ,Germany
jannedy@zas.gwz-berlin.de; weirich}@zas.gwz-berlin.de; jana.brunner@uni-potsdam.de

ABSTRACT

In Berlin German, the identification of /ɕ/ as in *Fichte* 'spruce' versus the alveopalatal fricative /ʃ/ in *fischte* (3. sg. past tense) 'to fish' is influenced by what neighborhood in Berlin hearers believe a speaker comes from (KB: Kreuzberg: multi-ethnic, multi-cultural or ZD: Zehlendorf: mono-ethnic, upscale, affluent area) and the inferences made as to how the speaker is likely to realize these two variants. Our results indicate that listeners adjust their interpretation of synthesized acoustic continua in accordance with their expectation, strongly suggesting that both perceptual cues and inferred social factors play a role in the categorization of speech stimuli. Thus, the results show that the mere suggestion of where the speaker may come from even within the confines of a city is enough to trigger such inferences.

Keywords: *Kiezdeutsch*, inferences, cross-sociolectal categorization, palatal fricatives, attribution

1. INTRODUCTION

In many European cities, researchers have noticed and studied the emergence of linguistic variation and the grammatical innovations introduced by young speakers from multi-ethnic urban neighborhoods. *Kiezdeutsch* is a youth-style multi-ethnolect spoken predominantly in larger urban areas of Germany. It has several specific grammatical characteristics and phonetically, it is saliently characterized by the palatalization of /ɕ/ to /ʃ/ following a high front vowel within the same syllable. Some speakers alternate between these two variants while others seem to only use [ʃ]. This alternation is quite pervasive and noticeable, it is mocked and stigmatized and there is a general awareness in the Berlin population that many young speakers from multi-ethnic neighborhoods such as Kreuzberg show such an alternation. Our previous corpus and perception study [4] has shown though that Berliners from different age groups and social backgrounds, too, show this alternation, leading to the more general assumption that the /ɕ/ is a fairly unstable speech sound in this dialect area. This however seems largely unnoticed by the general public.

The aim of this study is to evaluate the effect of implicit social information (neighborhood with all its inferred social characteristics as the supposed origin of the stimuli) on the categorization of synthesized acoustic continua ranging from /ɕ/ to /ʃ/. We are looking for these inferred attributions with the confines of a single urban area.

It has been shown that inferred information on the origin of speech stimuli biases the perception and rating of native listeners [3, 7] along the lines of national or geographical identity. For example, in [7], Detroit listeners' responses to a perceptual vowel matching task showed more raised variants when listeners believed to hear Canadian speakers who produce *Canadian Raising* in their vowels as opposed to Detroit speakers who also have this raising but of which they are not aware. In [3], the authors replicated [7] in the context of Australian versus New Zealand vowel variants of /ɪ/ which is more raised in Australian English. In both of these studies, the only cue as to the origin of the speaker was the national identity (Detroit/Canadian vs. Australien/New Zealander) written on the response sheets. They also found such a perceptual divergence effect whereby listeners responded differentially depending on who they believed the speakers were. In a second follow-up study [3], they were able to show that priming listeners with symbols denoting national identity also affects categorization of speech.

These observations support the view that speech perception is strongly influenced by information bootstrapped from non-linguistic observations and from inferences about speakers and speaker groups by mere suggestion of their geographic origin. It appears as listeners adapt to speakers perceptually as part of a subconscious socio-cognitive learning process. Here, we are investigating the indirect inferences and attributions that Berlin and Brandenburg listeners

make when they are led to believe that they are listening to speech/stimuli from speakers from specific districts of Berlin, one known for its multi-cultural and multi ethnic composition, while the other is known as predominantly mono-ethnic, affluent and rich.

2. METHODS

Listeners from Berlin and Brandenburg were asked to identify the stimuli from two continua /fɪɕə/-/fɪʃtə/ and /fɪɕə/-/fʏʃtə/ as either *Fichte* or *fischte* by pressing the appropriate button on a response box. Continuum 2 reflects the rounding of the lax high front vowel before an alveopalatal fricative /ʃ/ typical for Berlin German.

To facilitate listeners and eliviate the cognitive load on memorizing which buttons to press, the response button for *Fichte* was always coded green and the one for *fischte* was coded blue.

