Stefanie Jannedy* and Melanie Weirich

Sound change in an urban setting: Category instability of the palatal fricative in Berlin

Abstract: The differential categorization of identical stimuli depending on the presence of a prime is described as a perceptual divergence effect. We examined whether native listeners of the Berlin vernacular of German categorized identical acoustic stimuli differently in the explicit context of the names of two different districts of Berlin, assuming that listeners infer social information and linguistic stereotypes based on the names of these neighborhoods (Kreuzberg vs. Zehlendorf). All listeners categorized natural acoustic stimuli with synthetic fricatives synthesized along a continuum ranging from /ç/ to /ʃ/ as either *Fichte* /fiçta/ (‘spruce’) or *fischte* /fiʃtə/ (1st person sg. ‘to fish’). This variable was chosen because many young multiethnic speakers of Berlin German pronounce /ç/ as [ʃ] or [ɕ], and this alternation is highly associated with speakers with a migrant background from Kreuzberg. Data were gathered in a forced-choice identification task, and, for a subset of the participants, reaction times (RTs) were also gathered. Results indicate a differential categorization pattern depending on (1) the co-presented information, i.e., Kreuzberg, Zehlendorf, or none (control), and (2) the age of the listeners, with older listeners being more affected by the co-presented information. While older listeners categorized significantly more /ʃ/ sounds in the context of Kreuzberg than in the control or Zehlendorf condition, younger listeners rated most /ʃ/ sounds in the control condition (no added information). Results are interpreted in terms of a potential sound change in progress: the loss of the phoneme contrast between /ç/ and /ʃ/.

*Corresponding author: Stefanie Jannedy: ZAS (Center for General Linguistics), Berlin. E-mail: jannedy@zas.gwz-berlin.de
Melanie Weirich: Friedrich-Schiller Universität, Jena. E-mail: melanie.weirich@uni-jena.de

1 Introduction

In many European cities, researchers have noticed and studied the emergence of linguistic variation and grammatical innovations introduced by young speakers from multiethnic urban neighborhoods (Multicultural London English: Torgersen et al. 2006; Kerswill et al. 2008; Straattaal (Netherlands): Appel 1999; Nortier

Nevertheless, most language research in Germany has neglected variation in speech based on social factors and has predominantly focused instead on lexical or pronunciation variation along geographical and dialectal dimensions (e.g., Deutscher Sprachatlas: Herrgen 2007; Kehrein 2009). While researchers have noticed the emergence of a speech style used predominantly by adolescent speakers in larger urban areas of Germany such as Berlin, Hamburg, and Mannheim (Auer 2003; Dirim and Auer 2004; Androutsopoulos 2006; Wiese 2006, 2012; Keim 2008; Jannedy 2010) and have made many interesting observations on the speech of these young speakers, there is relatively little quantifiable empirical data on this variation. Over the last couple of years, though, data-driven and corpus-based investigations on the Berlin variety of this youth style (here named Hood German) have been conducted (Jannedy 2010; Jannedy, Weirich, and Brunner 2011; Jannedy and Weirich 2012). For example, in our corpus study (Jannedy 2010), we analyzed over 1,400 tokens to explore the usage and function of the discourse particle so in this youth-style vernacular. Our analysis showed that it is primarily used for quoting as well as hedging and non-committal references, much as like is used in English (D'Arcy 2007; Fox Tree and Tomlinson 2008).

So far, there is only some agreement about how this multiethnolect should be named. Both Kanak Sprak (Zaimoğlu 1995) and Türkenslang (Auer 2003) are unfortunate labels, as they have – for different reasons – derogatory connotations. This is because both labels reflect the presumed speakers of this vernacular, namely the children and grandchildren of the first Turkish migrants to Germany. The name Kiezdeutsch, a compound built from the noun Kiez meaning neighborhood plus the noun Deutsch meaning German, refers to the locale where this variety is spoken rather than the speakers (Wiese 2009, 2012). It is meant to denote the German spoken in local neighborhoods in urban areas. Unfortunately, though, Kiez is a term that is only locally interpretable and means different things in different places. For instance, in and around the urban area of the northern city Hamburg, Kiez refers solely to the red-light district of the city, while in Berlin, Kiez is an endearing term for a local neighborhood. Other urban areas such as Frankfurt, Munich, or Cologne do not use this term at all.

We would like to introduce the term (neighbor-) Hood German, which makes no reference to the ethnicity or cultural background of the speakers, or to the size of the site; it is locally interpretable and seems to be a more appropriate, less judgmental, and more geographically independent descriptor. Hood German refers to the style of German spoken by young adults and adolescents in multi-ethnic neighborhoods (we have made recordings of adolescents and young adults...
with ethnic and cultural roots in Lebanon, Palestine, Turkey, Russia, Croatia, Britain, Poland, and Vietnam) and their monoethnic monolingual German peers who socialize in the same neighborhood. Hood German is a youth-style multiethnolect spoken predominantly in larger urban areas of Germany.

Recently it was suggested (Wiese 2011) that Hood German is a new dialect of German. We disagree with this assessment, as per definition, a dialect is a regionally distributed variant of a language. Rather, we argue that Hood German is its own speaking style (Hay, Jannedy, and Mendoza-Denton 2010), which has been observed in Berlin (Wiese 2009, 2012; Jannedy 2012; Jannedy and Weirich 2013), but also in Böblingen, Munich, Nürnberg, and Urbach near Stuttgart, (Füglein 2000, cited in Auer 2003), Frankfurt (Tertilt 1996), Hamburg (Auer 2003; Dirim and Auer 2004), and Mannheim (Keim 2008). Descriptions include primarily only morphosyntactic alternations (Füglein 2000). Variation in the sound production domain is generally only impressionistically described (Auer 2003).

From a historical perspective, the development of Hood German is rooted in the recruitment of foreign workers from Turkey in the early 1960s. Germany and Turkey signed a foreign workers recruitment agreement that brought many Turkish men to Germany to support the rebuilding of German industry. It was intended that workers would return to their respective home countries within two years, but many stayed and brought their families from Turkey or started families in Germany. No formal language instruction was offered to the workers or to their families. Language interpreters helped workers manage on the job, or they learned enough German on their own to get by. Pressure to learn German was alleviated by the quickly developing infrastructure in the larger urban areas (grocery stores, insurance agents, travel agencies) that made it possible to converse in the native language and get by with rudimentary German (Aksoy 2010). This learner variety of German was termed Gastarbeiterdeutsch (Keim 1978).

Dirim and Auer (2004: 215) state that while the urban youth-style multiethnolect shows several grammatical features of Gastarbeiterdeutsch (Keim 1978; Aksoy 2010), it is not simply a fossilized version of this learner variety. Their claim is based on the variability found in the data, which shows that (1) often the correct grammatical features of more standard German are exchanged for the multiethnic variants, meaning that both variants are known to the speaker, and (2) in some situational contexts, multiethnolectal features are avoided all together. Moreover, they argue, the large number of monoethnic Germans using the urban youth-style variety speaks against a learner variety. That is, not only second, third, and fourth generation migrant speakers with different language backgrounds speak this primary ethnolect (Auer 2003) but also monolingual German adolescents who were born, raised, and socialized in Germany and live in the same neighborhoods (Dirim and Auer 2004; Wiese 2009, 2012). In this paper, we
will refer to the latter group of adolescents as monoethnic speakers. Multiethnic speakers are those who have at least one other strong language in addition to German and a strong cultural background other than German.

