Kleber et al. (2010), Journal of Phonetics, 38, 185 - 196.

Title: The implications for speech perception of incomplete neutralization of final devoicing in German.

Felicitas Kleber[a,*], Tina John[a], Jonathan Harrington[a]

*Corresponding author: kleber@phonetik.uni-muenchen.de, Tel: +49 89 2180 2454, Fax: +49 89 218 5790

[a]Institute of Phonetics and Speech Processing (IPS), Ludwig-Maximilian-Universität, Schellingstr. 3/II, 80799 Munich, Germany

Abstract

We investigated the perceptibility of stop voicing in a domain-final neutralizing context in German that according to various phonological models is completely neutralized in favour of the voiceless category but that according to various empirical studies is distinguishable phonetically. A primary aim was to determine whether acoustic cues that were available for the stop voicing distinction were perceptible in a neutralizing context. A secondary aim was to assess whether voicing perception was influenced by phonotactic frequency and the potential for resyllabification. Nineteen listeners of a Standard German speaking variety made forced-choice judgments to synthetic stimuli spanning a voiced–voiceless continuum containing domain-final alveolar and velar stops in various neutralizing contexts that differed in terms of phonotactic probability and the potential for resyllabification. Our results showed that voicing information could be distinguished but that the perceptibility of this distinction also depended on statistical properties of phoneme sequences and whether a domain-final stop could potentially be perceptually resyllabified as domain-initial. Our general conclusion is that a categorical neutralization model is insufficient to account for stop voicing perception in German in a domain-final context: instead, voicing perceptibility in these contexts depends on an interaction between acoustic information and phonological knowledge which emerges as a generalization across the lexicon.

Keywords

Incomplete neutralization, fine phonetic detail, perception, phonological frequency

1. Introduction

Recent studies have provided increasing evidence that fine phonetic detail, at segmental and prosodic levels is an integral part of speech communication in both production and perception (see e.g. for a detailed description by Hawkins, 2003). For example, vowels in frequently occurring words with fewer lexical neighbours tend to be produced with a more centralised quality as opposed to rare words with more lexical neighbours (Wright, 2003) and onset /l/ is longer before voiced vs. voiceless coda stops in /li:d/ vs. /li:t/ (Hawkins & Nguyen, 2004). These subtle phonetic differences are perceptible and used for word recognition (Hawkins & Nguyen, 2003; Davis et al., 2002; Manuel, 1995). The type of fine phonetic detail which is the main concern of the present study involves incomplete neutralization of the voicing contrast in oral stops which has been demonstrated experimentally in Dutch (Warner et al., 2004), German (Port & O'Dell, 1985), and Catalan (Charles-Luce & Dinnsen, 1987). In a number of languages such as Polish, Dutch, Catalan and German, obstruents contrast in voicing when they occur in a prosodically domain-initial or medial position[1], whereas in domain-final position this contrast is neutralized with a bias in favour of the voiceless component. The classic example of neutralization in German arises from word-final devoicing which causes *Rad* ('wheel') and *Rat* ('advice'), which differ underlyingly in the voicing status of the final consonant, to become surface homophonous. The process of final devoicing that results in this type of neutralization is often generalized phonologically by a rule such as:

[+obstruent] → [-voice]/_D (where D stands for domain-boundary)

However, implicit in this type of rule is that neutralization is complete, i.e. that neutralized and underlyingly voiceless forms are indistinguishable from each other.

The issue as to whether German final devoicing really is incomplete remains controversial and it has received considerable attention in the last 20–30 years. For German, acoustic analyses have revealed small but significant longer vowel durations preceding word-final underlying voiced obstruents than their voiceless counterparts (Port, Mitleb and O'Dell, 1981; O'Dell and Port, 1983, Port and O'Dell, 1985). That is, in these studies derived voiceless obstruents were shown to be acoustically intermediate between

[1] This phonological contrast in German is also described in terms of lenis versus fortis. For a better comparability with other languages we use throughout the paper the terms voiced and voiceless. See Kohler (1984) for a discussion of the lenis/fortis dichotomy.

voiced and voiceless obstruents in domain-initial or medial position – though they were far closer to voiceless than voiced stops. Other languages for which there is evidence in favour of incomplete neutralization include Catalan (Dinnsen & Charles-Luce; 1984), Polish (Slowiaczek & Dinnsen; 1985), and Dutch (Warner et al., 2004; Ernestus & Baayen, 2006), the latter being a final-devoicing language for which previous studies had shown complete neutralization (Jongman et al., 1992; Baumann 1995). The durational differences between stop categories in a neutralizing context varied depending on the language, but were always pervasive. Among all the acoustic correlates measured in these studies (e.g. voicing into and during closure), vowel duration stood out as the most important cue for preserving the voicing contrast to some extent, even in Polish (Slowiaczek & Dinnsen; 1985) where vowel duration is not the primary cue for differentiating voiced and voiceless obstruents word-medially (Keating 1979). Most of these have also shown individual speaker and dialect-dependent significances (see especially Piroth & Jancker, 2004 for German) while in others the extent to which the voicing contrast was neutralized was shown to be dependent on sentence position, semantic information and phonetic environment (Charles-Luce 1985, 1993; Slowiaczek & Dinnsen, 1985), pragmatics (Port & Crawford, 1989) as well as morphology (Ernestus & Baayen, 2006). Various studies showed that orthography and speaking style affect the degree of incomplete neutralization (sometimes resulting in hyperarticulation) and some authors concluded that incomplete neutralization was an experimental artefact of orthography (e.g., Fourakis & Iverson, 1984; Jassem & Richter, 1989; Warner et al., 2006; but see Port & Crawford, 1989 and Ernestus & Baayen, 2006 for a different interpretation).

One way of resolving these contradictory arguments is to test whether acoustic incomplete neutralizations are perceptible and if so to assess whether listeners make use of the different cues for distinguishing between words that differ underlyingly minimally in stop voicing. Some of these issues have already been addressed by Port and colleagues (Port & O'Dell, 1985; Port and Crawford, 1989) for German and by Slowiaczek & Szymanska (1989) for Polish. In these three studies, it was found that the words of minimal pairs differing in underlying voicing were discriminated better than chance in neutralizing contexts and it was concluded that listeners can make use of the acoustic cues in order to identify minimal word pairs in a forced-choice task but argued against the idea that this discriminability had a functional role. Since, however, these experiments were based on natural speech materials that contained various acoustic confounds, no clear conclusion can be drawn as to which acoustic cues listeners actually exploit for word identification and discrimination in these neutralizing contexts. In Slowiaczek & Szymanska's (1989), the same Polish material was also presented to English listeners whose performance was similar to that of the Polish listeners; they also found that both groups of listeners showed a perceptual bias towards the voiceless variants. According to Slowiaczek & Szymanska (1989), the combination of poorer-than-expected (though better-than-chance) identification and the perceptual bias in favour of the voiceless obstruent suggest that the acoustic cues are not reliably used in regular communication to differentiate between members of a minimal pair (and are hence not primary cues) and are "perceptually neutralized" (p. 211). Moreover, since the data from speaker groups of both languages were quite similar, it was concluded that the performance must have been a function of the acoustic information in the speech signal rather than being mediated by native listeners' knowledge of Polish phonological rules. By contrast, Broersma (2005) found that language background did influence listeners' responses. When a choice of cues was available, listeners preferred to use familiar or primary cues (e.g. cues they know from their language or from other syllable positions), but in the absence of such a choice listeners were able to exploit less familiar cues to improve their disambiguation performance (Warner et al., 2004; Broersma, 2005). Listeners are very well able to discriminate voicing even in positions where there usually is no contrast, as long as this contrast exists in their phonological system (cf. also the Perceptual Assimilation Model, Best, 1994).

One of the factors that has not been considered so far that may influence incomplete neutralization is phonological frequency. Indeed, since listeners' perception of fine phonetic detail to disambiguate domain-final obstruent voicing has been shown to be language-dependent (Broersma, 2005), it seems quite possible that the extent of these subtle acoustic differences may also be conditioned by language-dependent factors such as phonological frequency and effects of syllable position. As far as phonological frequency effects are concerned, various studies have shown how perception is influenced by phonotactic co-occurrences, word frequency, and neighbourhood density. For example, Hay, Pierrehumbert, and Beckman (2003) found that listeners often tend to misperceive statistically infrequent (e.g., /np/) as statistically frequent (e.g., /mp/) nasal-obstruent clusters. Nonsense words containing phoneme sequences that occur frequently in real words are better memorized (Frisch et al., 2000) and repeated faster (Vitevitch & Luce, 1999; Vitevitch et al., 1997) as opposed to non-words containing rare sequences. Compatibly, Pitt & McQueen (1998**)**, found evidence for a perceptual bias in listeners' responses to acoustically ambiguous consonants depending on the probability with which the consonant occurs in a consonant sequence. Following exemplar theory, all these examples of 'misperception' are very likely to come about because of a perceptual adjustment depending on lexically based expectations of the listeners.

Such frequency related predictions may also carry over to final devoicing in German in which the degree of neutralization could depend on properties of statistical frequency in the lexicon and more specifically on phonotactic frequency. In German, there is a phonological distinction between tense and lax vowels (e.g. /biːtən/ 'to offer' vs. /bɪtən/ 'to beg') and intervocalically also between voiced and voiceless obstruents (e.g. /miːdən/ 'to avoid' vs. /miːtən/ 'to rent'). However, the combination of vowel tensity and obstruent voicing is anything but equally distributed in the German lexicon. For example, lax vowels almost always precede voiceless stops. There are only a handful of words containing a lax vowel plus voiced stop sequence and most of these are loan words (e.g. /ɛbə/ 'tide', /klɛvə/ 'clever', /meːʃʊgə/ 'crazy') which, according to Féry (2003), are not part of the "truly core native German vocabulary" (p. 150). At the same time, underlyingly voiced labial or velar obstruents are almost always preceded by tense vowels. Thus surface /iːp/ is almost always a reflex of underlying /iːb/ (/liːbə/ 'love', /ʃiːbən/ 'to push') and, with the exception of one or two infrequent words (e.g., /piːksən/, 'to prick'), surface /iːk/ indexes underlying /iːg/ (/fliːgə/ 'fly', /kriːgən/ 'to get', /liːgən/ 'to lie'). The same predictable relationships hold for other tense vowels. Only alveolar stops that are preceded by tense vowels show a balanced distribution of both underlying voicing categories (e.g. /miːdən/ 'to avoid', /miːtən/ 'to rent', /baːdən/ 'to take a bath', /baːtən/ 'they asked for', /boːdən/ 'floor', /boːtən/ 'carriers').

Previous studies have controlled for *phonetic* vowel length in combination with the following obstruents (Broersma, 2005; Warner et al., 2004) but only as part of the investigation of vowel duration as a cue for voicing perception. By contrast, our aim in the present study is to test whether these types of imbalances in combinations of *phonological* vowel length and voicing influence perception. Thus we would predict that listeners are better able to perceive a domain-final /t, d/ contrast after a tense vowel than after a lax vowel, because the voicing contrast is only frequent in the lexicon for tense vowel plus alveolar stop sequences (e.g. /laɪdən/ 'to suffer' vs. /laɪtən/ 'to lead'). When, on the other hand, the contrast is infrequent, then voicing judgements should be guided by lexical statistics: thus alveolars are more likely to be perceived as voiceless following lax vowels since combinations of lax vowels and voiced stops are almost non-occurring in the lexicon.

