ISCA Archive
http://www.isca-speech.org/archive

INTERSPEECH 2012
ISCA's 13th Annual Conference
Portland, OR, USA
September 9-13, 2012

Phonology & the Interpretation of Fine Phonetic Detail in Berlin German

Stefanie Jannedy[1], *Melanie Weirich*[2]

[1]Laboratory Phonology Group, ZAS, Berlin (Germany)
[2]Institut für Germanistische Sprachwissenschaft, Friedrich-Schiller Universität Jena (Germany)
Jannedy@zas.gwz-berlin.de, melanie.weirich@uni-jena.de

10.21437/Interspeech.2012-38

Abstract

Young multi-ethnolectal speakers of Hamburg-German introduced an alternation of /ç/ to [ʃ] following a lax front vowel /ɪ/ [1]. We conducted perception studies exploiting this contrast in Berlin (Germany), a city with large multi-ethnic neighborhoods. This alternation is pervasive and noticeable, it is mocked and stigmatized and there is an awareness that many young speakers (including ethnically Germans) from neighborhoods with larger migrant populations like Kreuzberg (KB) substitute /ç/ with /ʃ/ while speakers from less stigmatized vicinities like Zehlendorf (ZD) do not.

The categorization of items on two 14-step synthesized continua from *Fichte* 'spruce' to *fischte* '3rd p. sg. to fish' by 99 listeners shows that the interpretation of fine phonetic detail is strongly influenced by the co-presentation of the label *KB* or ZD in contrast to no label (control). Analyses of the reaction times (RTs) show that significantly more time is needed to process stimuli in KB and less in ZD. Moreover, younger listeners (below 30 years) perceive more /ʃ/ variants than older listeners. Phonological generalization over phonetic input is dependent on associative information: perceptual divergence is found within the confines of a single large urban area [2,3,4].

Index Terms: sociophonetics; urban German; palatal fricative

1. Introduction

A perceptual divergence effect describes the differential categorization of identical stimuli in dependence to a prime. Such a perceptual divergence effect along the lines of national identity was first observed by [4] for an US-Canadian context and then replicated by [3] and [2] for an Australian-New Zealand context. The authors exploited the effect that Australian English has generally fewer raised variants of /ɪ/ while New Zealand English has more raised variants. In this New Zealand (NZ) – Australian context, they used a word-matching paradigm in which synthesized tokens from a continuum of more or less raised variants of the lax front vowel /ɪ/ had to be matched to natural recordings of a single speaker saying words containing this vowel. NZ-listeners were tested in two conditions and were rather subtly primed with objects denoting national identity: for the Australian condition, listeners saw a stuffed toy koala bear and for the NZ condition, listeners saw a stuffed kiwi bird. Listeners in the kiwi context matched more raised synthesized variants to the stimulus words than listeners in the koala-bear context. From the results of this word-matching paradigm, the authors conclude that the perception shifts as a function of which toy category was seen by the listeners and that listeners display a perceptual divergence effect along the lines of a national or geographical identity.

Questions arising from the results addressed in this study pertain to whether the effect is due to the potentially antagonistic relationship between people on two sides of a geo-political border and thus, can this effect only be replicated along the lines of national identity? Or is this effect a more general expression of identity whereby also local identities for example within the confines of a larger urban space can induce such diverging behavior?

The synchronic alternation of /ç/ and /ʃ/ is observed in a multi-ethnolect spoken by young multi-ethnic speakers in urban areas of Germany (documented in Hamburg and Berlin) where this alternation is (from a dialectal perspective) unexpected. We have followed up on this observation and conducted production and perception studies in Berlin where we also find larger multi-ethnic neighborhoods in which this alternation is quite pervasive and noticeable [5,6]. The question that we have been trying to address is if a perceptual divergence effect as found in the New Zealand/Australian and Canadian/US context can be replicated within the confines of a single city that consists of different neighborhoods whereby some of these areas are known for their multi-ethnic and culturally diverse composition and which are often characterized by low family income, higher unemployment rates or an unwillingness to integrate into German society.