Auer (2003) provides a model explaining the propagation and spread of this youth-style multietnolect through different kinds of speech communities. In these communities, the primary ethnolect shows a great deal of within- and across-speaker variability and yet displays many common features (see Auer [2003] for examples of morphosyntactic alternations and a list of phonetic/phonological variants). The group characterized by this ethnolect predominantly consists of adolescents with a multiethnic/multicultural background from different urban areas around Germany and peers who closely associate with them. Auer describes the secondary ethnolect (see also Androutsopoulos [2001] for a stylized version of the primary ethnolect typically used by comedians, in films, or in newspapers), which displays many features of the primary ethnolect but is often extended with features that are not part of the inventory of the primary ethnolect. The tertiary ethnolect is widely known to many school children and adolescents who do not have direct contact with speakers of the primary ethnolect and who even live outside larger urban areas or in different districts of the larger city. It is copied from the secondary ethnolect and used as a mockery of the speaker group rather than as an alignment with them. It serves to stereotype those from whom this style originates as uneducated and aggressive (Auer 2003). Through contact among different speaker groups— which often include monolingual Germans who live in the same neighborhoods, go to the same schools, or hang out at the same youth centers— some features of the primary ethnolect have been de-ethnitzed and are less strongly associated with the speech of people from certain districts of the city. This process is motivated by the wish to identify with street culture and distance oneself from mainstream culture (Auer and Dirim 2003). One salient phonetic feature of Hood German that we believe has undergone de-ethnization is the pronunciation of the palatal fricative /ç/ as the postalveolar fricative [ʃ] or, more frequently, as an unrounded alveopalatal fricative [ɕ] (Dirim and Auer 2004: 208; Jannedy et al. 2011) so that the pronunciation for the first person singular pronoun ‘I’ ich /ɪç/ becomes [ɪɕ]. According to our informal observations, some speakers of the primary ethnolect alternate between [ç] and [ɕ], even within the same breath cycle in fluent spontaneous conversations, while others seem to use only [ɕ]. Auer and Dirim (2003) also report that speakers from an all-male gang (Tertilt 1996) were able to switch register when being interviewed on the radio.

According to Mielke’s (2007, 2008) analysis of sound patterns (PBASE), only 23 languages have /ç/ in their phoneme inventory (only 5% of the world’s languages, according to Ladefoged and Maddieson 1996), and only 3 languages (Ger-
man, Polish, and Margi) contrast /ç/ and /ʃ/. In certain varieties of the central German dialects ranging from Cologne in the west to Dresden in the east, the phonemes /ç/ and /ʃ/ have merged to /ɕ/ (Herrgen 1986; Dirim and Auer 2004; Hall 2013). That is, in some dialects, the phonetic forms of phonological /ʃ/ and /ç/ can include [ɕ] and [ʃ] (Dirim and Auer 2004; Hall 2013).

Our ongoing work on this alternation also shows that some young Berlin speakers with a monolingual and monoethnic German background do not produce a contrast even when reading lists of minimal pairs. In more canonical iterations, differences are marked by rounding of the vowel before /ʃ/ (Lasch 1928; Schönfeld 1986: 225; Schlobinski 1987: 135; Schönfeld and Schlobinski 1995; Porrath 2007), so that we get [fyʃtə] or [fyɕtə] for *fischte* (3rd p. sg. ‘to fish’), but also [fɪʃtə] or [fɪɕtə] for some speakers. Rounding of the vowel in this context is presumed to be a secondary perceptual cue that is often found in the speech of genuine Berliners and is not necessarily associated with young multiethnic speakers from Kreuzberg.

Lasch (1928: 261) first noticed a /ʃ-ç/ merger in the speech of children in her early description of Berlin German. Almost 60 years later, Schönfeld (1986: 230) mentions in passing an alternation by older Berliners whereby they produce ‘ch’ as a sound intermediate between [ç] and [ʃ]. And at the turn of the millennium, Auer (2003) notes that this alternation is unknown in Hamburg but occasionally observed in Berlin. Dirim and Auer (2004: 208) notice that (some) Turkish adolescents and adults in Berlin produce /ç/ as [ç] or [ʃ]. They describe this variation as very unusual for Berlin, as Berlin does not belong to the middle German dialect area (Koronalisierungsgebiet) in which this variant is to be expected (Herrgen 1986; Auer 2003; Hall 2013). All in all, the allophonic realization of /ç/ as [ʃ] or [ç] has been observed at various times in the speech of Berliners (Jannedy et al. 2010) but has not yet been systematically explored and has not been described in any more recent descriptions of the Berlin dialect.

We hypothesize that young speakers from multiethnic neighborhoods are using this alternation as a social marker, which may or may not have sprung up independently of already existing rudimentary patterns from different sources such as Berlin German, central German dialectal influences, and youth-style multiethnolectal usage patterns. In summary, we find that many monolingual Germans and multiethnic adolescents with various language backgrounds who consider German to be their mother tongue display the alternation of /ç/ produced as [ç] and [c] in their speech.

The origin of this alternation is rather unclear. Neither Turkish nor Arabic have /ç/ or /ɕ/ in their phoneme inventories. Therefore, the usage of [ç] is rather unlikely to stem from language interference. Moreover, if multiethnic adolescents were copying, or influenced by, standard German, then they should produce [ç]
Stefanie Jannedy and Melanie Weirich

(which occurs in the standard) and not [ɕ] (which does not). Rather, the usage of [ɕ] conforms to the merged forms found in the central German dialects. There is sparse mention in the literature of speakers of the Berlin dialect palatalizing and producing mixed forms between /ç/ and /ʃ/; however, there is neither a recent nor a detailed description of this alternation.

We can only speculate why young multiethnic speakers use an allophone of /ç/ which at this point is uncommon and therefore ‘socially salient’ in the Berlin area. Fricatives with higher level frication noise like /s/ and possibly /ɕ/ seem to lend themselves to socially differentiated usage cross-linguistically (Strand 1999; Stuart-Smith 2007; Halonen 2012). From this, it seems that the young linguistic innovation by adolescent speakers from Berlin is one of the driving forces for sound change in this urban area. This internal driver of sound change in Berlin and its vicinity is compounded by an external driving force, the spread of the central German dialects into Berlin. We may be observing a new wave of the merger of /ç/ and /ʃ/ to /ɕ/, with a change possibly being due to the relative instability of this rare fricative category cross-linguistically, especially since the German fricative space is relatively crowded with six voiceless fricatives /f s ʃ ç x h/.

We will return to this in the discussion section.

