Journal of Phonetics 66 (2018) 185–216

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis

Andries W. Coetzee [a,b,*], Patrice Speeter Beddor [a], Kerby Shedden [a], Will Styler [a], Daan Wissing [b]

[a] University of Michigan, Ann Arbor, MI 48109, USA
[b] North-West University, Potchefstroom 2531, South Africa

ARTICLE INFO

Article history:
Received 17 October 2016
Received in revised form 28 September 2017
Accepted 28 September 2017
Available online 8 November 2017

Keywords:
Voicing
Fundamental frequency
Afrikaans
Speech perception
Production-perception relation
Cue weighting

ABSTRACT

This study documents the relation between f0 and prevoicing in the production and perception of plosive voicing in Afrikaans. Acoustic data show that Afrikaans speakers differed in how likely they were to produce prevoicing to mark phonologically voiced plosives, but that all speakers produced large and systematic f0 differences after phonologically voiced and voiceless plosives to convey the contrast between the voicing categories. This pattern is mirrored in these same participants' perception: although some listeners relied more than others on prevoicing as a perceptual cue, all listeners used f0 (especially in the absence of prevoicing) to perceptually differentiate historically voiced and voiceless plosives. This variation in the speech community is shown to be generationally structured such that older speakers were more likely than younger speakers to produce prevoicing, and to rely on prevoicing perceptually. These patterns are consistent with generationally determined differential cue weighting in the speech community and with an ongoing sound change in which the original consonantal voicing contrast is being replaced by a tonal contrast on the following vowel.

1. Introduction

The contrast between voiced and voiceless plosives is cued by multiple acoustic properties both within and across languages. Two widely recognized properties are the onset of voicing relative to release of plosive closure (voice onset time or VOT) and the fundamental frequency (f0) of the vowel following the plosive. Although VOT is often the primary information for this contrast, post-plosive f0 has been shown to provide reliable, even if typically less robust, information. The basic pattern is for vowels to have higher f0 after voiceless than after voiced plosives.

This study examines the contributions of VOT and f0 to voicing contrasts in contemporary Afrikaans. It investigates how older and younger speakers in an Afrikaans speech community produce word-initial plosives, and how they perceptually differentiate between the two plosive voicing categories. Anticipating one main outcome, we find that the VOT differences between phonologically voiced and voiceless plosives vary both within speakers and between generations, with younger speakers being especially likely to produce phonologically voiced plosives as voiceless unaspirated.[1] We explore whether this generational difference in plosive VOT production is accompanied by age differences in vocalic f0 production. We investigate as well whether production patterns for VOT and f0 align with listeners' perceptual use of the two properties. The results are interpreted relative to the literature on cue weighting and sound change, especially in terms of whether the findings are more indicative of a situation of variable cue weighting or of an ongoing sound change in which the historical voicing contrast may be in the process of being replaced by a tonal contrast.

In this introductory section, we review relevant background about cue weighting, about how small f0 perturbations due to consonantal voicing can over time be exaggerated and start functioning as independent phonemic tonal contrasts, about the history of the plosive voicing contrast in Afrikaans, and

[1] Throughout this paper, we will refer to the historically voiced plosives of Afrikaans (that are often realized as voiceless in terms of VOT) as either "historically voiced" or "phonologically voiced". These designations are used for the sake of convenience and are not intended to indicate that Afrikaans has lost plosive voicing or that the phonological contrast between the plosive categories is necessarily a voicing contrast for all speakers. As we document below, many, especially older, speakers regularly produce these plosives with prevoicing so that plosive voicing is not only a historical property of Afrikaans. On the other hand, many, especially younger, speakers rarely produce historically voiced plosives with prevoicing, so it is unclear whether these plosives still have the phonological representation of "voiced" for these speakers.

* Corresponding author.
E-mail address: coetzee@umich.edu (A.W. Coetzee).

https://doi.org/10.1016/j.wocn.2017.09.009
0095-4470/

about the relation between individuals and the speech community to which they belong in terms of variation and sound change.

1.1. Cue weighting

The multiple acoustic properties associated with a given speech contrast have often been shown to enter into a trading relation such that stronger information for one property can offset weaker information for another, both in production and perception. Of particular interest here is the relative weighting of VOT and f0 as information for voicing contrasts. These weights have been shown to vary for speakers of different languages (e.g., Llanos, Dmitrieva, Shultz, & Francis, 2013), first and second language speakers (e.g., Schertz, Cho, Lotto, & Warner, 2015), and speakers of the same first language (e.g., Massaro & Cohen, 1976, 1977; Shultz, Francis, & Llanos, 2012). In this study, we ask whether VOT and f0 weights differ for older and younger speakers within a speech community. Finding a generational difference within the same speech community could be indicative of an ongoing change in the community through reweighting of the relevant acoustic properties.