Instead of trying to replicate the effect with the same experimental paradigm, we used a forced-choice identification task with an artificially created acoustic continuum. The meaningful linguistic variable is the alternation of /ç/ to /ʃ/. If there is some level of awareness that makes hearers judge identical stimuli differentially in dependence on who they believe the speaker is, it strongly implies that phonology interprets fine phonetic detail and makes use of variability in speech rather than simply abstracts over it and regards it as noise. That is, variability can be pragmatically and perceptually extremely useful. More importantly, this is evidence that there is a much less categorical division between phonology and phonetics [7].

2. Methods

In this study, we build on and further explore our pilot work [5,6] and examine whether perceptual divergence effects are also found within an ethnically diverse and culturally rich urban context in Berlin. In addition to a forced-choice identification task with listeners in 4 age groups, we gathered reaction times (RTs) for a subgroup of listeners. We expect that immediate response measures such as RTs give us cues to whether different kinds of added information contribute to the processing load by increasing the time it takes for the listener to respond.

2.1. Description of Experiment

Listeners from Berlin and Brandenburg were asked to classify the stimuli from two continua as either *Fichte* versus *fischte*: 1. /fɪçtə/-/fɪʃtə/ and 2. /fɪçtə/-/fʏʃtə/. Both words form a minimal pair, whereby [fʏʃtə] is a dialectal alternation of /fɪʃtə/ in Berlin German. Thus, we were looking for the category boundary between /ç/ and /ʃ/. Listeners were instructed to press the suitable button on a response box. Continuum 2 (Fig. 2) reflects the rounding of the lax high front vowel before an alveopalatal

fricative /ʃ/ typical for Berlin German. To facilitate listeners and lighten 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 showed the CONDITION: *Kreuzberg* (KB) or *Zehlendorf* (ZD) or it did not list any condition (control, CO). 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 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 the group membership by casually mentioning the CONDITION under the assumption that s/he would derive inferences from that.

2.1.1. Subjects

In total, 99 (26 male and 73 female) native Berlin/ Brandenburg listeners participated in the study. They ranged from 19-61 with an average age of 27.6 (SD 8.3). None of them reported speech or hearing problems. Table 1 summarizes the information on the different age groups that were tested. As of now, there are only 6 listeners older than 45. More subjects in this age category are currently tested to have similar numbers of listeners in all age groups.

Table 1: *Number of Subjects and Responses by Age Group.*

Age Group	No. of Subjects	No. of Responses
Young (mean=22.9, SD=1.9)	61	8529
Middle Young (mean=29.1, SD=1.7)	23	3205
Middle Old (mean=35, SD=0.9)	6	840
Old (mean=50.1, SD=5.1)	9	1242

2.1.2. Stimuli

To create the stimuli, a 14 year old male native speaker of Berlin German was recorded saying the words *Fichte* [fɪçtə] and two variants of *fischte* [fɪʃtə] and [fʏʃtə] and an intermediate form of the fricative in *fischte* [fɪçtə] in between these two extremes. Two continua were generated. In the first continuum, only the fricative portion was varied while the stressed vowel was kept as in *Fichte* (see Fig1). In the second continuum, the fricative was shifted from [ç] to [ʃ] also, but in addition, the vowel was shifted from the [ɪ] in *Fichte* toward a more rounded lax front vowel like [ʏ] (see Fig2).

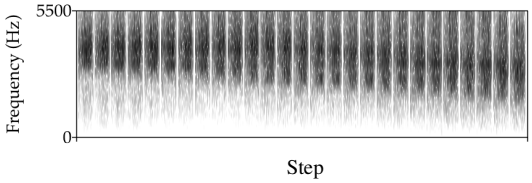

Figure 1: *Resynthesized 27-step continuum from [ç] to [ʃ].*

Both of the two continua had three anchor points, namely the endpoints of the continuum *Fichte* [fɪçtə] and *fischte* [fɪʃtə] or [fʏʃtə] and an intermediate unrounded fricative version such as [fɪɕtə] or [fʏɕtə]. The fricative continuum was created using the Klatt synthesizer [8]. The synthesis parameters (frequency and amplitude of each prominent spectral peak) were derived from the natural renditions of [ç, ɕ, ʃ] whereby [ɕ] is not a phoneme of German. The continua were created by morphing between the sibilant parameters [9]. The continuum from /ç/ to /ɕ/ had 10 steps, the one from /ɕ/ to /ʃ/ had 17 steps. The vocalic continuum was created by interpolating between the anchor points [10].