The second conditioning factor that we will consider here is the *degree* of neutralization which may vary depending on syllable position and the extent to which the sequence containing a syllable final but pre-consonantal obstruent is resyllabifiable. The starting point for this research question comes from two phonological approaches to the analysis of final devoicing: the *licensing-by-prosody* account (Itô 1986, 1989; Goldsmith, 1990; Rubach 1990; Lombardi, 1999; Beckman, 1997), in which final prosodic position is the main determiner of neutralization (see also e.g. Brockhaus, 1995; Hall, 1992; Vennemann, 1972; Wiese, 1996) and the *licensing-by-cue* approach (Steriade, 1997, 1999, 2000) in which acoustic properties of the potentially neutralizing context are primary. Based on the analysis of neutralization in different languages, Steriade predicts that phonological contrasts are neutralized first in environments in which the perceptual differentiation can only be maintained by additional articulatory effort. Irrespective of the differences between these models, an important question that is relevant for both is whether acoustic cues to a phonological contrast are perceptually *masked* depending on either the phonetic or prosodic context. This issue was touched upon by Cutler (2002) who reasoned that the phonological generalization that vowel length is neutralized utterance-finally in Japanese may in itself contribute to listeners' inability to hear the contrast, even when there is acoustic evidence for its distinction (Kubozono, 2002).

As far as the present study is concerned, the basis for this kind of perceptual masking is as follows. It is well known that the perceived differences between allophones of a phoneme are much less than those between different phonemes and this is consistent with findings from the child language acquisition literature (Werker, 1995; Werker & Tees, 1989) showing that the perceptual discriminability between allophones of the same phoneme diminishes in the first year of acquisition, presumably because children learn to focus on acoustic cues that are important for distinguishing between phonemes and words and also to ignore (or pay less perceptual attention to) those that are not. It is possible that this kind of perceptual masking operates not only allophonically but also between phonemes that are in a neutralizing context. Thus, perhaps listeners filter out perceptually any acoustic cues that might be present to the voicing contrast in domain-final position, because they interpret this as a neutralizing context in which there is usually no surface distinction according to their phonological knowledge (however, see Ernestus & Baayen, 2006 for some results in Dutch that are not consistent with this position). If the phonological grammar does exert a top-down influence on the acoustic signal in this way, then we might expect acoustic cues to be more perceptible in a context in which a domain-*final* stop has the potential to be interpreted as domain-*initial* (as a result of which it is no longer in a neutralizing context). More specifically, consider that in a /vowel-stop-l/ context, the stop is necessarily domain-final if it is alveolar because in German, as in English, initial */dl/ and */tl/ are excluded. Therefore, any potential cues to the voicing contrast might be perceptually strongly masked precisely because the alveolar stop is necessarily domain-final and therefore in a neutralizing context. But this might not be so if the stop is a velar because /kl, gl/ are legal onset clusters in German (and English). Thus, because velar stops could be interpreted by the listener to be domain-initial, the perceptual masking which is predicted to filter out acoustic information in the domain-final neutralizing context would not apply in this case. Consequently, the probability of hearing any distinguishing voicing cues should be greater for the velar than the alveolar context according to this perceptual masking hypothesis.

In this paper, we describe two experiments designed to test whether the domain-final voicing contrast in German obstruents is incompletely neutralized in perception and whether the degree of the incompleteness depends on phonotactics and statistical co-occurrences of phonemes in the lexicon. The first hypothesis was the starting point for all other hypotheses and therefore tested in both experiments; it can be summarised as follows.

(H1) The voicing contrast is incompletely neutralized in the perception of German domain-final obstruents.

Hypotheses H2 – H4 that gave rise to the experiments were all motivated by the relative frequency with which various patterns of segment sequences occur in the lexicon. Experiment 1 was conducted to investigate syllable internal sequences. In this experiment, hypotheses H2 and H3 were tested; they can be summarised as follows:

(H2) Listeners show a perceptual bias towards the more frequently occurring voiceless stop when preceded by lax vowels as opposed to tense vowels (e.g. /vɪd/ is predicted to be perceived as /vɪt/ because /ɪt/ co-occurs frequently and /ɪd/ does not).

(H3) The voicing distinction is perceptually less neutralized when there are analogous frequent contrasts in the lexicon. Thus since /iːd/ vs. /iːt/ is lexically frequent but /iːg/ vs. /iːk/ is not, then listeners should be better able to distinguish /d, t/ in the former sequence in a neutralizing context (such as in a domain-final position, e.g. before an obstruent in which the stop is necessarily domain-final).

The central research question of Experiment 2 was whether resyllabification affects the degree of incomplete neutralization: this is formulated in hypothesis H4.

(H4) Perception of the voicing contrast is more likely in a consonant cluster that can be resyllabified with the onset consonant of the following syllable (e.g. there is more neutralization of the voicing contrast in a /stop#l/ sequence for alveolars than for velars).

2. Experiment 1: Effect of probabilistic co-occurrences of phoneme sequences on incomplete neutralization

2.1. *Speech Materials*

We created four continua one each consisting of resynthesized stimuli of four minimal pair disyllabic compounds: *Widdlinn–Wittlinn* (henceforth /$V_{lax}C_{alv}$/), *Bigglinn–Bicklinn* (henceforth /$V_{lax}C_{vel}$/), *Niedlinn–Nietlinn* (henceforth /$V_{tns}C_{alv}$/), and *Mieglinn–Mieklinn* (henceforth /$V_{tns}C_{vel}$/). These compounds were hypothetical German town names. The reason for choosing different onset consonants for the compounds was as a reminder to the listener that s/he would be perceiving a tense vowel for example in 'nie[d/t]' but a lax vowel in 'wi[d/t]'. If we had used the same onset consonant for both, then listeners might have confused the tense and lax continua given that decreasing vowel duration, which is one of the variables manipulated here, is also a positive cue for a lax as opposed to a tense vowel. We have no reason to expect the different [v] vs. [n] onset consonants to affect voicing judgments. Analogously we chose different onset consonants for the minimal pairs differing in place of articulation of the syllable-final stop.

In order to create the continua in the compounds, a male speaker produced (together with the test words from Experiment 2 reported below) the trochaic words *Widden* (/vɪdən/), *Witten* (/vɪtən/), *Biggen* (/bɪgən/), *Bicken* (/bɪkən/), *Nieden* (/niːdən/), *Nieten* (/niːtən/), *Miegen* (/miːkən/), and *Mieken* (/miːkən/). Each of these words was repeated in isolation 10 times. We then chose one token of each of the 10 repetitions of the four voiced tokens and spliced out the first syllable at the stop release thus leaving /vɪd-/, /bɪg-/, /niːd-/, and /miːg-/. Our choice of this syllable was based on two criteria: firstly, the second syllable in the original trochaic word had to be produced as a syllabic /n/ with an elided schwa; and secondly, the durations of the /iː/, /ɪ/, /d/, and /g/ should be closest to the mean duration of these segments across all ten tokens. The second syllable /lɪn/ was spliced out of either a production of *Britlinn* or a production of *Ricklinn* (again both context words were German pseudo town names). We cut out these second syllables at the onset of periodicity of /l/ and, depending on the syllable-final stop's place of articulation, appended the -*linn* taken from *Britlinn* to /vɪd-/ and /niːd-/ and the -*linn* from *Ricklinn* to /bɪg-/ and /miːg-/ to create the spliced compounds with the voiced stop /vɪdlɪn/, /bɪglɪn/, /niːdlɪn/, and /miːglɪn/. The syllable final stops in the target words and the context word matched in place of articulation in order to avoid a disruption of the acoustic cues at the splice point (in front of the lateral). These spliced blends served as the endpoint stimuli at the voiced ends of the continua.

The tokens towards the voiceless end of the continuum, i.e., towards /vɪtlɪn/, /bɪklɪn/, /niːtlɪn/, and /miːklɪn/ were derived from these voiced endpoints by reducing the V/VC duration ratio (ratio of vowel duration to vowel plus closure duration, henceforth V/VC), where V = /iː/ or /ɪ/ and C = the following alveolar or velar stop closure. V/VC has been found to be the most powerful acoustic cue for disambiguating voiced from voiceless stops in a semi-intervocalic context (Kohler 1979). Before we applied this shortening, we had to determine V/VC at the voiceless *endpoint* (i.e. at the most extreme /vɪtlɪn/, /bɪklɪn/, /niːtlɪn/, and /miːklɪn/ tokens). This was calculated by weighting the V/VC averaged across all voiceless tokens by the VC duration in the selected voiced context. More specifically where $Rhyme_{voiced}$ is the duration of the voiced rhyme (e.g. duration of the selected /ɪd/, /ɪg/, /iːd/ or /iːg/ token) and $Ratio_{voiceless.m}$ is the mean V/VC duration ratio in voiceless consonants (mean duration of V/VC in /ɪt/, /ɪk/, /iːt/ or /iːk/), then the duration of the vowel preceding the voiceless stop, $Vowel_{voiceless}$, that was used for the voiceless endpoint in the synthesis continuum was calculated from:

(1) $$Vowel_{voiceless} = Rhyme_{voiced}\ Ratio_{voiceless.m}$$

The calculation by means of Eq. (1) ensured that the total duration of each stimulus item within a continuum remained constant.

We then derived four seven-step continua between these voiced and voiceless endpoints. To calculate the step size for the stimuli of each continuum, we divided the vowel duration difference between these endpoint stimuli by six. The step size was 7 ms for both the /$V_{lax}C_{alv}$/- and the /$V_{lax}C_{vel}$/- continuum, 14 ms for the /$V_{tns}C_{alv}$/-continuum, and 20 ms for the /$V_{tns}C_{vel}$/-continuum. In order to compare, analyze and evaluate the perception results for the various continua that differ both in phonological vowel quantity as well as in their segmental structure, it was necessary to use proportionally equal distances instead of absolute step sizes. The vowel durations of the selected voiced tokens were then progressively shortened and the stop closure durations were progressively lengthened by the calculated step sizes, so that the VC duration remained constant (see Table I for further details).

Insert Table 1 about here

The f0-contour was stylized such that there was a rise towards the midpoint of the accented syllable and a fall from there linearly over the rest of the test blend with all stimuli having the same pitch heights for the five f0 target points (Fig. 1). For all stimuli, any evidence of voicing during the closure was also removed by high-pass filtering. Listening tests with 10 subjects showed that the endpoints could be unambiguously distinguished in all continua.

Insert Figure 1 about here

All manipulations were done by means of the manipulation function and then resynthesized with the "overlap and add" function in Praat (version 5.0.27, Boersma & Weenink, 2008).

2.2. *Participants*

The 168 stimuli (6 repetitions x 4 continua x 7 steps) were made available in an online forced-choice identification experiment (together with the stimuli of Experiment 2 reported below as well as stimuli from another experiment which we will not report on here). Nineteen native speakers of Northern Standard German, all of them students at Kiel University, participated in the online experiment at the Institute of Phonetics and Digital Speech Processing, University of Kiel. Participants were paid a small amount for participation. None of the subjects reported any hearing, eye-sight, or reading problems.

2.3. *Experimental Procedures*

The subjects performed a two-alternative forced-choice task. All stimuli (including all other stimuli, which served as distracters in this experiment) were presented to the listeners over headphones in one session. Upon presentation of an auditory stimulus, the subject saw an orthographic representation corresponding to the minimal pair distinction. For example, upon being presented auditorily with one of the tokens from the /niː[d/t]lɪn/ continuum, the subject saw *Niedlinn* or *Nietlinn* on the screen and had to judge which of these was more similar to the perceived stimulus. The experiment was self-paced, i.e. the next item was only presented after the participant had made a decision. On average, the entire experiment took about one hour. The order of the stimuli was random for each participant to avoid any presentation effects. The responses were saved to a server located at the Phonetics Institute in Munich.

Two different types of statistical analyses were carried out in the programming language/environment R. We first analyzed our data by means of logistic regressions, i.e. we calculated the log of the ratio of the voiceless responses to the corresponding voiced responses for each stimulus in each continuum. This logit was the dependent variable and V/VC Ratio (stimulus 1 to 7, in which V/VC ranged from maximal to minimal) was the independent variable. Second, repeated measures generalized linear mixed models (GLMM – see Baayen, 2008) were fitted to predict incomplete neutralization as a function of the decreasing V/VC duration ratio and to determine whether the degree of incompleteness depended on the lexical frequency of the voicing contrast which was in this case dependent upon vowel tensity and place of articulation. The 'voiceless' responses served as our dependent variable. As predictors (independent variables) we entered Tensity (lax vs. tense), Place (alveolar vs. velar) and V/VC Ratio (stimuli 1–7). Subject was entered as a random effect factor.