Previous research has documented that, relative to phonologically voiced plosives, phonologically voiceless plosives have later onset of voicing and are followed by vowels with higher f0 (at least for the early portion of the vowel). In English, for example, phonologically voiceless plosives are realized as aspirated. The voicelessness of these plosives is hence cued by long VOT and high f0. If these two properties are in a trading relation, longer VOTs for voiceless plosives should be associated with lower f0 values, a pattern that was documented for American English by Shultz et al. (2012) and Dmitrieva, Llanos, Shultz, and Francis (2015). Similar results are also reported for perception, where studies have shown that, when post-plosive f0 is relatively high, English listeners require shorter VOTs to identify the plosive as voiceless (Abramson & Lisker, 1985; Kong & Edwards, 2016; Llanos et al., 2013; Pearce, 2009; Whalen, Abramson, Lisker, & Mody, 1993).

In so-called true voicing languages, such as French, Spanish and Italian, the phonological voicing contrast is realized as a difference between prevoiced and voiceless unaspirated plosives. In these languages, plosive voicing is therefore indicated by both voicing lead (negative VOT) and low f0 (with longer voicing lead and lower f0 being more prototypically voiced). A trading relation between these two cues would be realized as a long voicing lead patterning with higher f0. However, the evidence for a trading relation between VOT and f0 is less clear for prevoiced than for aspirated plosives. In a series of studies investigating the interaction between VOT and f0 in French and Italian, Kirby and Ladd (2015, 2016) looked for correlations between the duration of voicing lead and f0 of the following vowel, but found inconclusive evidence. Specifically, their 2015 study showed a positive correlation between voicing lead and f0 (i.e., longer negative VOT associated with lower f0), counter to what would be expected if these two cues were in a trading relation. Their 2016 study showed that the relation between f0 and VOT was conditioned by the word's syntactic prominence, and differed between French and Italian.

Compared to languages that contrast unaspirated and aspirated plosives (such as English), relatively little information is available about the perceptual interaction between f0 and VOT in true voicing languages (such as Spanish). Llanos et al. (2013), however, investigated Spanish- and English-speaking listeners' categorization of stimuli that co-varied VOT and vocalic f0. They found evidence of cue trading for both groups of listeners—but only when VOT was in the positive range (i.e., similar to the results mentioned above for English). No evidence for the use of f0 was found in the negative VOT range for either group of listeners.

Like Spanish, French and Italian, Afrikaans is, at least historically, a voicing lead language. Given the inconsistent results in the literature regarding a trading relation between VOT and f0 for this type of language, a clear prediction cannot be made for what to expect in Afrikaans. However, if cue trading were to be observed in Afrikaans production, it should be most clear in the comparison of f0 values for the different phonetic realizations of historically voiced plosives. For those plosives that are realized as voiced (negative VOT values), speakers would not have to rely as much on low f0 as an additional voicing cue, since prevoicing alone is an unambiguous cue for the plosive's voicing status. Devoiced plosives (0 ms or greater VOT values), though, lack the VOT information that signals their contrast with voiceless plosives, making a lower f0 more important. If there is a trading relation for speakers of Afrikaans, lower f0 values should therefore be found after voiceless than voiced realizations of historically voiced plosives.

In terms of perception, a prediction that would be in keeping with the results for Spanish reported by Llanos et al. is that Afrikaans-speaking listeners would ignore f0 in plosives that are realized with prevoicing. However, unlike the phonologically voiced plosives of Spanish, which are consistently realized as voiced, Afrikaans voiced plosives are frequently devoiced. Given this production difference, it is possible that VOT and f0 may also perceptually pattern differently in these two languages.

1.2. From post-consonantal f0 perturbations to tones

The correlation between consonant voicing and f0 of neighboring vowels was noted in the early acoustics literature (Haggard, Ambler, & Callow, 1970; House & Fairbanks, 1953; Lehiste & Peterson, 1961), and has been investigated in detail in many subsequent studies. That f0 is higher following phonologically voiceless consonants and lower following phonologically voiced consonants likely holds for all languages with voicing contrasts, although these consonantally induced f0 perturbations are generally small, and in particular are smaller than the f0 differences typically reported for contrasting tones. Table 1 gives a sample of the values that have been reported in the literature. Because some of these values represent the average f0 across the entire vowel, some the f0 at vowel onset, and others the f0 peak (as indicated in the second column of the table), comparisons should be done with care. As is clear from the table, however, even at vowel onset, where the influence of the neighboring consonant is likely to be largest, the f0 differences between post-voiced and post-voiceless vowels are relatively small, ranging from 8 to 16 Hz. This difference is, in particular, smaller than the typical difference observed between tones in languages with phonemic tonal contrasts. In Yoruba, for instance, the mean f0 difference

Table 1
Consonant voicing effect on following vowels (in Hz).