1.1 Perceptual learning: Convergence and divergence

Usage-based or experience-based models of language (for an overview, see Barlow and Kemmer 2000) operate under the assumption that there is a very close link between abstract linguistic representations and instances of language use. In such theoretical accounts, language performance and language competence are not cognitively separated – both are continually shaped over time through exposure to and experience with language (Pierrehumbert 2002; Bod et al. 2003). Both experience and exposure are vital shaping forces of the human language faculty and thus part of the perceptual learning process. Research from the visual perceptual domain (Saffran et al. 1996; Watanabe et al. 2001) shows that perceptual learning does not require attention; mere exposure below the level of awareness is sufficient to recognize patterns. We suggest that mechanisms of this sort also apply to the auditory perceptual system. This explains why probabilistic characteristics of language play a key role in explicating language variation and language change (Wedel 2012; Wedel et al. 2013).

It is well established by now that, even through adulthood, exposure to fine phonetic detail in ambient speech productions continues to subtly influence a speaker’s own speech and causes changes over time. For example, Harrington (2006, 2007) showed that over a span of about 50 years, the lax vowel /ɪ/ in happy,
as produced by HRH Queen Elizabeth II (in her annual Christmas addresses), developed tensing and thus converged towards the speech of ordinary people. Also, research shows that speakers mediate social distance (Giles and Coupland 1991) by adjusting their speaking styles. Pickering and Garrod (2004) found that speakers can shift their pronunciations and converge towards their addressees, the people they are conversing with. Bell (1984) argues that a speaker’s audience, including over-hearers (people listening in on a conversation) and also referees (third parties not present at an interaction but contextually salient), can have an effect on language choice. Hay, Jannedy, and Mendoza-Denton (2010) showed that the U.S. talk show host Oprah Winfrey displayed various degrees of monophthongization of /ay/ depending on the ethnicity of the person (referee) she was speaking about. Thus, we assume that there can be active as well as passive acts of convergence or entrainment and that passive instances can be extremely subtle and below the level of awareness. Gregory and Webster (1996) presented an analysis of interviews between a CNN talk show host and his guests and found that lower status partners accommodated in their voice spectra towards the higher status partners. While some (Goldinger 1998; Pickering and Garrod 2004) view convergence as automatic and mechanical, others believe that converging or entraining towards a model involves choices (Giles and Coupland 1991; Pardo 2006; Babel 2009) and can be deliberate.

It seems clear that listeners are very sensitive to linguistically and socially determined variation in the speech signal (Ladefoged and Broadbent 1957; Campbell-Kibler 2010; Drager 2010; Foulkes 2010). Speech signals are acoustically rich and encode speaker characteristics, the speaking situation, and possibly through whom and where the signal was encountered before (Pierrehumbert 2002). These acoustically rich signals are stored as indexed phonetic density distributions (exemplar clouds), which allow a listener to abstract over the encountered input and build connections between social attributes and linguistic cues. When linguistic events become – consciously or unconsciously – highly associated with specific groups or individuals, perceptual learning and interpretation of the acoustic stimuli in light of the inferred social characteristics of a speaker takes place.

Perceptual convergence or divergence effects have been investigated with various experimental perception paradigms (Strand and Johnson 1996; Niedzielski 1999; Hay et al. 2006; Brunelle and Jannedy 2007, 2013; Hay and Drager 2010). In general, listeners are asked to categorize identical stimuli across different priming conditions. These priming conditions influence the type of speaker inferred by the listeners.

In speech production, a divergence effect can be observed when a speaker reinforces his or her identity as different and separate from another speaker (Bourhis and Giles 1977). A perceptual divergence effect was found in a classic
speech perception study by Strand and Johnson (1996). They exploited the fact that women produce fricatives like /s/ and /ʃ/ at higher spectral frequencies than males. To evaluate listeners’ sensitivity to this fine phonetic detail, they tested listeners on very prototypical as well as very non-prototypical sounding male and female voices. They played a resynthesized nine-step acoustic continuum from /s/ to /ʃ/ to listeners and found that the prototypical sounding male and female voices were rated differently. Female voices pushed up the perceptual boundary towards higher-frequency /s/, while the male voices shifted down these boundaries towards /ʃ/. In an audio-visual integration task, matching the non-prototypical voices with pictures of male and female faces, Strand and Johnson found that listeners had a preference for /ʃ/ with the co-presentation of a male face and a preference for /s/ with the co-presentation of a female face. Their Face-Gender effect shows that auditory and visual information are integrated and that the expectation that listeners have of a speaker strongly influences the categorization of identical acoustical stimuli. The integration of auditory stimuli with information delivered through the visual channel seems to be an important point in associating speech forms with social categories or speaker groups. We explore this effect in the current study.

Niedzielski (1999) first exploited this effect and showed that listeners in a vowel matching experiment were biased by the inferred origin of the speaker, which they gathered from reading a label on a response sheet. Both speakers from Detroit (Michigan, USA) and those from across the border in Canada have Canadian raising of the diphthong /aʊ/. Nevertheless, speakers from Detroit seem to be unaware of this and strongly associate this raising with speakers from Canada. Niedzielski asked listeners from Detroit to match naturally recorded speech tokens (recorded from a Detroit speaker) containing the diphthong /aʊ/ to vowel tokens synthesized along a vowel continuum. There were two conditions in this experiment: either the word Detroit or Canadian was written on top of the response sheet. Detroit listeners selected fewer raised variants of the diphthong when they inferred from the response sheet that the speaker was from Detroit compared to when they inferred from the response sheet that the speaker was from Canada. Niedzielski called this a perceptual divergence effect along the lines of national identity and concluded that listeners use inferred social characteristics and expectations such as a speaker’s supposed dialect for the categorization and interpretation of speech stimuli.

Hay et al. (2006) conducted a follow up of this study using the same methodology as Niedzielski but in the Australian–New Zealand context. There is evidence that lax /ɪ/ (in ‘hid’) is more raised and fronted in Australian than in New Zealand (NZ) English (Bell 1997; Watson et al. 1998). Hay et al. (2006) asked listeners to match synthesized tokens from a continuum of more to less raised /ɪ/
variants to recorded words containing this vowel as spoken by a native speaker of NZ-English. Here too, when *Australian* was written on the response sheet, listeners selected more fronted versions of the /ɪ/, even though the majority of the NZ listeners stated that they knew the recordings were from a NZ speaker. Listeners integrated the read information from the response sheet with the interpretation of the auditory stimulus. It seems that listeners display a perceptual divergence effect even if they are aware of the geographic/national identity of the speaker.

This points to a mechanism that seems to take place somewhat automatically, even if the integration of all information in some instances misguides perception rather than facilitates it. What we mean is that the human cognitive system is built in such a way that perception works with and without awareness (Watanabe et al. 2001). Interpolating from the visual perceptual domain, listeners may be able to bootstrap information from stimuli that they do not consciously pay attention to or that they are not aware of.