2.4. Expectations and Predictions

To answer the question whether the post-vocalic domain-final voicing contrast is completely or only incompletely neutralized, we shall first make some assumption regarding different forms of psychometric curves representing the listener's judgments.

Fig. 2 is a schematic outline of five possibilities: (1) no neutralization of the contrast, showing a steep rise between the endpoints which signals the presence of two categories, (2) incomplete neutralization of the contrast with no bias towards either category, showing a slightly rising slope around the 50% cross-over point, (3) complete neutralization with no bias towards either category, (4) complete neutralization of the contrast with a bias towards the voiceless category in which responses are above the 50% cross-over point, (5) as (4), but in which the bias is towards the voiced category.

Insert Figure 2 about here

2.5. Results and Discussion

Fig. 3 (a) gives the proportion of voiceless responses as a function of decreasing V/VC duration ratio for the four continua. In order to test whether there were discernible trends in the proportional responses along the continua, logistic regression lines were calculated (Fig. 3 (b)). All continua showed slightly, but gradually increasing identification functions, which indicated that the voicing contrast remained to some extent perceptible though there were no abrupt changes which would point to the perception of two distinct categories. The increases in the voiceless responses along the continua from left to right followed significant trends in all continua: /$V_{lax}C_{alv}$/ ($\chi^2(1) = 7.2$, $p < 0.01$), /$V_{lax}C_{vel}$/ ($\chi^2(1) = 42.1$, $p < 0.001$), /$V_{tns}C_{alv}$/ ($\chi^2(1) = 28.2$, $p < 0.001$), and /$V_{tns}C_{vel}$/ ($\chi^2(1) = 61.8$, $p < 0.001$). The responses to most stimuli of all continua were in the voiced range, i.e. the percentage was below the 50% boundary. As Figs. 3 (a) and (b) show, listeners labelled more stimuli as voiceless when the vowel was lax (36.6%) than tense (10.9%), which suggests that the lexical frequency of the tensity plus stop voicing combination influenced voicing judgments: this is consistent with H2. Listeners also labelled more stimuli as voiceless when the place of articulation was velar as opposed to alveolar, which is not compatible with H3.

Insert Figures 3 (a, b) about here

The results of the GLMM in which the voiceless responses to all four continua were included as the dependent variable showed significant main effects for the independent variables Tensity ($z = -11.8$, $p < 0.001$) and V/VC Ratio ($z = 3.0$, $p < 0.01$) but not for Place. The significant main effect for V/VC Ratio is compatible with H1 that neutralization is incomplete. The significant main effect for Tensity is compatible with H2 that judgements of voicing are biased by the frequency distributions in the lexicon of vowel tensity plus stop voicing combinations. The non-significant effect for Place, however, indicates that there is no difference in the degree of perceptual neutralization of the voicing distinction when there are analogous frequent contrasts in the lexicon, which is prima facie not compatible with H3.

Insert Figure 4 about here

In order to assess further the validity of H3, we compared the effect of changing V/VC on lax vs. tense vowels and alveolar vs. velar stops. According to H3, there should be less neutralization and therefore a steeper rise in the regression curve for tokens with a voicing contrast that is lexically frequent. Recall from the Introduction that the post-vocalic voicing contrast is lexically frequent in the /$V_{tns}C_{alv}$/-continuum but not in the other three continua. Fig. 4, in which the difference between the first and the last stimuli of each of the continua are linearly interpolated, shows that the rise of the /$V_{tns}C_{alv}$/-tokens' identification function is steeper than that of the /$V_{lax}C_{alv}$/-tokens' function but slighter than that of the velar tokens' functions.

Insert Figure 5 about here

Fig. 5 illustrates that the difference in the proportion of voiceless responses between stimuli 1 and stimulus 7 is, on the one hand, greater for /$V_{tns}C_{alv}$/ (21.9%) than for /$V_{lax}C_{alv}$/ (10.5%), but, on the other hand, greater for /$V_{tns}C_{vel}$/ (31.6%) than for /$V_{tns}C_{alv}$/ and greatest for /$V_{lax}C_{vel}$/ (33.3%). The four identification functions diverge significantly and this is consistent with the significant interaction effects V/VC Ratio x Tensity ($z = 2.7$, $p < 0.01$) and V/VC Ratio x Place ($z = 3.2$, $p < 0.01$). This result means that V/VC had a significantly different impact on the four continua depending on vowel tensity and place of articulation. To test for the significance of the rise of the identification functions – defined by the difference between the voiceless responses to stimuli 1 and 7 – we ran *post-hoc* analyses separately for the lax, tense, alveolar, and velar continua. There was a significant interaction between V/VC Ratio x Place for the lax continua ($z = 3.0$, $p < 0.01$), but no such significant interaction for the tense continua. This significant interaction confirmed that the difference between stimuli 1 and 7 in the proportion of voiceless responses was greater for velar than for alveolar tokens in the lax vowel context, but not in the tense vowel context. The non-significant interaction in the tense continua is not compatible with H3 that there is less perceptual neutralization for contexts in which the voicing distinction is lexically frequent. There was also a significant interaction between V/VC Ratio and Tensity for the alveolar context ($z = 3.0$, $p < 0.01$), but no such significant interaction for the velar context. That means that the difference between the two endpoint stimuli in the proportion of voiceless responses was significantly greater for tense vowels than for lax vowels but only in the alveolar context. This is compatible with H3, because tense vowel plus alveolar sequences frequently co-occur with both underlying voiced and voiceless stops while lax vowels are almost always followed by underlying voiceless stops. Since, on the other hand, velar stops are either underlying voiced when preceded by tense vowels or underlying voiceless when preceded by lax vowels (i.e. there is more or less complementary distribution of voicing in velar stops depending on vowel tensity), then velars should pattern differ from alveolars: indeed, as our results showed, for velars, in contrast to alveolars, the difference in voicing responses between stimuli 1 and 7 was about the same in lax vs. tense vowel contexts.

The *post-hoc* analyses partly support H3 that the same acoustic cues to the voicing distinction are less effective in contexts in which the voicing contrast is lexically infrequent (i.e. in contexts in which only a handful of lexical items are distinguished by post-vocalic voicing). This prediction however, is only applicable with respect to vowel tensity. In this case, listeners seem to collapse the voicing distinction in favour of the more likely category, when the voicing contrast is lexically infrequent, whereas the acoustic cues to the voicing distinction are much more effective for contexts in which the voicing distinction is lexically frequent.

Our results for place of articulation were, however, inconsistent with H3. According to H3, there should be a sharper discrimination between the endpoint stimuli if the voicing contrast occurs in a context which is lexically frequent, i.e. listeners should have been better able to hear the voicing distinction in a tense vowel plus alveolar context (analogously to frequent contrasts such as /laɪtən/ 'to lead' vs. /laɪdən/ 'to suffer') than in a tense vowel plus velar context (in which the velar is almost always underlyingly voiced) but this is not what we found. One possible explanation for this result could be the potential for resyllabification in the velar continua. In both velar continua, the syllable-final stop is resyllabifiable with the sonorant onset consonant of the second syllable and can also be interpreted perceptually as syllable initial, while in the alveolar sequences the stop is not resyllabifiable and can therefore be only interpreted as syllable final. In the second experiment, we, therefore, sought to shed more light on this possible explanation for our non-significant result by investigating whether the potential for resyllabifying the syllable-final stop influences the extent of perceived incomplete neutralization. To do this, we re-tested H1 and H4. A by-product of this experimental analysis is also a test of the *licensing-by-cue* vs. *licensing-by-prosody* hypotheses of neutralization, as outlined in the Introduction.

3. Experiment II: Effect of resyllabification on incomplete neutralization

3.1. Speech Materials

In Experiment 2, we reused the *Niedlinn-Nietlinn* (henceforth /C_{alv}-l/) and the *Mieglinn-Mieklinn* (henceforth /C_{vel}-l/) continua from Experiment 1 as well as two newly created continua each formed from resynthesized stimuli of two minimal pairs that were once again based on hypothetical German town names: *Niedstein – Nietstein* (henceforth /C_{alv}-ʃt/) and *Miegstein – Miekstein* (henceforth /C_{vel}-ʃt/). As in Experiment I, the blends were derived by combining the first syllable of a trochaic target word with the second syllable of two other context words. The target syllable of the test compounds were derived from the intervocalic productions of the two trochaic target words that were selected for Experiment 1: *Nieden* (/niːdən/) and *Miegen* (/miːgən/). The second syllable of the compound was, as before dependent on the syllable-final stop's place of articulation, either the suffix *–stein* taken from *Wirtstein* (/vɪrtʃtaɪn/) or the suffix *–stein* taken from *Birkstein* (/bɪrkʃtaɪn/). The velar place of articulation was chosen because, as described earlier, it is potentially resyllabifiable in a *–linn* context (since /gl/ and /kl/ are legal onset clusters in German) but not in a *–stein* context (because /kʃt/ or /gʃt/ are illegal onset cluster in Standard German). The method for generating the stimuli was the same as in Experiment 1. The vowel and stop closure durations as well as the step sizes of the decreasing V/VC duration ratios were the same as for the stimuli of Experiment 1 (cf. Table 1). The alveolar place of articulation was also chosen to test the *licensing-by-cue* hypothesis. In Standard German both /tl, dl/ and /tʃt, dʃt/ do not occur in syllable-initial position. Therefore, according to the *licensing-by-prosody* hypothesis, /d-t/ is necessarily domain-final and therefore categorically neutralized preceding both /l/ and /ʃt/. On the other hand, according to the *licensing-by-cue* hypothesis, the /d-t/ distinction should be more perceptible preceding the sonorant /l/ than preceding the obstruent /ʃt/ cluster since, according to this theory, sonorants but not obstruents provide a favourable context for the perceptibility of the voicing contrast.

3.2. Participants and Experimental Procedures

The experimental and analysis procedure was the same as in Experiment 1 and was run with the same subjects. Again, we first analyzed our data by means of logistic regressions, in which the logit of the voiceless responses was the dependent variable and the V/VC Ratio was the independent variable. Second, repeated measures generalized linear mixed models (GLMM) were fitted to predict incomplete neutralization as a function of the decreasing V/VC duration ratio and to determine whether the degree of incompleteness also depended on the potential for resyllabification. The 'voiceless' responses served as our dependent variable. As predictors (independent variables), we entered Place (alveolar vs. velar), Manner (sonorant vs. obstruent), and V/VC Ratio (stimuli 1–7). Subject was entered as a random effect factor.

3.3. Results and Discussion

Fig. 6 (a) gives the proportion of voiceless responses as a function of a decreasing V/VC duration ratio (i.e. stimulus number) for the four continua. In order to test whether there was a discernible trend in the proportional responses along the continuum, the corresponding logistic regression lines were calculated, as shown in Fig. 6 (b). These lines show that the identification functions for three of the four continua increase slightly, but gradually. Although these perceptual changes were not categorical, the increases in the voiceless responses along the continuum from left to right followed significant trends in /C_{vel}-l/ ($\chi^2(1) = 61.8$, $p < 0.001$), /C_{vel}-ʃt/ ($\chi^2(1) = 22.4$, $p < 0.001$), and /C_{alv}-l/ ($\chi^2(1) = 28.249$, $p < 0.001$), but not in /C_{alv}-ʃt/. The responses to /C_{vel}-ʃt/ were predominantly voiceless, but predominantly voiced to the other three continua. The results also show that listeners were much more likely to perceive a voiceless stop preceding an obstruent than a sonorant. From Figs. 6 (a) and (b) it is also clear that velars were more likely to be perceived as voiceless than alveolars.