	Location	Post-voiceless	Post-voiced	Difference
English (House & Fairbanks, 1953:108)	Average	126	122	6
English (Lehiste & Peterson, 1961: Table IV)	Peak	175	163	12
English (based on Hombert, 1978: Fig. 1)	Onset	135	119	16
Swedish (Löfqvist, 1975: Table VI)	Peak	156	146	10
German (Jessen, 1999:100)	Onset	169	161	8
Dutch (van Alphen & Smits, 2004: Fig. 9)	Onset	176	160	16
Italian (based on Kirby & Ladd, 2015: Fig. 4)	Onset	±206	±191	±16
French (based on Kirby & Ladd, 2015: Fig. 4)	Onset	±202	±192	±10

between high and low tones at vowel onset is approximately 25 Hz (based on Hombert, 1977: Fig. 2). For Mandarin Tone 1 (high) and Tone 3 (mid-falling-rising), Liu (2015:43–44) reported a difference of 53 Hz at vowel onset.

In addition to being smaller than the contrastive f0 differences typical of tonal languages, these non-contrastive f0 perturbations are also usually localized to vowel onset rather than continuing throughout the vowel, as is more typical of actual tonal contrasts. Kirby and Ladd (2015), for instance, found that f0 of vowels after voiced and voiceless plosives in both Italian and French differs significantly at vowel onset, but that the difference collapses quickly and is no longer significant by vowel midpoint. Dmitrieva et al. (2015:89) reported that the f0 difference between vowels after voiced and voiceless plosives ceases being significant at approximately 53 and 75 ms into the vowel for Spanish and English, respectively. Because they measured f0 up to 100 ms into the vowel, their results indicate that the f0 difference often stops being significant for Spanish no later than halfway into the vowel and for English no later than three quarters into the vowel. Similarly, Hanson (2009:430) reported, for American English obstruents, that "[w]hile *F*0 values at midvowel were not always observed to be the same across the voiced and unvoiced obstruents, they also did not seem to have systematic differences . . . This leads us to believe that in general, English speakers have the same *F*0 target at midvowel for vowels following voiced and unvoiced obstruents." In contrast to this localized pattern, f0 differences between tones in languages in which tone functions phonemically typically last throughout the vowel. In fact, it has been documented for many tonal languages that the f0 peak associated with a high tone often appears late in the syllable with which it is associated, or even in the following syllable (e.g., Myers, 1999 for Chichewa; Myers, 2001 for Kinyarwanda; Xu, 1999, 2001 for Mandarin), resulting in a rising f0 contour on the vowel associated with a high tone. It can therefore be expected that f0 effects may not only last throughout the vowel in tonal languages, but that they may even increase in magnitude across the vowel. For Yoruba, for instance, Hombert found that the f0 difference between high and low tones increases from approximately 25 Hz at vowel onset to 55 Hz at vowel midpoint (Hombert, 1977), while Liu reported the average difference between Tones 1 and 3 of Mandarin to increase from vowel onset to midpoint from 53 Hz to 96 Hz (Liu, 2015:44). For Seoul Korean, where the consonantally induced f0 perturbations have been phonologized as a tonal contrast, Silva (2006) also reported that the f0 effect is not only present throughout the post-plosive vowel but can even carry over onto the vowel of the following syllable.

Although the effect of consonantal voicing on the f0 of following vowels is well established in the literature, the effect's precise source is subject to debate. Both perception- and production-driven sources have been proposed. On the perceptual side, Kingston and Diehl (1994; see also Kingston, 2005, 2011) argued that the f0 effect does not have an automatic physiological source; rather, speakers actively induce these differences in order to maximize the contrast between plosives that function phonologically as voiced vs. voiceless. Being predicated on a principle of contrast enhancement, this explanation could be considered phonological. On the production side, Hombert, Ohala, & Ewan (1979; see also Abramson, 1975; Ohala, 1973), for instance, suggested that the higher transglottal flow associated with voiceless plosives results in faster vocal fold vibration after voiceless than voiced consonants. Other sources have ascribed the f0 effect to less horizontal vocal fold tension for voiced than voiceless plosives (Ewan & Krones, 1974; Halle & Stevens, 1971; Honda, Hirai, Masaki, & Shimada, 1999; Löfqvist, Baer, McGarr, & Story,1989; Westbury, 1983; etc.). Because these production-oriented explanations assume the f0 effect to be an automatic consequence of physiological conditions, these explanations can be considered phonetic rather than phonological.[2]