Hay and Drager (2010) conducted a second study in the New Zealand–Australian context, showing that even abstract objects denoting national identity have an effect on categorization: for the Australian condition, listeners briefly saw a stuffed toy koala, and for the NZ condition, listeners saw a stuffed kiwi bird. The koala and the kiwi bird are national emblems of Australia and New Zealand, respectively. As expected, listeners in the koala context matched more raised synthesized variants to the stimulus words than listeners in the kiwi context. From the results of this word-matching paradigm, the authors conclude that 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 national/geographical identity. It seems that the forces that govern categorization and interpretation of speech input are unclear.

Moreover, perceptual learning effects can also be found in suprasegmental parameters of speech. Brunelle and Jannedy (2007, 2013) asked listeners to categorize Vietnamese lexical tones as belonging to either the northern (6 tones) or southern (5 tones) dialect. Instructions were given in the local southern dialect or in the northern dialect. Generally, categorization was better when the experimenter instructed the listeners in the local dialect and when the listeners had increased exposure to the experimenter.

Inferred social characteristics based on labels (for example, the name of a neighborhood known for its multiethnic and young population) can facilitate listeners’ inferences regarding the social characteristics of a supposed speaker, but might vary between age groups. This is because, as the variationist literature shows, generally adolescent speakers use the most innovative phonological forms (Trudgill 1974; Ash 1982; Eckert 1997; Kirkham and Moore 2013).
1.2 Aim of the study

In this paper, we investigate a merger in progress by means of a perception task in which listeners (in different age groups) were implicitly asked to associate specific speech forms with districts of Berlin. One district – Kreuzberg (KB) – is predominantly multiethnic, is on average rather young, has a low income and a high unemployment rate, and is considered hip and fashionable. The other – Zehlendorf (ZD) – is monoethnic, almost rural with many single family homes, and middle-class with a higher mean income and a low unemployment rate. It is also an area where people do not use the Berlin dialect but adhere to more standard forms of German (Schlobinski and Wachs 1983). The decision to contrast these two neighborhoods was driven by wanting to maximize the social contrast between the two West Berlin neighborhoods and to exploit the fact that Zehlendorf was rated prior to this study as middle class. A third group with no mention of any area of Berlin serves as the control condition (CO). The assumption is that the control condition will serve as a baseline for an interpretation of the acoustic stimuli in the absence of an immediate concept to bootstrap from or context within which to judge the stimuli. We expect listeners in the Kreuzberg (= Hood German) condition to categorize more stimuli as /ʃ/ than those in the Zehlendorf and control conditions. Moreover, we suggest that age plays a role, with older listeners being more susceptible to the prime than younger listeners, pointing to a sound change in progress.

Thus, the aim of this study is to evaluate the effect of implicitly coded social information (relating to neighborhood in our case) on the categorization of acoustic stimuli with synthesized fricatives ranging from /ç/ to /ʃ/. The alternation in the production of Fichte with [ɕ] or [ʃ] is salient and, for some listeners, socially marked. The merger of /ç/ and /ʃ/ in this vernacular is discussed in the printed media as the demise of the German language (Berliner Kurier, 21.10.2010). In the perception of the general public, the variable production of /ç/ as [ʃ ɕ] is attributed to speakers from Kreuzberg in general, and to young speakers with a migrant background in particular.

Based on the idea that cognitive processing of a speech stimulus is affected by non-speech observations, we carried out a detailed investigation within the confines of a single large urban area to determine whether there was any evidence for perceptual divergence. We expected that linguistically naïve listeners would interpret identical stimuli differently in the context of co-presented information that lets them infer speaker heritage and social background. The implication would be that listeners have a fairly detailed awareness of supposed linguistic similarity and linguistic difference and which geographical regions these are
affiliated with. In addition, the potential effect of the listener’s age on perceptual divergence was investigated.

2 Method

A forced-choice identification task with listeners in three age groups (young, middle, old) was carried out. This experimental paradigm combines features of the methods from Niedzielski (1999), Hay et al. (2006), and Hay and Drager (2010) in a simple word-identification response task. Listeners from Berlin and Brandenburg were asked to classify synthesized stimuli from two acoustic continua as either *Fichte* (/fiçtə/) (‘spruce’) or *fischte* (/fyʃtə/ or /fyʃtə/) (1st and 3rd p. sg. past tense of ‘to fish’). The two continua differed with respect to the vowel quality in *fischte*, since rounding of the vowel is a feature of Berlin German (Schönfeld 1986: 225; Schlobinski 1987: 135) but less likely to be associated with speakers of Hood German. Listeners were randomly assigned to two different priming conditions (i.e., two neighborhoods from Berlin) and a control condition with no prime.

2.1 Priming conditions

The first priming condition for the following perceptual divergence experiment was the name of the neighborhood Kreuzberg (KB). The second prime or condition was the name of the district Zehlendorf (ZD), which was previously used as a stimulus by Schlobinski and Wachs (1983). The third condition was a control condition (CO) with no additional information given. We will refer to these priming conditions as different Hood (neighborhood) groups.

2.2 Stimuli

The two German words *fischte* and *Fichte* were used as stimuli for the perception test. They were chosen because, whereas in standard German, as defined by the Duden (Mangold 2006), and particularly in northern German dialects, *Fichte/fischte* is a minimal pair, in Hood German the contrast is often not produced even when speakers read lists of minimal pairs.

In addition, the stimuli varied in the vowel quality of *fischte*: in one *fichte–fischte* continuum, the vowel quality before the fricative was maintained as the lax high front vowel [i], while in the other continuum, the vowel quality changed from unrounded to rounded, as in [ɪç] to [ʏʃ].
2.2.1 Creation of the stimuli

To create the stimuli for the two continua, synthetic fricatives were generated using the Klatt protocol and then spliced into natural speech. To do this, a 14-year-old male native speaker of Berlin German was recorded at a sampling frequency of 48 kHz saying the German word \textit{Fichte} [fɪçtə], the two variants of \textit{fischte}, [fɪʃtə] and [fʏʃtə], and an intermediate form of the fricative, as in [fɪçtə], that was somewhere between the two extreme renditions. It was important to use the voice of an adolescent, since Hood German is mainly a youth style. Two continua were generated. In the continuum from [fɪçtə] to [fɪʃtə] (which we call continuum \textit{ii}, where the \textit{ii} stands for the vowels in the two extreme steps), only the fricative portion was varied (including one of the steps shown in Figure 1), while the stressed vowel was kept as in \textit{Fichte} [fɪçtə]. We varied the fricative to make a [ç] – [ɕ] – [ʃ] continuum using the Klatt-Synthesizer (Klatt and Klatt 1990). First, the synthesis parameters (frequency and amplitude of each prominent spectral peak) were derived from the natural renditions of [ç, ç, ç]. Then these peaks were taken as anchor points, and the intermediate sounds between these points were created by morphing between the measured peaks in 10 steps from [ç] to [ç] and in another 16 steps from [ç] to [ʃ] (cf. Ghosh et al. [2010] for more details). The differences in the number of steps between the anchor points results from the fact that it turned out to be more difficult to synthesize in a smooth way from [ç] to [ʃ] than from [ç] to [ç].