Insert Figures 6 (a, b) about here

The results of the overall GLMM in which the voiceless responses to all four continua were included as the dependent variable showed significant main effects for Place ($z = 15.0$, $p < 0.001$) as well as for Manner ($z = -10.4$, $p < 0.001$), but none for V/VC Ratio. There were also significant interaction effects for Place x Manner ($z = -5.5$, $p < 0.001$), Place x V/VC Ratio ($z = 3.3$, $p < 0.001$), and Manner x V/VC Ratio ($z = 3.6$, $p < 0.001$). The non-significant main effect for V/VC Ratio together with the significant interaction effects for V/VC Ratio x Place and V/VC Ratio x Manner support the idea that the perceptibility of the voicing contrast was influenced by both place and manner of articulation of the two consonants.

Insert Fig. 7 about here

In order to assess the validity of H4, we compared the different effects that the acoustic cue V/VC duration ratio had upon the disambiguation of the four continua's stimuli. According to H4, there should be less neutralization for resyllabifiable /C_{vel}-l/ than for the other three non-resyllabifiable continua. As can be seen in Fig. 7, in which the difference between the first and the last stimuli of each of the four continua is linearly interpolated, the rise for the /C_{vel}-l/ identification function was steeper than for those of the other continua. Fig. 8 also shows that the difference in the proportion of voiceless responses between stimuli 1 and 7 was greater for resyllabifiable /C_{vel}-l/ than for the other non-resyllabifiable continua, but the difference was also greater for stops preceding sonorants than obstruents.

Insert Figure 8 about here

In order to test whether identification functions differed significantly with respect to place and manner of articulation, we ran *post-hoc* analyses separately for the velar, alveolar, sonorant, and obstruent continua. There was a significant main effect for V/VC Ratio ($z = 3.3$, $p < 0.001$) and also for Manner ($z = -9.0$, $p < 0.001$) and a significant interaction effect for V/VC Ratio x Manner in the velar continua ($z = 3.1$, $p < 0.01$), which means that the /C_{vel}-l/ and /C_{vel}-ʃt/ slopes differed from each other significantly. More specifically, the *post-hoc* analysis showed that the difference between stimuli 1 and 7 in the proportion of stimuli judged to be voiceless was greater in the sonorant than in the obstruent context.

There was also a significant main effect for Manner ($z = -4.5$, $p < 0.001$) and a significant interaction effect for V/VC Ratio x Manner ($z = 3.1$, $p < 0.01$) but no significant main effect for V/VC Ratio in the *post-hoc* analysis for the alveolar continua. Therefore, the divergence between the sonorant and the obstruent series in the proportion of stimuli 1 vs. 7 that were judged to be voiceless was about the same for both places of articulation.

On the other hand, there was a significant main effect of V/VC Ratio in the sonorant continua ($z = 4.1$, $p < 0.001$), but neither a significant main effect for Place, nor a significant interaction effect for V/VC Ratio x Place. This means that /C_{vel}-l/ and /C_{alv}-l/ did not differ significantly in their identification functions. Neither did /C_{vel}-ʃt/ differ from /C_{alv}-ʃt/ with respect to V/VC Ratio: That is, there was neither a significant main effect for V/VC Ratio nor a significant interaction for V/VC Ratio x Place. However, there was an overall significant main effect for Place ($z = 8.4$, $p < 0.001$). Thus, there was a greater probability of perceiving the voicing contrast in a sonorant context than in an obstruent context irrespective of the final stops' place of articulation. But the potential for resyllabification also played a role in the perceptibility of the voicing contrast: resyllabifiable clusters with a sonorant onset consonant showed the least degree of perceptual neutralization whereas non-resyllabifiable clusters with an obstruent in syllable onset showed the highest degree of neutralization.

The results provide some support for H4 that the voicing contrast is less perceptible given the same acoustic cues in non-resyllabifiable than in resyllabifiable clusters. The *post-hoc* tests showed that listeners exploit the acoustic cue V/VC duration ratio to a greater extent for the differentiation of voiced from voiceless stops when the resyllabifiable cluster contained a sonorant onset consonant as opposed to an obstruent onset consonant.

4. General Discussion

Our aim has been to establish whether the fine phonetic detail in the neutralizing context of German final obstruents, for which evidence has been presented in a substantial number of production studies (e.g. Port & O'Dell, 1985; Charles-Luce, 1985, Port & Crawford, 1989) is perceptible and moreover whether these subtle acoustic differences are conditioned by factors such as phonological frequency and effects of syllable position.

Our first hypothesis (H1) was that listeners perceive fine phonetic differences in the speech signal – in our experiment fine phonetic differences in the V/VC duration ratio – but that its power to differentiate voiced from voiceless obstruents in neutralizing contexts is substantially diminished compared with non-neutralizing contexts. Our results were consistent with this hypothesis. Although listeners were able to perceive differences between voiced and voiceless stops in a neutralizing context when acoustic cues were available for their distinction, listeners' judgements between the stimulus endpoints shifted continuously rather than categorically. These results are consistent with the idea that there is incomplete perceptual neutralization to the stop voicing contrast, or at least that the fine phonetic details of the incompletely neutralized contrast are perceptible in this kind of neutralizing context, but it is far from clear whether the cues are sufficiently powerful to distinguish unambiguously between stops. Previous research (e.g. Port & O'Dell 1985) on the perception of word-final German obstruents using natural speech showed that listeners discriminate between derived and underlying voiceless obstruents better than chance. Our results are consistent with findings from other final-devoicing languages in which the voicing contrast was shown to be perceived gradually (Warner et al., 2004) or even categorically (Broersma, 2005) when obstruents were in a neutralizing context. Our results are also consistent with those for Dutch listeners who were shown to 'borrow' the intervocalic duration cue to obstruent voicing word-finally (Warner et al. 2004). Taken together, all these results show that acoustic information in a neutralizing context can be used for disambiguating voicing. There is further support from our results that incomplete neutralization is not an artefact of the experimental design (Fourakis & Iverson, 1984).

According to H2, there should be a greater probability of identifying stops as voiceless after lax than after tense vowels (i.e. *Widdlinn* and *Bigglinn* in our experiment should be perceived more often as *Wittlinn* and *Bicklinn*, respectively), since lax vowels are so rarely preceded by underlying voiced obstruents in German. Our results confirmed that there was a greater probability of listeners identifying more stops as voiceless when they occurred after lax vowels. More specifically, for the same V/VC duration ratio, listeners identified a greater proportion of syllable-final consonants as voiceless after lax than after tense vowels irrespective of the syllable-final stop's place of articulation. This finding supports the idea that incomplete neutralization also depends on the frequency with which a vowel and following consonant co-occur and that the phonological [±tense] feature is not syntagmatically independent of a following [±voice] in syllable-final vowel plus consonant sequences in German. That is, lexical frequencies of phoneme sequences strongly affect listeners' phoneme identification and categorization. In our materials, the lax vowel was a phonological cue that evoked the listener's expectation of a following voiced stop, i.e. listeners made predictions about the voicing category of the following stop based on their knowledge of phonotactic constraints. Our results are in line with findings that listeners adjust to high frequent categories (Hay et al. 2003, Pitt & McQueen 1998) and therefore perceive more tokens as containing a frequent cluster (as opposed to an infrequent one).

As far as tense vowels and the stops' place of articulation are concerned, H3 predicted that the voicing contrast should be more perceptible for alveolar stops than for velar stops given that stop voicing occurs frequently in the former (e.g. /laɪdən/ 'to suffer', /laɪtən/ 'to lead') but not the latter context. Velar stops after tense vowels should be perceived predominantly as voiced because of the paucity of underlying tense vowel plus voiceless velar sequences (i.e. sequences like /piːksən/ 'to prick' with an underlying /iːk/ are very rare in German). Our results were not compatible with this position (H3) and showed instead that stop voicing was facilitated in a velar compared with an alveolar context. However, compatibly with H3 we did find that the stop voicing contrast was more perceptible in alveolars when they were preceded by tense than by lax vowels. This presumably comes about because the voicing contrast in alveolar stops preceded by lax vowels is infrequent, i.e. the stop is almost always a reflex of underlying /t/, whereas, the voicing contrast in alveolar stops preceded by tense vowels is frequent (e.g. /miːdən/ 'to avoid' vs. /miːtən/ 'to rent'). In the latter case, listeners are not biased by the lexical voicing distribution and therefore cannot make predictions about the following stop's voicing status. Instead, the voicing disambiguation has to be based on the acoustic information that is available. After lax vowels, the perceptibility of the acoustic cue is constrained due to the perceptual bias towards voiceless stops. Within this interpretation, the listener's meta-linguistic knowledge facilitates or restricts the perceptibility of a strong acoustic cue at least in certain environments (in this case in *Niedlinn – Nietlinn* and *Widdlinn – Wittlinn*, respectively) and this together with H2 partially supports the claim made in a number of statistically based phonological models (Bybee, 2001, 2004; Pierrehumbert 2001, Coleman 2003) that probability has to be taken into account when modelling phonological neutralization.

The starting point for H4 was to test whether resyllabification could explain our finding (which ran counter to H3) that stop voicing was facilitated in a velar compared with an alveolar context. H4 also allowed us to adjudicate between the two diverging phonological approaches to modelling final devoicing – *licensing-by-cue* (Steriade, 1997, 1999, 2001) and *licensing-by-prosody* (e.g. Itô 1986, 1989; Goldsmith, 1990; Rubach, 1990). According to the latter, final devoicing is controlled by prosodic position, whereas according to the former, segmental context is decisive. More specifically, according to the *licensing-by-prosody* hypothesis, there should be a perceptual advantage for the voicing distinction in a velar as opposed to an alveolar context. This is because, since velars but not alveolars can form legal onset clusters with /l/ in German (e.g., /ɡl/ vs. /kl/, /ɡlɔtsən/ 'to stare' vs. /klɔtsən/ 'to slog away'), listeners should be able to identify the voicing status of a velar stop in /V_1-stop-l-V_2/ in which the stop can potentially be interpreted as part of an onset cluster of the second /l-V_2/ syllable and is therefore not subject to domain-final neutralization. By contrast, since velars and alveolars precede the same sonorant context, then according to the *licensing-by-cue* hypothesis, voicing distinctions should be equally favourable in both contexts. In addition, voicing distinctions should be more perceptible preceding the sonorant /l/ context than preceding the /ʃt/ context because cues for the voicing distinction are less likely to be obscured preceding sonorants than obstruents.

We found some support for the idea that the voicing distinction is more perceptible in resyllabifiable velar plus lateral sequences as opposed to the non-resyllabifiable alveolar plus lateral sequences which is compatible with the prediction from *licensing-by-prosody*. We also found that voicing distinctions were in general more perceptible for both alveolars and velars preceding sonorants than fricatives. This lends some support to the *licensing-by-cue* hypothesis, since sonorants provide a more stable context than fricatives for the voicing distinction to be realised. It runs counter to the *licensing-by-prosody* hypothesis, however, which predicts that the perceptibility of the voicing distinction in alveolars preceding either sonorants or fricatives should be the same, since in both cases the alveolar is necessarily domain-final (because it is not resyllabifiable).

In general, our results provide some evidence that listener's knowledge of phonotactic constraints may perceptually mask lexical redundant acoustic cues (Cutler, 2002). Perceptual masking is the presumed process by which listeners ignore cues in the signal if they are irrelevant phonologically. We propose that perceptual masking varies with the extent to which neutralization is complete. For example, there was less neutralization in the resyllabifiable /kl/ cluster than in the non-resyllabifiable /kʃt/ cluster which suggests that listeners take advantage of the cues as long as they are consistent with phonological distributions.