A laminated card with color pictures of a spruce and of a young man holding a fishing rod against the background of a blue sky was attached to the button box. In hand-writing the card also showed the CONDITION: *Kreuzberg* or *Zehlendorf*. Half of the listeners were presented with *Fichte* on the left side of the response box and the other half with *Fichte* on the right side. For all listeners, identical procedures were applied. Because our experimental set up depended crucially on listeners noticing the experimental CONDITION that they were in, we also wrote the words *Kreuzberg* or *Zehlendorf* on the subject information sheet that they had to fill out prior to starting the experiment. While the listener was filling out the form, the experimenter asked to have a look at the form to cross check the group and then read out the CONDITION so that it was audible to the listener. The listener's attention was implicitly and subtly directed to his group membership by casually mentioning the CONDITION under the assumption that s/he would derive inferences from that.

2.1. Mode of presentation

In a sound attenuated room (over headphones), listeners were presented with 11 Blocks (each in a different random order) of 27 stimuli. Block 1 served as training and was excluded from the analysis. Each continuum consisted of 13 steps from *Fichte* /fɪɕə/ to fischte /fʏʃtə/. The starting step of the *Fichte* side of the continuum was presented only once, however, responses to Step1 were included in the statistical analysis of both continua. The inter-stimulus interval was 2000ms but the next stimulus was played sooner when the response button was pressed quickly. Hearers were asked to rate each stimulus as either *Fichte* or *fischte*.

2.2. Stimulus construction and selection

A 14 year old male native speaker of Berlin German was recorded saying the words *Fichte* /fɪɕə/, *fischte* /fɪʃtə/ and an unrounded variant /fɪɛtə/ in between these two extremes. Two continua were created. In the first continuum, only the fricative portion was varied while the stressed vowel was kept as in *Fichte* (see Fig1 and upper panel of Fig4). In the second continuum, the fricative was shifted from /ɕ/ to /ʃ/ also, but in addition, the vowel was shifted from the /ɪ/ in *Fichte* toward a more rounded lax front vowel like /Y/ (see Fig2).

Figure 1: Resynthesized 27-step continuum from /ɕ/ to /ʃ/.

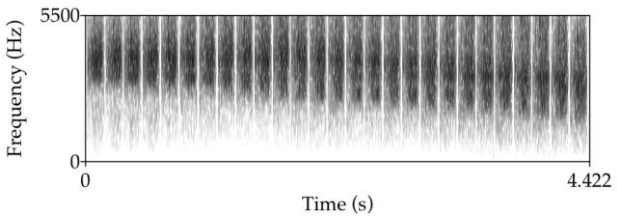
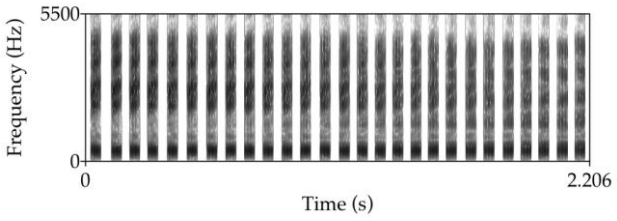


Figure 2: Resynthesized 27-step continuum from /I/ to /Y/.



Each of these two continua had three anchor points, namely the endpoints of the continuum *Fichte* /fɪɕə/ and *fischte* /fɪʃtə/ or /fʏʃtə/and a intermediate unrounded fricative version such as /fɪɛtə/ or /fʏɛtə/. The fricative continuum was created using the Klatt synthesizer [5]. The synthesis parameters (frequency and amplitude of each prominent spectral peak) were derived from the natural renditions of /ɕ ɛ, ʃ/.

The continua were created by morphing between the sibilant parameters [1]. The continuum from /ɕ/ to /ɛ/ had 10 steps, the one

from /ɕ/ to /ʃ/ had 17 steps. The vocalic continuum was created by interpolating between the anchor points [6]. The synthesized vowels and fricatives were then spliced into the original *fichte*-utterance. From these 27 steps, we selected 14 test items through pre-tests by evaluating what step-range listeners seem to be specifically sensitive to. In these tests, the stimuli were rated to sound naturally by native listeners.

2.3. Subjects

44 (14 male and 30 female native Berlin/ Brandenburg listeners recruited from the Humboldt University student pool) participated in the study. They ranged from 19-31 with an average age of 24.54 (SD 3.16). Males were on average 23.63 (SD 2.29) and females 24.84 (SD 3.38) years old. None of them reported speech or hearing problems.

2.4. Statistical analysis

In total, there were 12.320 responses (44 subjects x 28 Steps x 10 Blocks) from which we had to exclude 122 items due to logging errors (~1%), thus 12.198 data points were subjected to the analyses. Identification results were analyzed using separate linear mixed effects models (lmer) in R (2.10.1) for the two different continua. The dependent variable is the type of response given (*Fichte* or *fischte*), the independent variables and fixed factors are the CONDITION, STEP and GENDER, the random factor is BLOCK.