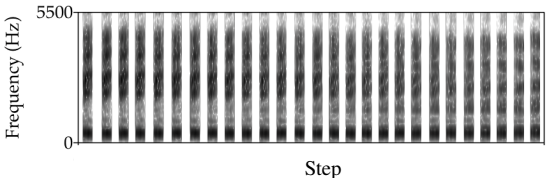

Figure 2: *Resynthesized 27-step continuum from [ɪ] to [ʏ].*

The synthesized vowels and fricatives were then spliced into the original *Fichte*- and *fischte*-recordings. From these 27 steps, we selected 14 test items through pilot-tests, evaluating the step-range listeners were especially sensitive to. In these tests, the stimuli were rated to sound naturally by native listeners.

2.1.3. Statistics

In total, there were 27.720 responses (99 subjects x 14 Steps x 10 Blocks x 2 continua) which were submitted 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 (KB, ZD or Control), STEP of the continuum, AGE and GENDER, the random factors are LISTENER and BLOCK.

3. Results

A linear mixed effects model (lme) with response as the dependent variable revealed significant results for the factors STEP in continuum ($p<.001$), CONDITION (KB vs. ZD) ($p<.05$) and AGE (old vs. young) ($p<.05$), for the /fɪçtə/ - /fɪʃtə/ continuum. In addition reaction time differed significantly between KB and ZD for both continua.

3.1. Factor CONDITION: KB vs. ZD vs. Control

The significant effect of CONDITION (KB vs. ZD) for the /fɪçtə/-/fɪʃtə/ continuum suggests that listeners adjust their interpretation of synthesized acoustic stimuli in accordance with their expectation: they perceive more /ʃ/ variants in CONDITION KB (Kreuzberg) than in CONDITION ZD (Zehlendorf). Although for the /fɪçtə/ - /fʏʃtə/ continuum CONDITION failed to show significance marginally (KB vs. ZD, $p = .06$), the tendency is the same: more /ʃ/-items were perceived in CONDITION KB than in CONDITION ZD (Fig.3). We take this as evidence that both perceptual cues and inferred social characteristics play a role in the categorization of speech stimuli. This evidence suggests that the mere allusion to where the speaker comes from is enough to trigger such inferences. Thus, we have replicated the perceptual

divergence effect within the confines of a larger urban area in a non-anglo-saxon context.

Fig. 3 shows that the distribution of responses differed between the two continua. A linear mixed effects model on all data showed an effect of the continuum on the responses: there were significantly more /ʃ/-responses (coded as "7" in Fig. 3,) in the continuum containing the rounded vowel /ʏ/ ($p<.05$) (top panel). This can be explained by the fact that the alveopalatal fricative is often accompanied by a more rounded vowel in spoken Berlin German. Thus for the /fɪçtə/ - /fʏʃtə/ continuum there are two cues (the change in the fricative spectrum and the change in the vowel formants) that trigger the perception of the alveopalatal variant. In contrast, in the /fɪçtə/ - /fɪʃtə/ continuum the absence of the expected change in formant values that goes hand in hand with a rounded palatalized fricative causes the lower portion of /ʃ/ responses.

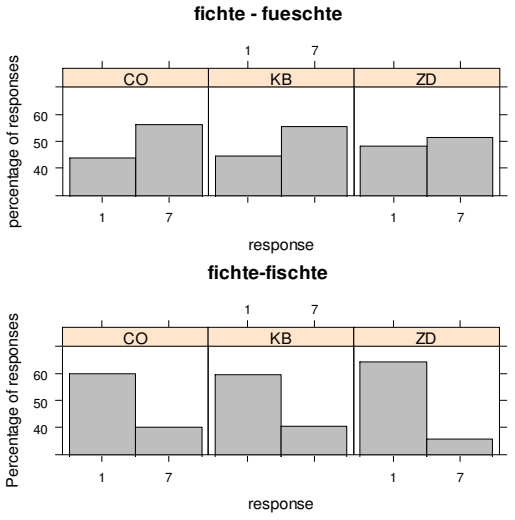

Figure 3: *Distribution of Responses for continuum 1: /ɪç/-/ʏʃ/ (top panel) and continuum 2: /ɪç/-/ɪʃ/ separated by Condition (in %): 1 = Fichte (left); 7 = fischte (right).*