A comparison of the results of both experiments shows predominant /g/ responses after tense vowels only preceding the sonorant -*linn*, but not preceding the obstruent -*stein* suffix. If listeners' responses had been guided by phonotactic probability, then they should have perceived /g/ preceding both -*linn* and -*stein*. The fact that they perceived /g/ before -*linn* and /k/ before -*stein* might instead suggest that their perceptions were guided by regressive voicing assimilation and not by phonotactic probability: that is, the voiced sonorant context induced (regressively) primarily /g/ perceptions in the preceding stop whereas the voiceless obstruent context elicited mostly preceding /k/ perceptions. But on the other hand, if right context is the only factor in determining voicing perception of the preceding stop, then listeners should have perceived predominantly /t/ when the tense vowel plus alveolar sequence preceded voiceless obstruents, but this is not what we found: they instead perceived /d/. Therefore, while phonotactic probability (H3) cannot entirely explain why tense vowel and following velar sequences should be perceived as /g/ before -*linn*, but as /k/ before -*stein*, neither in view of the results from alveolars can this be explained by regressive assimilation alone.

The different perceptual responses to the continua in this study – which depend to some extent also on probabilistic co-occurrences of V+C sequences, the phonetic environment, and the potential for resyllabification – cannot be easily modelled in a generative framework. Phonological rules in generative theories are either applied or not. For that reason these rules are less able to give expression to gradient changes in phoneme perception (unless they include a vast number of constraining rules). In probabilistic phonological theories such as exemplar or episodic models of speech perception (Pierrehumbert 2001, 2003), a phonological category such as voicing is defined by the density distribution in an acoustic-perceptual space. This distribution is continuously expanded by new remembered exemplars, which are compared with previous stored exemplars and added to the neighbourhood of the most similar exemplars within that space. In such a usage based model, the density distribution for lexically frequent sequences (e.g. /bɪtən/ 'to beg') is very likely to be high and enlarged across the perceptual space and very low and narrowed for rare sequences (e.g. /ɛbə/ 'tide'/), i.e. language-dependent lexical frequency distributions are stored in the listener's mental lexicon. As a consequence listeners re-interpret rare as frequently occurring sequences (e.g. Pitt & McQueen, 1998, Hay et al, 2003). On the other hand, in phoneme sequences that show a balanced distribution across the lexicon (e.g. /miːdən/ 'to avoid' vs. /miːtən/ 'to rent'), a misinterpretation of a phonological feature is less likely since similar density distribution are to be expected. These predictions from exemplar theory are compatible with our findings for Experiment 1. Similarly, an exemplar framework predicts that the voicing contrast in velar stop + /l/ sequences should be readily perceptible analogous to the frequently occurring /kl/ vs. /gl/ contrast in German; and it also predicts that the voicing contrast should be much less perceptible before obstruents because there are no instances of a voicing contrast in this context in the lexicon. These predictions are compatible with our findings.

Moreover, an exemplar model also explains the finding from Experiment 2 that listeners exploit the acoustic cue V/VC duration ratio to distinguish voiced from voiceless velar stops in the non-resyllabifiable /C_{vel}-ʃt/ context. Despite the obstruent context and the non-resyllabifiable cluster, the discrimination performance of the stimuli of this continuum was rather good. One partial reason for this fairly good discrimination could be that forms such as [g̊ʃtoːsn] 'pushed' and [g̊ʃtoːln] 'stolen' in which the schwa from the first syllable is elided (the citation forms have initial /gə/ in both cases) are quite common in spontaneous and colloquial speech in German: consequently, native German listeners are likely to have been exposed a good deal to onset clusters such as /g̊ʃt/. According to exemplar theories, remembered exemplars for these alternative pronunciations with /g̊ʃt/ might be stored as alternative pronunciation forms in the mental lexicon. Furthermore, the discrimination performance for the obstruent and non-resyllabifiable /C_{alv}-ʃt/ continuum was very poor. Besides the unfavourable phonetic context and the missing potential for resyllabification – which were no such strong impediments to the perception of the voicing contrast in the /C_{vel}-ʃt/ cluster – the /tʃt/-cluster is, in contrast to the corresponding form with initial velars, a more or less non-occurring onset cluster in German even in spontaneous and colloquial speech. Therefore, we expect that German listeners cannot draw on stored /tʃt/ or /dʃt/ exemplars when discriminating voicing in this cluster. In the case of /gʃt/, however, German listeners will only have been exposed to devoiced [g̊ʃ] (because of the high frequency of occurrence of the past tense marker /gə/ with the reduced or elided schwa) but scarcely to /kʃ/ onset clusters (since /kə/ is far less frequent than initial /gə/), so that we are not dealing with a potentially frequent voicing contrast in velar stop plus /ʃ/ sequences, which would, according to H3, have yielded a greater voice contrast perceptibility.

To conclude, we have found that the extent of perceptual neutralization is influenced by phonological frequency. Furthermore, the potential for resyllabification of the final stop enhances the perceptibility of the voicing contrast in domain-final obstruents. The phonetic environment is another important factor that affects the perceptibility. Final devoicing very likely cannot only be characterized in terms of either prosodic position or phonetic context. Instead prosodic position, phonological and phonotactic frequency as well as phonetic context have to be included to model adequately the neutralization of the final voicing contrast in German. But there may well be other contextual factors (which have not been addressed in the present paper) that affect the degree of incomplete neutralization such as, for example, placement of morphological boundaries in relation to the obstruent (e.g. Iverson & Salmons, 2007). Evidence has been provided that incomplete neutralization is an instance of fine phonetic detail that is perceptible when listeners are forced to distinguish underlying voiced from voiceless stops in a neutralizing context. Listeners exploit acoustic cues but only as long as these are consistent with phonotactic generalizations. Our findings so far show that incomplete neutralization plays a “substantial role in language processing” (Ernestus & Baayen, 2006: p. 27), but whether it is morphological functional in German as has been shown for Dutch where “it appears to be a subphonemic cue to past-tense formation” (Ernestus & Baayen, 2006: p. 27) still needs to be addressed.

Acknowledgments

We are very grateful to Jennifer Hay, Robert Port, and an anonymous reviewer for their helpful comments. We thank Michaela Adlwart and Julia Orlitzky for help with stimulus generation, Benjamin Hamer for help with recruiting subjects and running the experiments, and Ulrich Reubold for help with figure preparation. This research was supported by German Research Council Grant no. HA3512/4-1, "Acoustic, articulatory and perceptual analyses of the post-vocalic voicing contrast in two varieties of German." to Jonathan Harrington. The project is part of the German Research Council Priority Programme 1234: "Phonological and phonetic competence: between grammar, signal processing, and neural activity".

References

Baayen, H. R. (2008). Analyzing Linguistic Data. A Practical Introduction to Statistics Using R. Cambridge: Cambridge University Press.

Baumann, M. (1995). *The production of syllables in connected speech*. Unpublished Ph.D. dissertation, University of Nijmegen.

Beckman, J. (1997). Positional faithfulness, positional neutralisation, and Shona vowel harmony. *Phonology*, *14*, 1–46.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman, & H. C. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 167–224). Cambridge, MA: MIT.

Boersma, P., & Weenink, D. (2008). Praat: doing phonetics by computer (Version 5.0.27) [Computer program]. Retrieved June 28, 2008, from http://www.praat.org/

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America, 117*, 3890-3901.

Brockhaus, W. (1995). *Final Obstruent Devoicing in the Phonology of German*. Tübingen: Niemeyer.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, J. (2004). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*, 261 – 290.

Charles-Luce, J. (1993). The effects of semantic context on voicing neutralization. *Phonetica*, *50*, 28–43.

Charles-Luce, J., & Dinnsen, D. (1987). A reanalysis of Catalan devoicing. *Journal of Phonetics*, *15*, 187–190.

Charles-Luce, J. (1985). Word-final devoicing in German: Effects of phonetic and sentential contexts. *Journal of Phonetics*, *13*, 309–324.

Coleman, J. (2003). Commentary: probability, detail and experience. In J. Local, R. Ogden, & R. Temple (Eds.) *Papers in Laboratory Phonology VI* (p. 88–100). Cambridge: Cambridge University Press.

Cutler, A. (2002). Phonological processing: comments on Pierrehumbert, Moates et al., Kubozono, Peperkamp & Dupoux, and Bradlow. In C. Gussenhoven & N. Warner (Eds.). *Laboratory Phonology 7* (pp. 275-296). Cambridge: Cambridge University Press.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden-path: segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28(1),* 218–244

Dinnsen, D. A., & Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics 12*, 49–60.

Ernestus, M., & Baayen, H. (2006). The functionality of incomplete neutralization in Dutch: the case of past-tense formation. In L. M. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory Phonology 8* (pp. 27–49). Berlin: Mouton de Gruyter.

Féry, C. (2003). Final Devoicing and the stratification of the lexicon in German. In van de Weijer, J. M., van Heuven, V. J., & van der Hulst, H. (Eds.). *The Phonological Spectrum: Segmental Structure* (pp. 145–170). Amsterdam: John Benjamins.

Fourakis, M., & Iverson, G. K. (1984). On the ‘incomplete neutralization’ of German final obstruents. *Phonetica, 41*, 140-149.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on processing of nonword sound patterns. *Journal of Memory and Language, 42*, 481–496.

Goldsmith, J. A. (1990). *Autosegmental and Metrical Phonology*. Oxford: Blackwell.

Hall, T. A. (1992). *Syllable Structure and Syllable-Related Processes in German*. Tübingen: Niemeyer.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*, 373-405.

Hawkins, S. & Nguyen, N. (2004). Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *Journal of Phonetics*, *32*, 199-231.

Hawkins, S., & Nguyen, N. (2003). Effects on word recognition of syllable-onset cues to syllable-coda voicing. In J. K. Local, R. A. Ogden, & R. A. M. Temple (Eds.), *Papers in Laboratory Phonology VI* (38–57). Cambridge: Cambridge University Press.

Hay, J., Pierrehumbert, J., & Beckman, M. E. (2003**.** Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Laboratory Phonology VI*. Cambridge: Cambridge University Press.

Itô, J. (1989). A prosodic theory of epenthesis. *Natural Language and Linguistic Theory*, *7*, 217-259.

Itô, J. (1986). Syllable theory in prosodic phonology. Ph.D. thesis, University of Massachusetts at Amherst.

Iverson, G. K. & Salmons, J. S. (2007). Domains and directionality in the evolution of German final fortition. *Phonology*, *24*, 121-145.

Jassem, W., & Richter, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics*, *17*, 317–325.

Jongman, A., Sereno, J. A., Raaijmakers, M., & Lahiri, A. (1992). The phonological representation of [voice] in speech perception," *Language and Speech*, *35*, 137–152.

Keating, P. (1979). *A Phonetic Study of a Voicing Contrast in Polish*. Unpublished Ph.D. dissertation, Brown University.

Kohler, K. J. (1984). Phonetic explanations in phonology: the feature fortis/lenis. *Phonetica*, *31*, 150–174.

Kohler, K. J. (1979). Dimensions in the perception of fortis and lenis plosives. *Phonetica*, *36*, 332–343.

Kubozono, H. (2002). Temporal neutralization in Japanese. In C. Gussenhoven & N. Warner (Eds.) *Laboratory Phonology 7* (pp. 171–2002). Cambridge: Cambridge University Press.

Lombardi, L. (1999). Positional faithfulness and voicing assimilation in optimality theory. *Natural Language and Linguistics Theory*, *17*, 267–302.

Manuel, S.Y. (1995). Speakers nasalize /o/ after /n/, but still hear /o/. *Journal of Phonetics*, *23*, 453–476.

O'Dell, M. & Port, R. F. (1983). Discrimination of word-final voicing in German. *Journal of the Acoustical Society of America*, *73(S1)*, S31(A).

Pierrehumbert, J. (2003). Probabilistic phonology: discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.) *Probabilistic Linguistics* (pp. 177-228). Cambridge: MIT Press.

Pierrehumbert, J. (2001). Exemplar dynamics: word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.) *Frequency and the Emergence of Linguistic Structure* (pp. 137-158). Amsterdam: John Benjamins.

Piroth, H. G. & Janker, P. M. (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics*, *32*, 81–109.