3. RESULTS

Fig3 shows the response data for each STEP in the /ɪ ɻ-/ɪʃ/ (top panel) versus /ɪ ɻ-/Yʃ/ (bottom panel) continuum for both experimental conditions. For the ZD group (solid lines), hearers generally rate stimuli longer to be *Fichte* (grey) compared to the KB group (dashed lines) where a slightly faster rise before the cross over point indicates more ratings for *fischte* (black). It is particularly apparent in the lower panel that the cross-over point for the KB group precedes that for the ZD group.

Fig4 shows the identification results for each step of the two continua depending on what CONDITION the listener was exposed to. Listening to identical series of stimuli, hearers judged/rated the steps of the continuum differentially depending on which group they belonged to.

The bottom panel clearly shows that the coloring of the vowel preceding the /ʃ/ helps listeners identify the fricative category. In the top panel, ratings for *Fichte* seize more rapidly when listeners were exposed to the concept *Kreuzberg* as opposed to *Zehlendorf*.

For both continua, the model reflects significant effects of CONDITION (z= -5.797, p<.001 for /ɪ ɻ-/Yʃ/ and z = -4.45, p<.001 for /ɪ ɻ-/ɪʃ/). GENDER however is only significant for the /ɪ ɻ-/ɪʃ/ continuum (z = 2.91, p<.01) where there is only a single auditory cue.

Figure 3: Responses in % for each STEP for the two continua /ɪ ɻ-/ɪʃ/ (top panel) and /ɪ ɻ-/Yʃ/ (bottom panel). The solid lines represent the data for the CONDITION Zehlendorf while the dashed lines represent the data for CONDITION Kreuzberg.

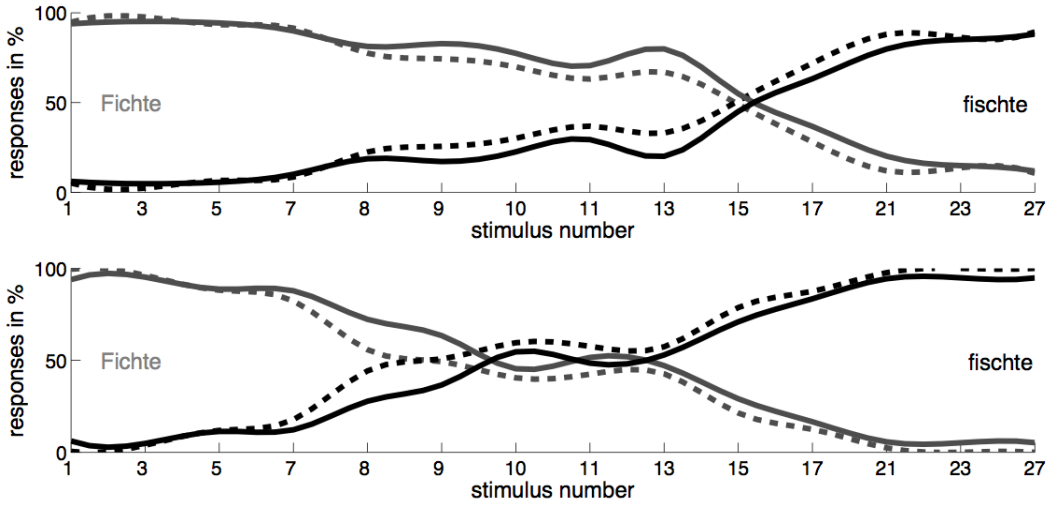
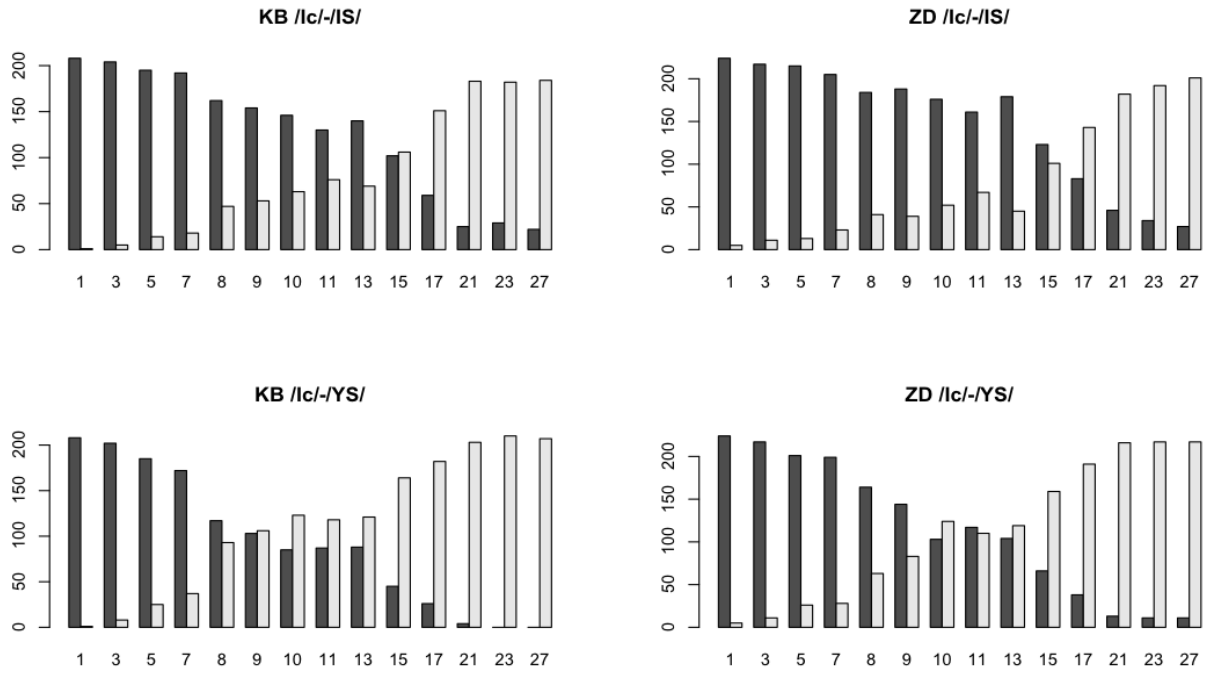