In the second continuum (continuum \textit{iy}), both the fricative and the vowel were synthesized. In addition to the 27 steps for the [ç] – [ç] – [ʃ] continuum, the vowel was shifted from the [i] in \textit{Fichte} towards a more rounded lax front vowel like [y] (see Figure 2). This was also done in a stepwise fashion, with a result of 27 test items for the second continuum (cf. Mitterer et al. 2011).

The remaining sounds, /f/, /t/, and /a/, were taken from one of the recorded words for all steps and both continua. The synthesized vowels and fricatives

Fig. 1: Resynthesized 27-step continuum from [ç] to [ʃ] used for both continua.
were spliced into this recording. To avoid listener fatigue, not all 27 steps of each continuum were included in the perception test. The first and last steps of each continuum were selected to build the most extreme end points, [ç] and [ʃ]. Six steps were selected from the part directly neighboring the middle anchor [ɕ] (three to the left and three to the right), since this part of the continuum was the most ambiguous and difficult to categorize, and thus most interesting for this study. For the remaining five stimuli, we selected two after the start point [ç] and three before the end point [ʃ]. In pilot tests with native Berlin listeners from Humboldt University in Berlin, the stimuli were rated as sounding very natural.

2.3 Description of the experiment

A forced-choice identification task was run to test for the perceptual category boundary between /ç/ and /ʃ/. Listeners heard a stimulus from one of the two continua over headphones (Sennheiser HD 280 professional with a frequency range from 8–25,000 Hz) and were instructed to press the corresponding button on a response box (see Figure 3). 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. The left and right buttons were labeled with Fichte or fischte, with the order balanced across subjects. As an aid to listeners and in order to lighten the cognitive load on memorizing which button to press, the response button for Fichte (‘spruce’) was always coded green and the one for fischte (‘fished’) was always coded blue. Since we wanted to test for perceptual divergence, three conditions were set up for three different listener groups: the card on the response box showed in handwriting either the word Kreuzberg (KB) or the word Zehlendorf (ZD) or was left blank and did not list any condition at all (control, CO). A sample response box setup is shown in Figure 3.
For all listeners, identical procedures were applied. Because our experimental set-up crucially depended on listeners noticing the experimental condition they were tested in, we also wrote the words Kreuzberg or Zehlendorf (or nothing) on the subject information sheet that listeners 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 the condition aloud once so that it was audible to the listener. In this way the listener’s attention was implicitly and subtly directed to his/her group membership under the assumption that s/he would derive inferences from that.

Listeners were presented with 11 blocks of data in which each token from each continuum occurred once. They rated the stimuli from both continua in a randomized order. There was no break while listeners were tested. The first block served as a familiarization with the task and was therefore excluded from the analyses.

We also gathered reaction times (RTs) for a subset of the participants (81 listeners), since we expected additional information to contribute to the processing load and increase the time it takes for the listener to respond. RTs were measured from the beginning of the stimulus to be rated, which included a 40 ms silence prior to the 700 ms stimulus. The inter-stimulus interval was 2000 ms. Our experiments were run on Superlab 4 with the response box connected to a Macbook Pro.
2.4 Listeners

In total, data from 132 (32 male and 100 female) native monoethnic Berlin/Brandenburg listeners were included in the analyses. None of them showed features of Hood German in their speech. We also collected a variety of meta-data on the subjects, such as the neighborhood they are now living in and the neighborhood they grew up in, the language(s) they speak at home, music taste, favorite hangouts and neighborhoods, jobs and education, as well as whether they had children or grandchildren in school or kindergarten, to see if any of this could have an effect on the outcome of the data.

Listeners ranged from 19–61 years old, with an average age of 32.04 years ($SD = 12.1$). None of them reported speech or hearing problems. Table 1 summarizes the information on the three age groups tested. The separation of listeners into three age groups was done to evaluate the effect of age in conjunction with the effect of the prime on the categorization of the stimuli. The number of responses in the table includes the ratings for both continua ($ii$ and $iy$).

<table>
<thead>
<tr>
<th>Age group</th>
<th>No. of listeners</th>
<th>No. of responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young (mean: 22.7, $SD$: 1.7)</td>
<td>59</td>
<td>16,472</td>
</tr>
<tr>
<td>Middle (mean: 30.2, $SD$: 3.4)</td>
<td>37</td>
<td>10,352</td>
</tr>
<tr>
<td>Old (mean: 50.7, $SD$: 5.02)</td>
<td>36</td>
<td>10,098</td>
</tr>
</tbody>
</table>

2.5 Analyses

In total, 36,960 responses (132 subjects $\times$ 14 steps $\times$ 10 blocks $\times$ 2 continua) and reaction times for a subset of 81 subjects were analyzed. All analyses were conducted for each continuum separately. First, a sigmoid function was fitted to the proportion of /ʃ/ ratings as a function of the acoustic step in the continuum for the different prime conditions. The sigmoid function is a special case of an S-shaped logistic function. It was used to visualize the distribution of the perceptual data and in particular the different cross-over points for the different groups. The sigmoid function (from which the 50% perceptual cross-over points averaged over all listeners was derived) and the statistical analysis were carried out with a generalized linear mixed model (GLMM) as implemented in the lme4 package.
(Pinheiro and Bates 2000) of the R software (version 2.14.1, R Development Core Team 2008). Categorical Response (/ç/ or /ʃ/) was the dependent variable, Continuum was the (numeric) fixed factor, and Listener was the random factor (see also Kleber et al. [2012] for a more detailed description of this method).

In addition to the identification results (responses), reaction times (RTs) were analyzed using separate linear mixed models for the two different continua. Here the dependent variable was the time it took listeners to make a decision (RT). The independent variables and fixed factors for both cases were Hood (KB, ZD, or CO), Step of the continuum, and Age of the listener; the random factors were Listener and Block. Other factors, such as musical taste, gender, favorite neighborhood, etc., were not included as factors because they did not provide significant explanatory power for the data. Note that in linear mixed models for categorical factors (such as Hood and Age), a particular level has to be chosen as the default level to which all other levels of the factor are being compared. However, this default level can be changed to get all other comparisons of interest.

3 Results

In this section, we will first look at the categorization of the stimuli as either *Fichte* or *fischte* depending on the three different priming conditions. In addition to the factor Hood, the potential influence of the Age of the listeners and the interaction of these two factors will be highlighted. The impact of these two factors will be analyzed for each continuum separately.

In the second part of the analysis, reaction times will be shown depending on the given Response (*fischte* vs. *Fichte*), on the Step of the continuum and on the Hood condition.

3.1 Responses

The psychometric curves in Figure 4 show the percentage of /ʃ/ ratings over the range from step 1 to step 14 for the *ii*-continuum and for the two priming conditions KB (marked by the darker circles and line) and ZD (marked by the lighter circles and line). As expected, the curves rise with increasing steps towards /ʃ/. The horizontal line marks the 50% rating boundary. The vertical lines (dark for KB and light for ZD) represent the turning point on the step scale (from 1 to 14) where the curves cross the 50% boundary, and thus, the categorical rating changes from more palatal fricative (/ç/) to more postalveolar fricative (/ʃ/) judgments.
Figure 4 shows that the crossover point in the Kreuzberg condition (KB) is to the left of the crossover point for the ratings in the Zehlendorf (ZD) condition. The crossover point was averaged over all subjects but for each continuum and each hood condition separately. This turning point or sensory threshold increases from the control condition CO (9.4) to KB (9.6) to ZD (9.9). For simplicity, the results for the control group are not included in the figure. The different turning points suggest that listeners rated the identical stimuli differently with respect to their priming condition. In other words, listeners changed their responses earlier to /ʃ/ judgments in the CO and KB conditions compared to ZD.