Pitt, M. & McQueen, J. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, *39*, 347–370.

Port, R., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics 17*, 257-282.

Port, R. F., & O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics 13*, 455– 471.

Port, R. F., Mitleb, F. & O'Dell, M. L. (1981) Neutralization of obstruent voicing in German is incomplete. *Journal of the Acoustical Society of America*, *70*, S13, F10.

Rubach, J. (1990). Final devoicing and cyclic syllabification in German. *Linguistic Inquiry*, *21*, 79–94.

Slowiaczek, L., & Szymanska, H. (1989). Perception of word-final devoicing in polish. *Journal of Phonetics*, *17*, 205–212.

Slowiaczek, L., & Dinnsen, D. (1985). On the neutralizing status of polish wordfinal devoicing. *Journal of Phonetics*, *13*, 325–341.

Steriade, D. (1997). Phonetics in phonology: the case of laryngeal neutralization. UCLA manuscript copy at: http://www.humnet.ucla.edu/humnet/linguistics/people/steriade/steriade.htm

Steriade, D. (1999). Alternatives to the syllabic interpretation of consonantal phonotactics. In O. Fujimura, B. Joseph, & B. Palek (Eds.) *Proceedings of the 1998 Linguistics and Phonetics Conference* (pp. 205–242). Prague: The Karolinum Press,.

Steriade, D. (2000). Paradigm uniformity and the phonetics-phonology boundary. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V* (pp. 313–334). Cambridge: Cambridge University Press.

Vennemann, T. (1972). On the Theory of Syllabic Phonology. *Linguistische Berichte*, *18*, 1–18.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*, 374–408.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactic and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*, 47–62.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, *32*, 251–276.

Warner, N., Good, E., Jongman, A. & Sereno, J. (2006). Orthographic versus morphological incomplete neutralization effects. *Journal of Phonetics*, *34*, 285-293.

Werker, J.F. (1995). Age-related changes in cross-language speech perception: standing at the crossroads. In Strange, W. (1995). *Speech Perception and Linguistic Experienc*e (pp. 155–169). Baltimore: York Press.

Werker, J. F. & Tess, R. C. (1998). Influences on infant speech processing: toward a new synthesis. *Annual Review of Psychology*, *50*, 509–535.

Wiese, R. (1996). *The Phonology of German*. Oxford: Clarendon Press.

Wright, R. (2003). Factors of lexical competition in vowel articulation. In J. K. Local, R. A. Ogden, and R. A. M. Temple (Eds.), *Papers in Laboratory Phonology VI.* Cambridge: Cambridge University Press.

Figure 1

Schematic representation of the f0-contour manipulation.

Figure 2

Schematic representations of five possible psychometric curves representing the percentage of voiceless responses as a function of decreasing V/VC duration ratio.

Figures 3 (a, b)

Proportion of ‘voiceless’ responses as a function of decreasing V/VC duration ratio (stimulus number) to the four continua differing in vowel tensity and place of articulation (a) and the corresponding regression curves (b): lax (grey), tense (black), alveolar (solid), and velar (dashed).

Figure 4

Linear interpolations between the proportional ‘voiceless’ responses to stimuli 1 and 7 of the four continua differing in vowel tensity and place of articulation: lax (grey), tense (black), alveolar (solid), and velar (dashed).

Figure 5

Proportional differences of ‘voiceless’ responses between stimuli 1 and 7 of the four continua.

Figures 6 (a, b)

Proportion of ‘voiceless’ responses as a function of decreasing V/VC duration ratio (stimulus number) to the four continua differing in place and manner of articulation and the corresponding regression curves (b): alveolar (grey), velar (black), sonorant (solid), and obstruent (dashed).

Figure 7

Linear interpolations between the proportional ‘voiceless’ responses to stimuli 1 and 7 of the four continua differing in the offset’s place and the onset’s manner of articulation: alveolar (grey), velar (black), sonorant (solid), and obstruent (dashed).

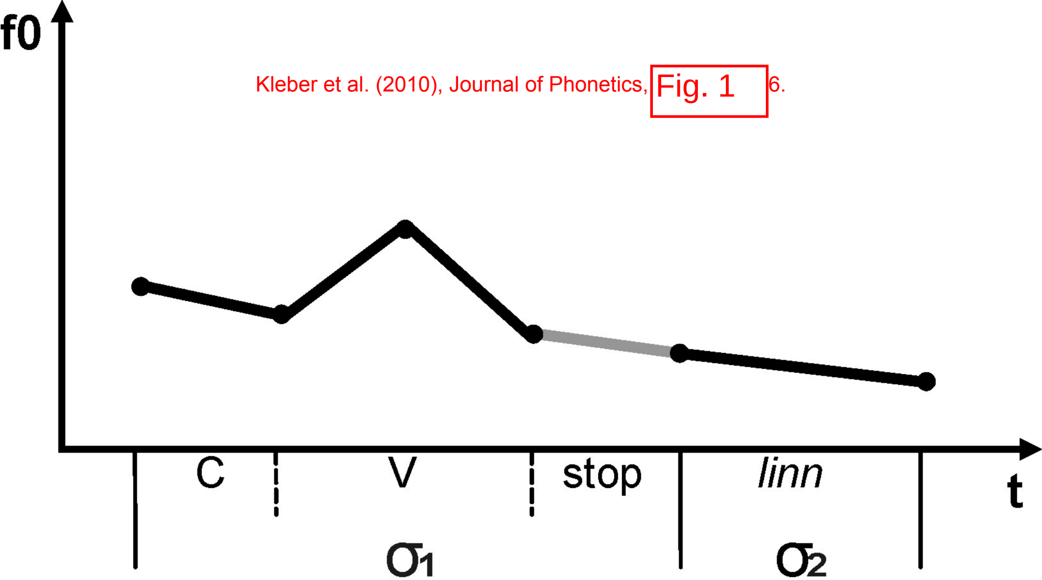
Figure 8

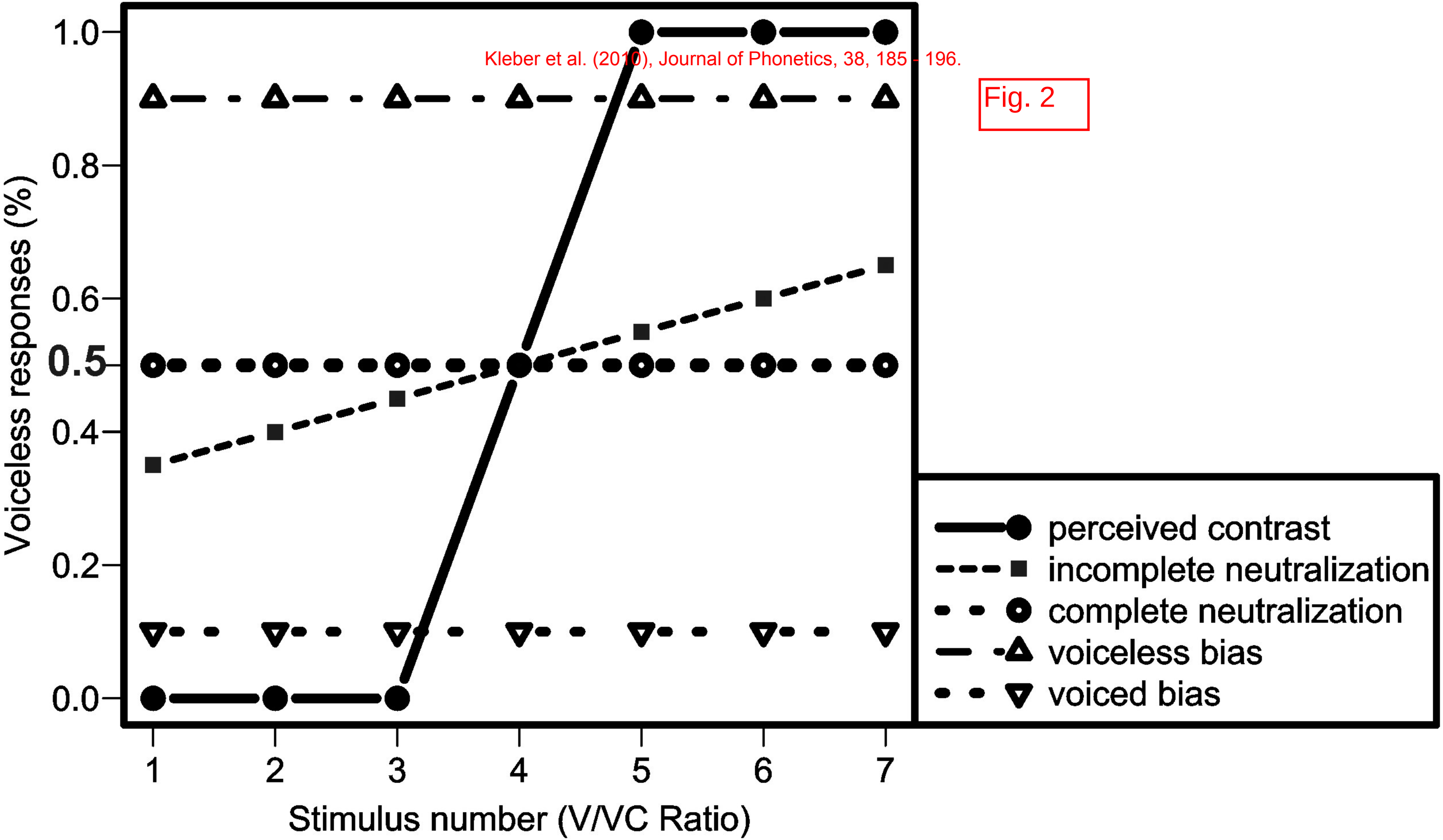
Proportional differences of ‘voiceless’ responses between stimuli 1 and 7 of the four continua.

Table 1 ***Segment durations and V/VC duration ratios for each stimulus of each continuum***

VC-Sequence	Segment	Stimulus number						
		1[a]	2	3	4	5	6	7
/ɪd/ - /ɪt/	V	90	83	77	70	63	56	50
	C	122	129	135	142	149	156	162
	VC	212	212	212	212	212	212	212
	V/VC duration ratio	0.42	0.39	0.36	0.33	0.30	0.27	0.23
/ɪg/ - /ɪk/	V	75	68	62	55	48	41	34
	C	96	103	109	116	123	130	137
	VC	171	171	171	171	171	171	171
	V/VC duration ratio	0.44	0.40	0.36	0.32	0.28	0.24	0.20
/i:d/ - /i:t/	V	192	178	165	151	137	124	110
	C	73	87	100	114	128	141	155
	VC	265	265	265	265	265	265	265
	V/VC duration ratio	0.72	0.67	0.62	0.57	0.52	0.47	0.42
/i:g/ - /i:k/	V	192	172	153	133	114	94	75
	C	71	91	110	130	149	169	188
	VC	263	263	263	263	263	263	263
	V/VC duration ratio	0.73	0.66	0.58	0.51	0.43	0.36	0.28

[a] Selected voiced token.


f0
Kleber et al. (2010), Journal of Phonetics,
Fig. 1
6.
C
V
stop
linn
t
σ1
σ2



Kleber et al. (2010), Journal of Phonetics, 38, 185 - 196.
Voiceless responses (%)
1.0
0.8
0.6
0.5
0.4
0.2
0.0
1
2
3
4
5
6
7
Stimulus number (V/VC Ratio)
perceived contrast
incomplete neutralization
complete neutralization
voiceless bias
voiced bias