Figure 4: Identification results (counts: y-axis) for each step (x-axis) of both continua depending on the CONDITION (KB left, ZD right) with dark bars representing ratings for Fichte' and light bars representing fischte responses.



For STEP, in the /ɪɕ-/ɪʃ/ continuum (top panel), the model finds a significant difference between Step1 and Step7 ($z = 4.31$, $p<.01$), whereas for the /ɪɕ-/Yʃ/ continuum (lower panel), there is a significant difference already between Step1 and 5 ($z = 4.68$, $p<.01$), highlighting the importance of the additional cue of rounding. The probability of identifying a stimuli as *fischte* in /ɪɕ-/ɪʃ/ is more than 50% as of Step 15 for KB and Step 17 for ZD.

In contrast, for the /ɪɕ-/Yʃ/ continuum perception shifts as of Step 9 for KB and Step 10 for ZD. These differences point towards the fact that the coloring in the vowel cues listeners in the identification of the palato-alveolar item (*fischte*) of the minimal pair.

4. SUMMARY & CONCLUSION

First, there is a difference in response to the two continua, mirroring the effect of the additional acoustic cue for the rounded vowel. Second, the significant effect of CONDITION suggests that hearers adjust their interpretation of an acoustic stimulus in accordance with a triggered expectation: it matters if hearers heard and read the words *Kreuzberg* as opposed to *Zehlendorf*. Both words trigger different associations that lead to differing responses and differing perceptual categorizations of the phoneme contrast between /ɕ/ and /ʃ/.

5. ACKNOWLEDGEMENTS

The work was supported by the German Federal Ministry for Education and Research (BMBF) – Grant Nr. 01UG0711.

6. REFERENCES

[1] Ghosh, S., Matthies, M., Maas, E., Hanson, A., Tiede, M., Ménard, L., Guenther, F., Lane, H., Perkell, J.S. 2010. An investigation of the relation between sibilant production and somatosensory and auditory acuity. *J Acoust Soc Am.* 125(5), 3079-3087.

[2] Hay, J., Drager, K. 2010. Stuffed toys and speech perception. *Linguistics* 48(4), 865-892.

[3] Hay, J., Nolan, A., Drager, K. 2006. From fush to feesh: Exemplar priming in speech perception. *Linguistic Review* 23, 351-379.

[4] Jannedy, S., Weirich, M., Brunner, J. 2010. Perceptual evidence for allophonic variation of the palatal fricative /ɕ/ in spontaneous Berlin German. *J Acoust Soc Am.* 128, 2458.

[5] Klatt, D.H., Klatt, L.C. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am.* 87, 820-857.

[6] Mitterer, H., Chen, Y., Zhou, X. 2011. Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science* 35, 184-197.

[7] Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(1), 1-18.