A generalized linear mixed model including Hood, Age, and Step as fixed effects and Listener and Block as random effects was run. A likelihood ratio test comparing the model with the three-way interaction term Hood*Age*Step to the model without the interaction term was conducted and revealed a better fit of the data for the more complex model with the interaction term included \((p < .001)\). Therefore, for the further analyses, the data were split by the three age groups. Figure 5 shows the percentage of fischte responses separated by Age and Hood. It is apparent that the age groups vary strongly depending on the influence of the different neighborhood names (Hood groups).

Separate GLMMs were run for each age group. For the middle age group, results were as expected: there was a significant difference in categorization ratings...
between KB and ZD ($p < .05$), while CO was between these two but did not differ significantly from either of them (cf. Table 2 for a summary of the fixed effects).

For the oldest age group, a likelihood ratio test revealed a better fit to the data of the model including the interaction term Hood*Step ($p < .001$). As Table 3...
shows, the effect of Step interacted with the comparisons of KB vs. CO and KB vs. ZD. Responses gathered from the CO group differed from those gathered in the KB condition \((p < .001)\) but not from those gathered in the ZD condition. In addition, there was a significant difference between the Hood groups (KB and ZD), as in the middle age group \((p < .001)\).

In contrast to the other age groups, for the youngest group, the priming conditions KB and ZD did not differ in their percentage of \textit{fichte/fischte} ratings. However, listeners in the CO condition rated significantly more stimuli as \textit{fischte} than in the KB condition \((p < .001)\), and also (although to a lesser degree) than in the ZD condition \((p = .05)\) (cf. Table 4).

Turning to the 
\(iy\)-continuum, the psychometric curves in Figure 6 show the proportion of /ʃ/ ratings for KB and ZD. As can be seen, the curves are steeper and the turning points for more /ʃ/ ratings are earlier than for the \(ii\)-continuum (Figure 4). This can be explained by the rounded vowel following the fricative, which serves as a secondary cue for the identification of /ʃ/. However, as in Figure 4, the crossover points for KB (dark vertical line) and ZD (light vertical line) differ, with KB turning first (KB = 7.3, ZD = 7.6). Here again, the crossover point for CO (not indicated in the figure) lies before KB (6.7).

Here too, the GLMM with the interaction of Hood, Age, and Step revealed a significantly better fit to the data than the model without the interaction \((p < .001)\). The corresponding models conducted for the different age groups revealed similar results as for the \(ii\)-continuum, but the factor Hood was not as strong: for the middle age group, the comparisons between ZD vs. CO and ZD vs. KB marginally failed to show significance \((p = .07\) and \(p = .08)\). For the older age group, KB vs. ZD differed significantly \((p < .05)\), while KB vs. CO did not reach significance \((p = .09)\). In the younger age group, no significant differences between Hood groups were found; only KB vs. CO nearly reached significance \((p = .054)\).

|                         | Estimate | Std. error | z-value | Pr(>|z|) |
|-------------------------|----------|------------|---------|----------|
| (Intercept)             | -4.15    | 0.257      | -16.163 | <.001*** |
| Hood (CO vs. KB)        | -1.45    | 0.379      | -3.843  | <.001*** |
| Hood (CO vs. ZD)        | -0.76    | 0.393      | -1.948  | .051     |
| Hood (KB vs. ZD)        | 0.69     | 0.408      | 1.693   | .091     |
| Step                    | 0.57     | 0.020      | 28.050  | <.001*** |
| Hood (KB vs. ZD)*Step   | -0.05    | 0.029      | -1.768  | .077     |
| Hood (CO vs. KB)*Step   | 0.11     | 0.026      | 4.295   | <.001*** |
| Hood (CO vs. ZD)*Step   | 0.06     | 0.026      | 2.314   | <.05*    |

Table 4: Summary statistics of the GLMM for the young age group \((N\) of observations: 8239, listeners: 59).
3.2 Reaction times

Reaction times (RTs) were gathered for a subset of 81 listeners including 24,374 observations. Stimuli with RTs below 350 ms (0.47% of the data) were excluded to guarantee that listeners had heard at least half of the stimulus before rating it. The mean RT over all data was 959.7 ms with a standard deviation of 109.3 ms. The two continua differed by a small but significant amount (paired $t$-test, $p < .001$, $t = 4.9$, df = 88). RT was higher, meaning that listeners were generally slower in the $ii$-continuum (966.5 ms) than in the $iy$-continuum (952.8 ms), suggesting that the additional auditory cue of the rounded vowel facilitates and speeds listeners’ decisions. Therefore, data were again split for the two continua for the further analyses.

We expected reaction times to be influenced by the continuum-step, with higher reaction times for the more ambiguous stimuli around the center of the continuum. Therefore, we divided the continuum into three subgroups (start, central, end) that reflect areas of relative stability (start and end) and areas of rapid change (central). For each continuum, RTs from steps around the 50% crossover point were pooled for the central group. Thus, the central group for the $ii$-continuum consisted of steps 7–11 (crossover point between 9.4 and 9.9 depending on the priming condition). The central group for the $iy$-continuum consisted

![Fig. 6: Sigmoid function fitted to the proportion of /ʃ/ ratings as a function of step for KB (darker circles) and ZD (lighter circles) for the continuum $iy$. The horizontal line marks the 50% rating boundary; the vertical lines mark the turning points for KB (dark) and ZD (light).](image-url)
of the steps 5–9 (crossover point between 6.7 and 7.6 depending on the priming condition). For the remaining steps, RTs were pooled into start and end, respectively, for each continuum.

Figure 7 shows the categorization of the stimuli as *fichte* or *fischte* dependent on the step group of the continuum (start, central, end) separately for the two continua. Remember that the stimuli from the start of the continuum are more /ç/-like (palatal), and the stimuli from the end of the continuum more /ʃ/-like (postalveolar). For the *iy*-continuum (left panel), as expected, listeners needed less time to categorize a stimulus from the beginning of the continuum as /ç/ and from the end of the continuum as /ʃ/, compared to a stimulus from the center around the category boundary. In contrast, for the stimuli from the *ii*-continuum (right panel), listeners only needed less time to categorize the more /ç/-like stimuli from the beginning of the continuum but not less time to categorize the more /ʃ/-like stimuli from the end of the continuum. Even though the stimuli from the end of the continuum comprise a more /ʃ/-like fricative and thus are less ambiguous, they were rated nearly as fast as either *fichte* or *fischte*. Here the missing cue of a rounded vowel seems to hamper reaction time for correctly rating the stimuli from the end (comprising /ʃ/-like fricatives) to be *fischte*.