Fig. 2

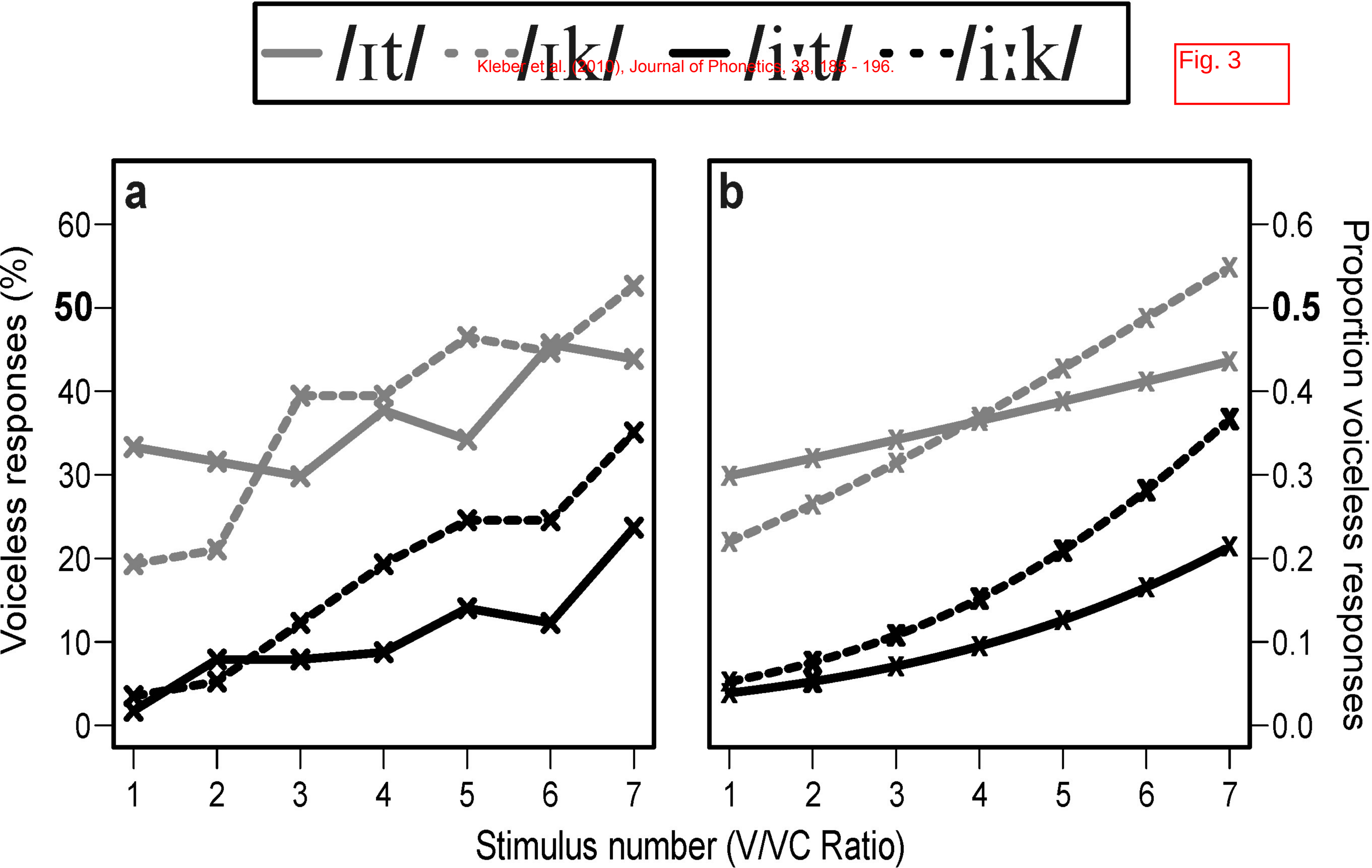
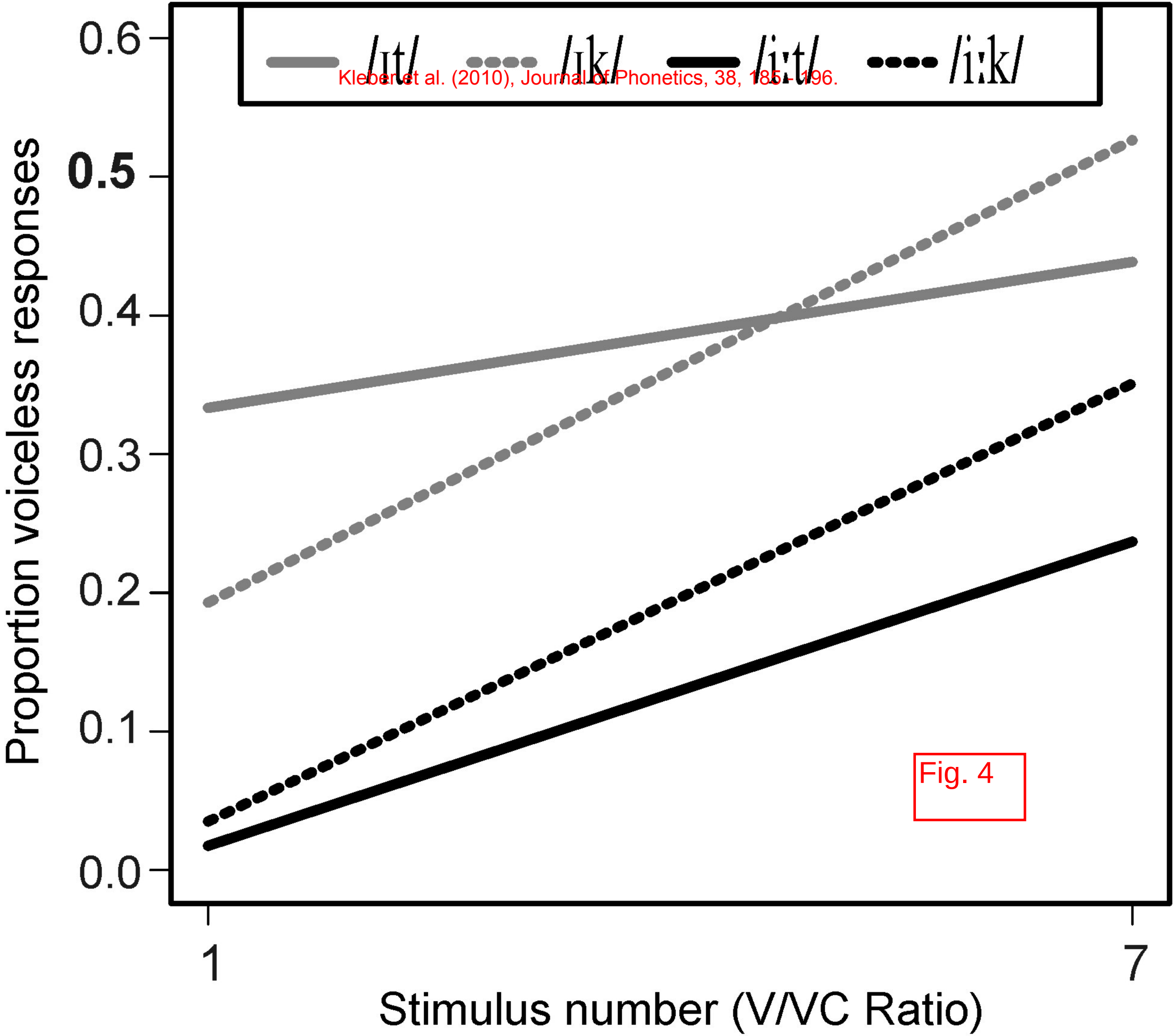


Fig. 3

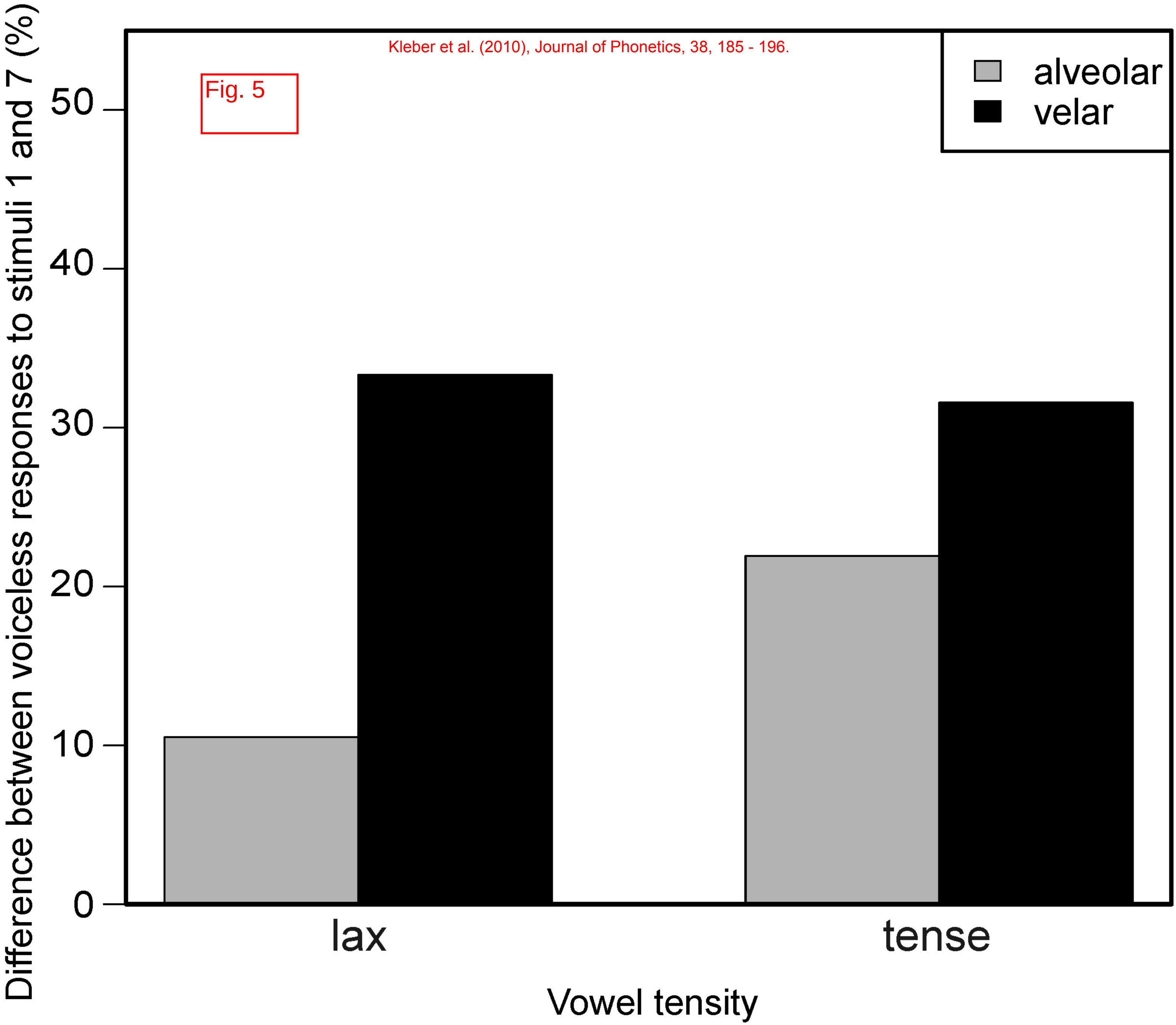
Kleber et al. (2010), Journal of Phonetics, 38, 185 - 196.



/ɹt/
/ɹk/
/i:t/
/i:k/
Kleber et al. (2010), Journal of Phonetics, 38, 185-196.
Proportion voiceless responses
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1
7
Stimulus number (V/VC Ratio)
Fig. 4

Kleber et al. (2010), Journal of Phonetics, 38, 185 - 196.