Generally, listeners in the CONTROL group behaved more similarly to those in the KB condition compared to listeners in the ZD condition. This was the case for both continua. The expectation was that neither *Zehlendorf* nor CO should have an effect on the responses: CO does not offer any added information in which to interpret the stimuli and *Zehlendorf* was not expected to facilitate one response over another. The concept of *Zehlendorf* however does not appear to be completely neutral in the sense that it seems to be associated with an even more "standard German". Listeners in the ZD group hear more /ç/ variants than listeners in the CONTROL condition. For the continuum containing the rounded vowel (upper panel in Fig.3), the difference between ZD and CO was significant while the difference between ZD and KB was not ($p=.06$). For the other continuum, the difference between ZD and CO was not significant, and neither was the difference between CO and KB. There was however, a significant effect in the comparison of ZD with KB. We interpret this finding in terms of a sound change in progress with a locus in urban centers as Berlin and Hamburg whereby the pronunciation of /ç/ vanishes in favor of the rounded alveopalatal variant /ʃ/.

3.2. Factor AGE: old vs. young

For the /fɪçtə/ - /fɪʃtə/ continuum, younger listeners perceive more /ʃ/- variants than older listeners (Fig. 4, light grey bars) presumably because of the prevalence of these variants in their environment.

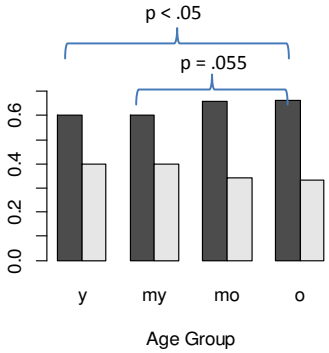

Figure 4: *Distribution of Responses by Age Group (proportion): dark grey = Fichte , light grey = fischte.*

For the /fɪçtə/ - /fʏʃtə/ continuum there was no AGE effect. This may be explained by the fact that the [fʏʃtə] variant is frequently used also by (older) speakers from eastern districts of Berlin.

3.3. Reaction Time

As for the RTs, results indicate that subjects needed statistically significantly more time for deciding which category was heard when they were assigned to group *Kreuzberg* compared to group *Zehlendorf.*

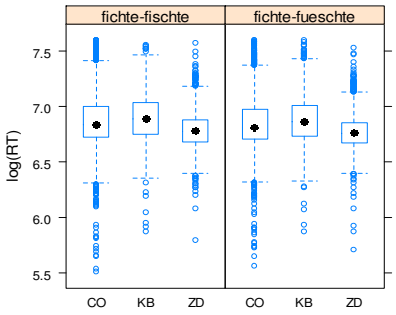

Figure 5: *Logarithmic Reaction Times (logRT) for the two continua plotted by CONDITION (Control (CO), Kreuzberg (KB) and Zehlendorf (ZD)).*

While this difference seems minute from the graph, the result is significant for both continua ($p<.05$). Moreover, in the CONTROL condition in which listeners were not co-presented with labels denoting Berlin districts, listeners were slightly yet insignificantly faster than in the KB condition and slightly but insignificantly slower than in the ZD condition (see Figure 5). Even though ZD was indicated by pre-tests not to trigger any associations, it seems that listeners also deviated from a baseline (Control). It remains to be tested if the association with 'negative' primes implies a larger cognitive load and thus requires more time, than the co-processing of 'positive' information.

4. Summary

Our results indicate that perceptual categorization of acoustic stimuli in a forced choice perception task is not context free – listeners use associations and any information available to them to analyze and make sense of the speech presented to them. Results indicate that listeners hear /ʃ/ more often than /ç/ when co-presented with the label Kreuzberg, a neighborhood highly associated with speakers who produce this alternation. However, a comparison to the CO-condition does not reach significance for either continuum. For the continuum with the rounded vowel, listeners in the ZD group deviated from the CO-group. Most importantly though, listeners in the Zehlendorf group always heared fewer /ç/ compared to listeners in the Kreuzberg group. While Zehlendorf is not immediately associated with any linguistic stereotype, it may be the lack of any accessible indexical information, marking it as a district in which speakers adhere to the standard. Our data also shows an effect of AGE whereby speakers up to the age of about 29 hear more often /ʃ/ than older speakers. We are currently adding more data to evaluate this claim.