In addition to the factor Step of continuum, we assumed that the different Hood conditions would have an effect on RTs, with longer RTs in the priming conditions ZD and KB than in the control condition. Figure 8 shows the relation between the Hood conditions and the RTs separated by the three step groups and the two continua. While, as expected, listeners in condition KB showed a slight tendency toward the longest RT, listeners in condition ZD responded faster than listeners in conditions KB and CO, regardless of the type of continuum (*ii* vs. *iy*) or the part of the continuum (start, central, end).

For the statistical analysis, the logarithmic values of the RTs were used, which was done to normalize the residuals (cf. Atkinson 1985; Pinheiro and Bates.
Separate GLMMs were run for the two continua. We ran models with the fixed factors Response (fichte or fischte), Step Group (start, central, end), Hood (CO, KB, ZD), and Age (young, middle, old), and the random factors Block and Listener. A likelihood ratio test did not reveal a better fit to the data with the factor Age included, but it did with the interaction term Response*Step Group*Hood. For the further analysis, we concentrated on the data for which the response was fischte. Again, likelihood ratio tests were conducted to find the model with the best fit to the data. For neither continuum did the interaction of Step Group*Hood or the inclusion of the factor Age improve the model.

Table 5 gives a summary of the fixed effects Step Group and Hood for the iy-continuum and for the ii-continuum. As expected from Figure 8, Step Group had

### Table 5: Summary statistics of the GLMM for the iy-continuum (N of observations: 6079, listeners: 81) and the ii-continuum (N of observations: 4278, listeners: 81)

<table>
<thead>
<tr>
<th>Continuum</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>pMCMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>iy</td>
<td>(Intercept)</td>
<td>6.886</td>
<td>0.015</td>
<td>433.1</td>
</tr>
<tr>
<td></td>
<td>Step (central-start)</td>
<td>−0.03</td>
<td>0.012</td>
<td>−2.3</td>
</tr>
<tr>
<td></td>
<td>Step (end-central)</td>
<td>−0.10</td>
<td>0.004</td>
<td>−22.9</td>
</tr>
<tr>
<td></td>
<td>Hood (CO-KB)</td>
<td>0.03</td>
<td>0.026</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>Hood (CO-ZD)</td>
<td>−0.06</td>
<td>0.028</td>
<td>−2.1</td>
</tr>
<tr>
<td></td>
<td>Hood (KB-ZD)</td>
<td>−0.09</td>
<td>0.033</td>
<td>−2.87</td>
</tr>
<tr>
<td>ii</td>
<td>(Intercept)</td>
<td>6.92</td>
<td>0.015</td>
<td>441.1</td>
</tr>
<tr>
<td></td>
<td>Step (central-start)</td>
<td>−0.01</td>
<td>0.009</td>
<td>−1.3</td>
</tr>
<tr>
<td></td>
<td>Step (end-central)</td>
<td>−0.05</td>
<td>0.005</td>
<td>−10.0</td>
</tr>
<tr>
<td></td>
<td>Hood (CO-KB)</td>
<td>0.02</td>
<td>0.027</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Hood (CO-ZD)</td>
<td>−0.06</td>
<td>0.029</td>
<td>−2.1</td>
</tr>
<tr>
<td></td>
<td>Hood (KB-ZD)</td>
<td>−0.08</td>
<td>0.023</td>
<td>−3.6</td>
</tr>
</tbody>
</table>
a significant influence on RT. RTs for rating a stimulus from the end of the continuum as *fischte* were faster than for the central part, and this effect was twice the size for the *iy*-continuum than for the *ii*-continuum (i.e., −0.10 vs. −0.05). Also, for the *iy*-continuum, the start and central step groups differed in their RTs for the *fischte* responses, while this was not the case for the *ii*-continuum. This can be explained by the additional cue of the rounding vowel in the *iy*-continuum facilitating listeners’ *fischte* responses. As is already apparent in Figure 8, the factor Hood also had an effect: RTs were faster in condition ZD than in conditions CO and KB, while KB and CO did not differ significantly in either of the continua.

4 Discussion and conclusion

Results show that listeners categorize identical acoustic stimuli on a continuum differently with the co-presentation of the names of two different Berlin neighborhoods. The significantly different responses in the Kreuzberg versus Zehlendorf conditions show a perceptual divergence effect in the classification and categorization of the stimuli. There is also evidence of age-graded differential perception across the three age groups tested with the perceptual divergence paradigm.

The middle age group behaved just as we had hypothesized. There were significantly more /ʃ/ ratings in the Kreuzberg condition than in the Zehlendorf condition, with values for the control condition somewhere in between (not varying significantly from KB or ZD). This group does not seem to associate /ʃ/ with the middle-class neighborhood Zehlendorf, but rather the regional standard variant /ç/. At the same time, these listeners in their thirties responded most often with /ʃ/ in the Kreuzberg condition, showing that this label triggers different expectations, which are then reflected in their categorizations. Whether these expectations are learned from their own experiences and observations is unclear; the data suggest, though, that this group of listeners as a whole seems to associate /ʃ/ with the concept of Kreuzberg.

The oldest group of listeners (in their fifties) also rated significantly more /ʃ/ in the Kreuzberg condition than in the Zehlendorf condition, indicating that they also interpreted the labels linguistically, meaning that they projected certain speech styles to specific neighborhoods. However, in contrast to the middle age group, they also responded significantly more often with /ʃ/ for the Kreuzberg condition than for the control condition, showing that here too their expectation was guiding their categorization. Ratings in the control condition and in the Zehlendorf condition did not differ significantly, possibly indicating an unmarked
'standard’ – with the main trigger for association being Kreuzberg. Therefore, it seems that older listeners still make a strong distinction between the categories Kreuzberg and Control, while these two categories seem to merge for the young and middle age groups. One reason primes are interpreted differently by older than by younger listeners may be a lack of contact between these two groups of listeners. Our attempts to find an effect of any of the meta-data factors that we gathered (for example, listeners’ interaction with children or grandchildren) did not generate any significant results.

Probably the most interesting results were obtained for the youngest group of listeners. They did not behave as expected: the label Kreuzberg did not trigger more ratings of the postalveolar /ʃ/ than the label Zehlendorf. In fact we found fewer ratings of /ʃ/ in the Kreuzberg condition than in the control condition. This evidence suggests that young listeners generally (expect and) hear more /ʃ/ fricatives independently of the context. That is, young listeners labeled the stimuli in the control condition with more /ʃ/ than the stimuli in KB and ZD. This indicates that these forms have already become their new standard. Interestingly, the prime Kreuzberg only had the effect of reducing the number of categorized /ʃ/. This may be a counter reaction to the common stereotype that speakers in Kreuzberg often produce /ç/ more like [ɕ].