Fig. 5



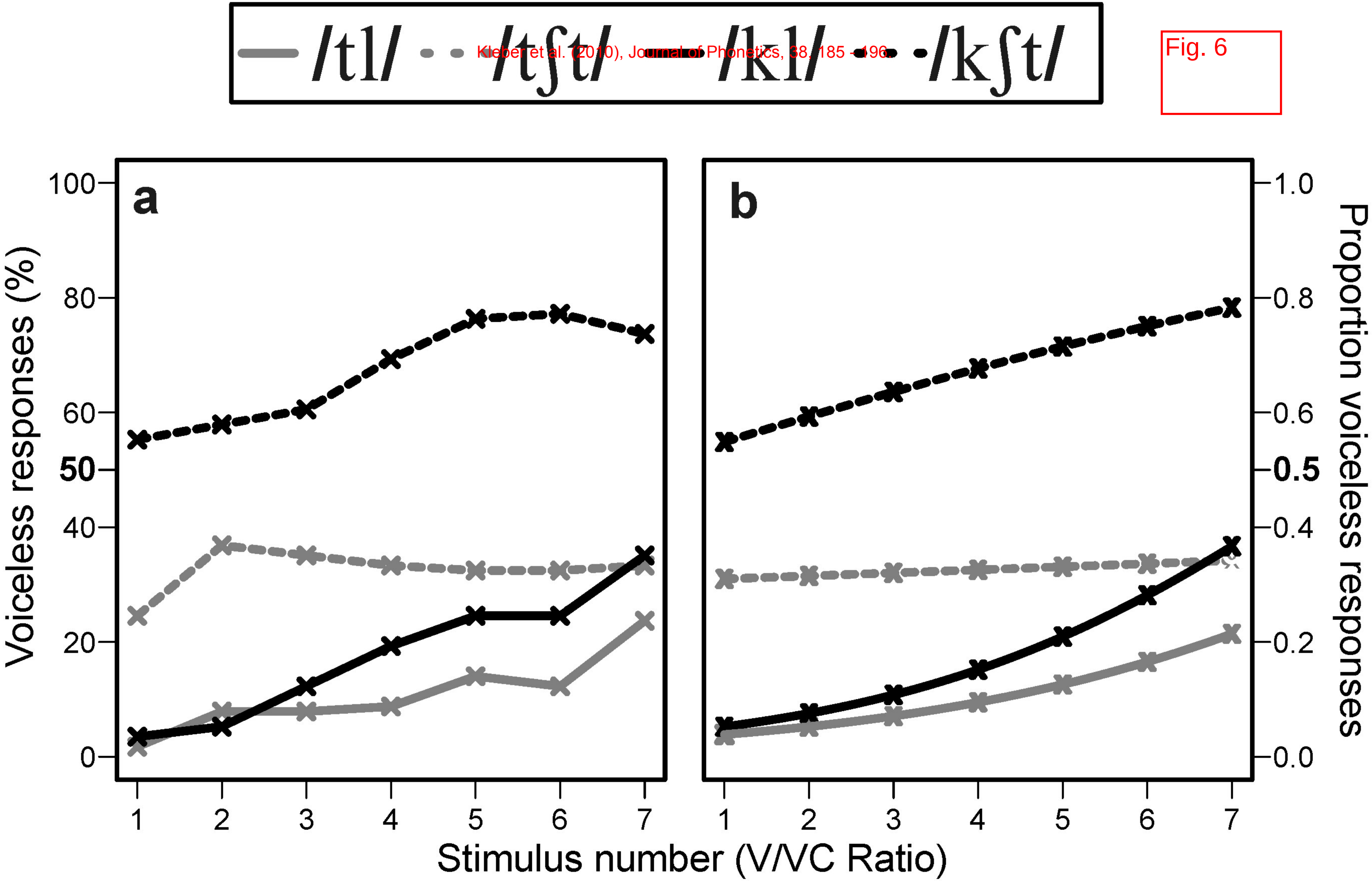
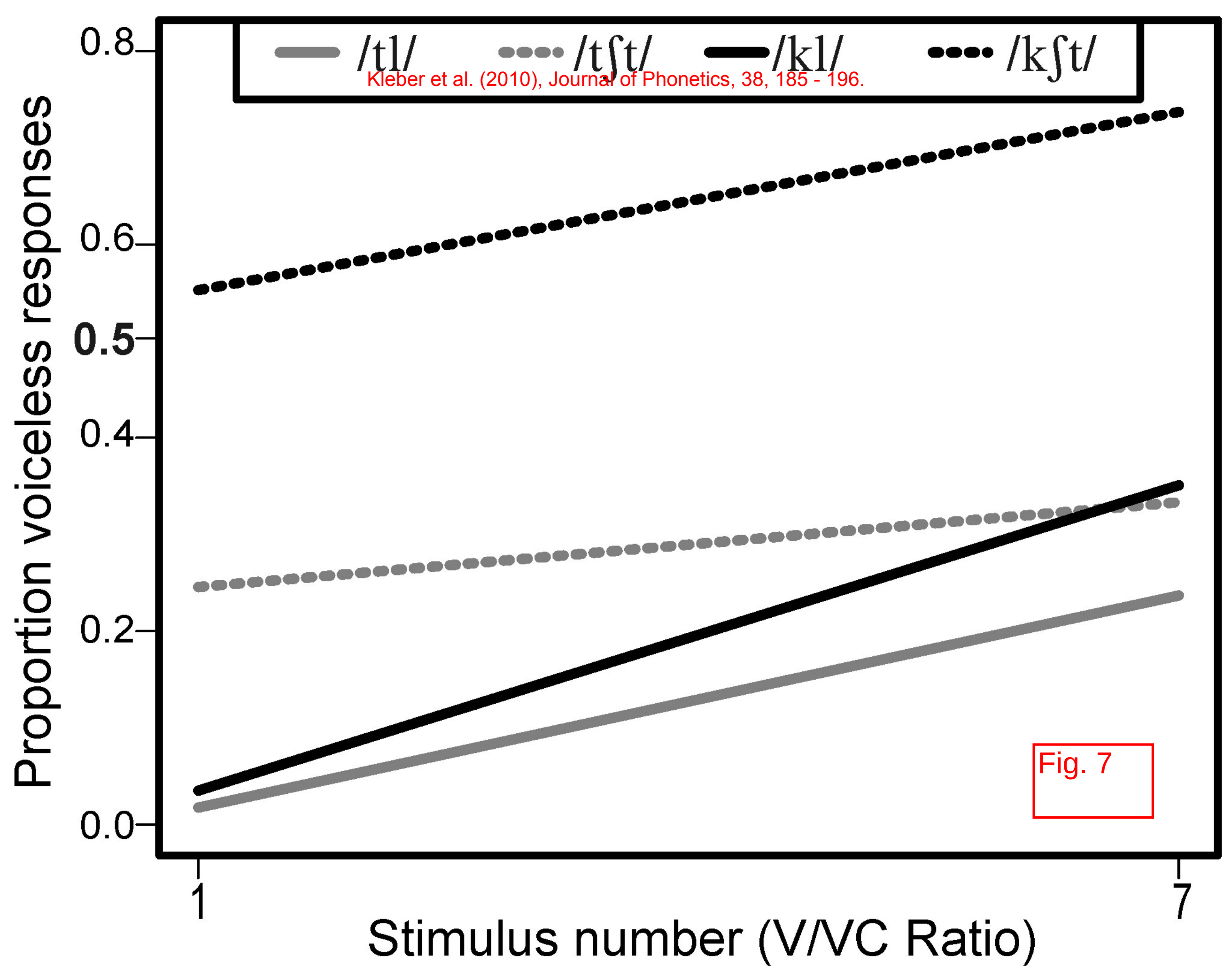
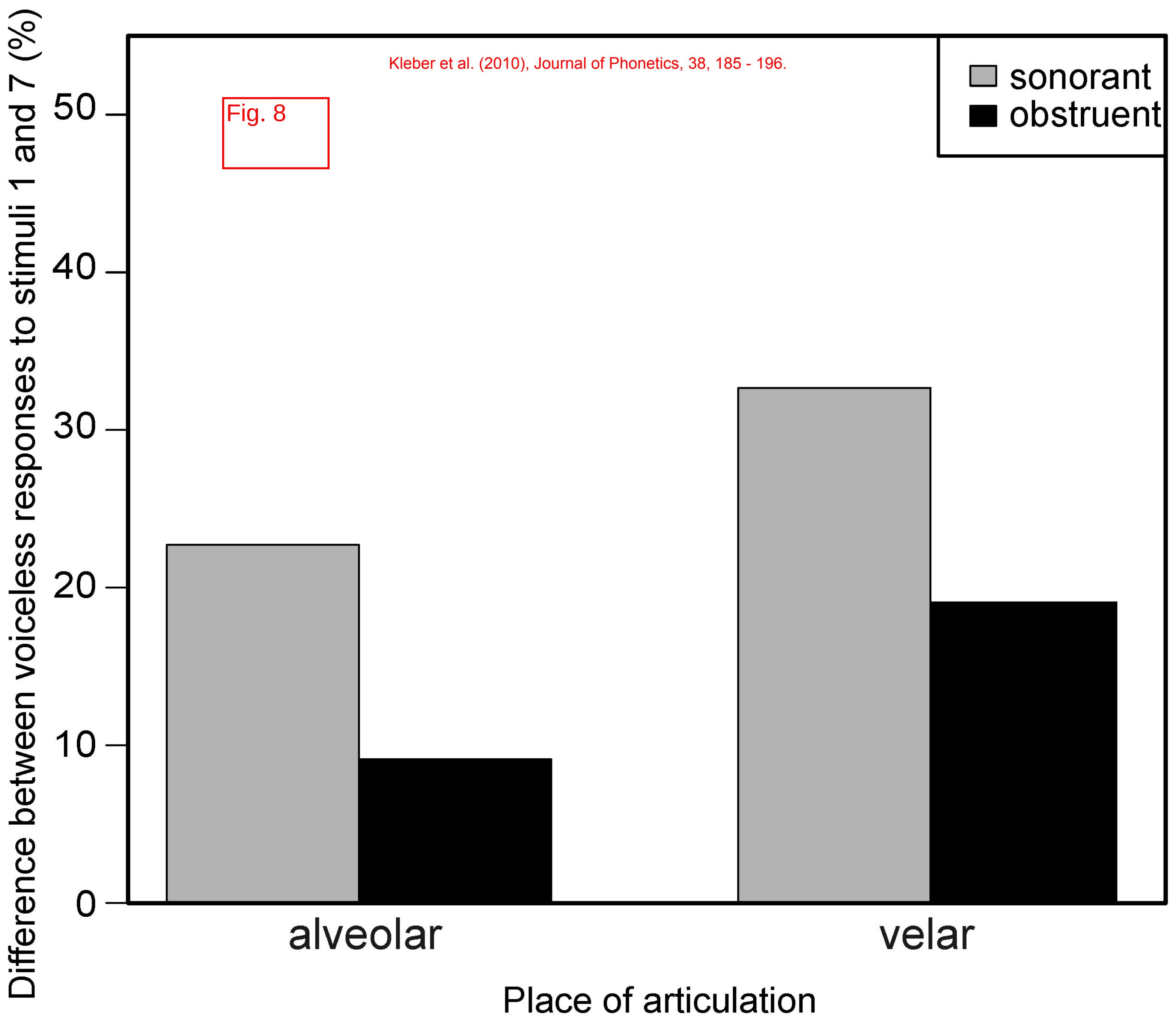

/tl/
/tʃt/
/kl/
/kʃt/
Kleber et al. (2010), Journal of Phonetics, 38, 185 - 196.
a
b
Voiceless responses (%)
Proportion voiceless responses
Stimulus number (V/VC Ratio)
100
80
60
50
40
20
0
1.0
0.8
0.6
0.5
0.4
0.2
0.0
1
2
3
4
5
6
7

Fig. 6



/tl/
/tʃt/
/kl/
/kʃt/
Kleber et al. (2010), Journal of Phonetics, 38, 185 - 196.
Proportion voiceless responses
0.8
0.6
0.5
0.4
0.2
0.0
1
7
Stimulus number (V/VC Ratio)
Fig. 7



Kleber et al. (2010), Journal of Phonetics, 38, 185 - 196.
Fig. 8
sonorant
obstruent
Difference between voiceless responses to stimuli 1 and 7 (%)
0
10
20
30
40
50
alveolar
velar
Place of articulation