For a subset of listeners we also gathered reaction times. These results indicate that listeners need slightly yet significantly more time to respond to test items when they were in group Kreuzberg, presumably when they had more (negative) associations to process, compared to listeners in group Zehlendorf. Moreover, for continuum 2 listeners in ZD heard more often /ç/ compared to the control group, indicating that priming generally has an effect of the phonological categorization of acoustic stimuli. In summary, categorization differences for listeners in KB versus ZD differed significantly only within the /fɪçtə/- /fɪʃtə/ continuum. For the /fɪçtə/-/fʏʃtə/ continuum, the redundant and enhancing cues to the phonological quality of the fricative, and the uncommon pronunciation pattern for adolescents from Kreuzberg prevent listeners to select more Fichte responses. Through added cues to the quality of the fricative, the lexical effect prevails and the indexical effect is overridden. We nevertheless conclude that generally, phonological (lexical information) and also indexical information or associations are used to process speech.

5. Discussion

The results shown here are much in line with exemplar theoretic accounts [see 8 for an overview] positing that acoustically rich signals are stored in memory along with linguistic and social indexing. Identical stimuli are interpreted differently when indexical information [12] enters into the categorization process. It appears that listeners need significantly more time to process the accessible memory traces of the indexical information in conjunction with the lexical information when the information or the linguistic stereotype is negative. Listeners adjust their categorization of synthesized acoustic continua in accordance with their expectation, suggesting that the interpretation of fine phonetic detail is also phonological in the sense that it can be recognized by listeners and that they categorize speech stimuli differently depending on who they believed they listened to.

The data presented here strongly points at a phonological categorization process that is context dependent. Yet, the differences in results for the two continua suggest that further work is necessary to evaluate the trading relationships between lexical and indexical information in speech (when do listeners respond to lexical information and when to indexical information). Nevertheless, our data provides evidence for the idea that phonetics is the meaningful and socially interpretable expression of phonology ("phonetics is defined as the semantics of phonology", [7]). Within phonological categories, there is much room for the interpretation of the malleable middle ground that speakers make use of to show their group affiliation or to distance themselves from other groups. This in itself may be a reason as to why sound change can happen [13]. The perceptual divergence effect shown here strongly points at a very close relationship between linguistic structures and instances of use. Both are continuously shaped over time and get updated throughout the lifetime of a speaker. Exemplar models offer an integrated account of the interplay and link between production and perception and explain our result very well.

6. Acknowledgements

This work was supported by the German Ministery of Education and Research (Grant Nr. 01UG0711). The authors would like to thank Claudia Blankenstein, Anna Theis and Sophie Arendt for their help with recording and coding data. We sincerely thank Jana Brunner for the creation of the synthesized stimuli.

7. References

[1] Auer, P. & Dirim, I. (2004) "Türkisch sprechen nicht nur Türken. Über die Unschärfebeziehung zwischen Sprache, und Ethnie in Deutschland", Walter De Gruyter.

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

[3] Hay, J., Nolan, A. & K. Drager (2006) "From Fush to Feesh: Exemplar Priming in Speech Perception", The Linguistic Review 23: 351-79.

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

[5] Jannedy, S., Weirich, M. & Brunner, J. 2010 "Perceptual Evidence for Allophonic Variation of the Palatal Fricative /ç/ in Spontaneous Berlin German", J Acoust Soc Am., 128, 2458.

[6] Jannedy, S., Weirich, M. & Brunner, J. (2011) "The Effect of Inferences on the Perceptual Categorization of Berlin German Fricatives." In Proceedings of the International Congress of Phonetic Sciences (ICPhS), Hong Kong; 962-965.

[7] Beckman, M., and J. Pierrehumbert (2004) "Interpreting 'phonetic interpretation' over the lexicon", Papers in Laboratory Phonology VI, Cambridge University Press, Cambridge UK, 13-38.

[8] Barlow, M. & Kemmer, S. (2000) "Usage Based Models of Language", Stanford: CSLI Publications.

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

[10] 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.

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

[12] Foulkes, P. (2010) "Exploring socio-indexical knowledge: A long past but a short history", Journal of Laboratory Phonology 1, 5-39.

[13] Ohala, J. (1981) "The listener as a source of sound change", In C. S. Masek, R. A. Hendrick & M. F. Miller (eds.), Papers from the Parasession on Language and Behavior. Chicago: Chicago Ling. Soc., 178-203.