We believe that the age-graded data that we gathered with the perceptual divergence paradigm shows that there is a sound change in progress, with the youngest listeners and speakers leading the change and the older ones either following (i.e., the middle age group) or resisting. The resistance may be due to lack of contact with the primary speakers of this style who propagate the use of [c] or [ʃ], or a negative attitude towards Hood German in general. It seems that, particularly for the listeners in the youngest and middle age groups, the [c] or [ʃ] variant has become an accepted form. This may be due to their respective environments where this speech variant is not socially marked anymore. This is particularly evidenced by the fact that young listeners rated more /ʃ/ in the control condition than in the other two conditions. In the spreading of this /ç/-/ʃ/ alternation for the youngest listeners, the feature is becoming less and less associated with a particular group, and the use of this feature by others is no longer a transgression and socially marked. In Auer’s (2003) sense, this feature has been de-ethnicized. As our ongoing work shows, the fronting of the fricative fits well with the observation that the realization of the diphthong /øy/ is also fronted by the Hood German speaker group (Jannedy and Weirich 2013); there seems to be a tendency to front or centralize articulation.

Although we did consider the effects of different kinds of meta-data on our listeners, we were not able to find a good predictor for our results other than age. The age effect turned out to be strongest in the control condition. We believe that
this might point to instability of the palatal fricative and to sound change in progress. The results are in line with the rarity of /ç/ in the world’s languages and the merger of /ç/ and /ʃ/ to /ɕ/ in the central German dialects.

Thus, there appear to be three driving forces of this sound change in Berlin. First, geographically, Berlin is close to the middle German dialect belt in which [ç] is being replaced by [ɕ] or even [ʃ]. It is feasible that due to geo-political obstacles (the Wall), the spread of this change was halted for several years and has only within the past 20 years or so started to progress again. Secondly, some speakers across the different age groups who live and were born in Berlin produce [ɕ], and it is thus not completely unobserved in the Berlin dialect, even though there is only sparse evidence of this in the literature from different times and speaker groups (Lasch 1928; Schönfeld 1986; Auer 2003). The third contributing factor may be the youth-style Hood German that can be frequently heard in public, such as on public transportation, in shopping malls, or on the streets. It has become a widely distributed youth speech style throughout Berlin, although it was formerly strongly associated with Kreuzberg. These three factors might provide the basis for this synchronic change.

Throughout this paper, we have used category descriptions such as /ç c f/. By no means, though, do we mean to imply that these are stable categories rather than fricatives that vary on some acoustic and articulatory dimension. This is particularly true for /ç/, which is typologically relatively rare in the world’s languages. Moreover, there is a strong distributional constraint on its occurrence in German (Hall 1992; Labov 1994: 350). In Pierrehumbert’s (2001) exemplar-based model of mergers, she shows that marked and less frequent categories such as /ç/ are subject to a persistent bias towards a more frequent and unmarked category such as /ʃ/. Moreover, Labov (1994) explains that mergers tend to occur toward the more common of the two phonemes, but that the result may also be intermediate between them. This explains why the resulting sound may resemble /ɕ/ more than /ʃ/, even though /ɕ/ is also relatively rare cross-linguistically. Wedel (2012) and Wedel et al. (2013) provide an additional language-internal reason for the loss of this contrast, which is based on the idea that the relatively small number of minimal pairs in German contrasting /ç/ and /ʃ/ facilitates the merger between these two sounds.

The quality of the preceding vowel is another factor that influences fricative classification. The rounded vowel quality characteristically occurring in a /ʃ/ context produced within the Berlin dialect region may be the cause of the faster RTs for the iy-continuum. For the ii-continuum, the extra cue of vowel rounding is missing and makes the categorization of /ʃ/ more difficult, which in turn results in longer RTs. Furthermore, the factor Hood has an effect on the RTs in both continua and is independent of the age of the listener: participants are faster in
condition ZD than in the conditions CO and KB, while KB and CO do not differ significantly. Further studies will have to provide more evidence, but our current results may indicate that positive associations lead to faster categorization responses than lack of association or negative associations. It remains to be tested whether the association with negative primes implies a larger cognitive load and therefore requires more time than the co-processing of positive information. This stipulation seems to hold true for research on positive versus negative trait processing in social psychology (Dovidio et al. 1986; Pratto and John 1991; Baumeister et al. 2001). For example, Pratto and John (1991) showed that the processing of color terms was faster when positive primes (such as honest vs. sadistic) were printed on the colored cards. In the priming experiments of Dovidio et al. (1986) on the relationship between racial categories (black vs. white) and positive and negative traits, they found that reaction times were generally faster for positive traits than for negative traits. Furthermore, subjects showed faster responses when positive personality traits such as ambitious, smart, and clean were paired with white as compared to when undesirable personality traits such as stupid, lazy, or welfare were paired with black.

Many recent experimental results on the categorization of speech in different languages and cultures suggest that listeners do not categorize speech in a context-free way but by attributing speaker characteristics, which may be manipulated through explicit and even very subtle implicit information relating to the origin of the stimulus (Strand and Johnson 1996; Niedzielski 1999; Hay et al. 2006; Hay and Drager 2010; Brunelle and Jannedy 2007, 2013). It is therefore possible that the young voice from which we generated the experimental stimuli may have had an effect on the outcome of our categorization data – that listeners overall favored a /ʃ/-type response. Speech categorization is highly influenced by non-speech observations; it is automatic and unmediated. Categorization shows great sensitivity to existing social stereotypes and social groups and seems to be at least one of the driving forces behind language variation and sound change.

In summary, we did find variation in the classification of identical acoustic stimuli. However, this variation was not random but appeared to be systematically affected by the listeners’ age and the prime condition (i.e., the co-presented information). This is a strong indication that in the processing of speech sounds, not only the sound itself but also a strong interpretive component is at play. This needs to be explored in much greater detail. More importantly, we need to explore what types of triggers can drive the interpretive component of the perceptual system. We know that the dialect of the experimenter can have an effect on the differential perception of stimuli: this is information interpreted through the auditory system. We also know that visually perceived objects denoting national
identity can lead to differential perception. And of course, visually perceived language – written words – that evokes concepts can result in differential perception. If expectations resulting from previous experiences and stereotypes calibrate a perceptual system, then any perceptual experience related to a linguistic experience, from sight, sound, or even smell could lead to associations that affect the categorization of speech sounds.

Acknowledgments: This work was supported by the German Federal Ministry for Education and Research (BMBF) – Grant No. 01UG0711. We thank Jana Brunner for creating the continua for this study. We are very grateful for all the help we have received from our experiment participants and for the time that the young men and women from Kreuzberg, Wedding, and Neukölln spent with us. Our Research Assistant, Sophie Arndt, helped process this data. We are also grateful to colleagues at various conferences who advised and taught us and gave us helpful feedback. We also thankfully acknowledge the great contribution of two anonymous reviewers and the editorial team for asking just the right questions and for giving us very valuable feedback on our work. All shortcomings remain our own.

References

Androutsopoulos, Jannis. 2001. From the streets to the screens and back again. On the mediated diffusion of ethnolectal patterns in contemporary German. LAUD Linguistic Agency A522.


Wiese, Heike. 2011. so as a focus marker in German. Linguistics 49(5). 991–1039.