Of course, more than one factor may contribute to these small consonantally induced f0 effects. (See Kingston & Diehl, 1994 and Dmitrieva et al., 2015 for more detailed discussion of the extensive relevant literature.) Of importance to the current study is that, whatever their original source, these effects can become exaggerated and have the potential to develop into tonal contrasts. It has been claimed, for instance, that tonal contrasts were introduced via this route into Chamic (Kingston, 2011; Maddieson & Pang, 1993), Vietnamese (Thurgood, 2002), Tibeto-Burman (Matisoff, 1973), Mon-Khmer (Svantesson & House, 2006), the Chadic language Kera (Pearce, 2005, 2009), and Seoul Korean (e.g., Jun, 1996; Kang, 2014; Kang & Han, 2013; Silva, 2006). A prerequisite for the development of these small f0 effects into a full tonal contrast is that the differences be detectable. Perceptual evidence from listeners whose languages lack clear tonal contrasts and have relatively small f0 differences points to the use of f0 as a secondary cue, that is, as information on which listeners rely when the primary VOT cue is ambiguous. Abramson and Lisker (1985), for instance, found that English

[2] The trading relation described in Section 1.1, in which longer (positive) VOTs are associated with lower f0, might appear be in conflict with proposed phonetic sources. Trading relations, though, function as a cue enhancement, when information along one dimension is relatively weak, and can thus be distinct from an original phonetic source. (See also discussion of Korean later in this section.)

listeners use f0 to differentiate plosives in the ambiguous VOT region, but that VOT values at the unambiguous endpoints override f0 differences, i.e., listeners classify a token with long VOT as voiceless even if the vowel following it has a lower f0 (for similar results see also Haggard et al., 1970; Llanos et al., 2013; Pearce, 2005). However, even when VOT values are unambiguous, listeners are sensitive to small f0 differences, as indicated by the influence of these differences on English listeners' reaction times when making voicing judgments (Whalen et al., 1993).

There are multiple possible routes by which these consonantally induced f0 perturbations can develop into independent tonal contrasts. Ohala (1981, 1993) has proposed that exaggeration of a contextual phonetic variant might occur when a listener, who would normally be expected to adjust or compensate for predictable coarticulatory variants, instead undercompensates (possibly due to failure to detect the coarticulatory source) and interprets the phonetic variant as inherent to the target vowel or consonant rather than as the result of contextual influences. Beddor (2009) has suggested that, because multiple interpretations are fully consistent with coarticulated input, some listeners may assign heavier weight to the coarticulatory effects than to their source. Under both views (see also Harrington, Kleber, & Reubold, 2008), the source of what was originally a contextual variant (e.g., a voiced or voiceless plosive in the case of f0 perturbations) would no longer be primary for these listener-speakers. As that variant becomes enhanced, one possibility is for the source to nonetheless be retained by (some) speaker-listeners, creating a situation in which two robust cues for the original contrast are perceived and produced. Alternatively, some members of the speech community may not reliably perceive or produce the original source (an especially likely outcome in Ohala's approach), in which case the historical contrast would be marked by only one property, different from the original. Ohala (1993:247) suggested that although these two patterns may sometimes represent two sequential stages in a sound change (see also Stage III and Stage V in the sound change model of Maran, 1973), it is also possible for both the exaggeration of a contextual variant and the loss of the contextual source to happen simultaneously. Since sequentiality is not necessary, these two scenarios will be referred to below as Pattern I and II, rather than Stage I and II.

The development of consonantal f0 perturbations into tonal contrasts has been described in detail for Seoul Korean. Some of the results reported in the literature are consistent with both the source and the exaggerated contextual effect being present in the newly developed system (Pattern I), while others are more consistent with a situation where exaggeration of the contextual f0 effects goes hand-in-hand with reduction or loss of the consonantal VOT differences (Pattern II). In Seoul Korean, the contrast between lax and aspirated plosives that used to be conveyed primarily by VOT (with longer VOT for aspirated plosives) has been replaced for younger speakers by an f0 contrast (these speakers no longer differentiate the two plosive categories in terms of VOT, but rather by low f0 after historically lax plosives and high f0 after historically aspirated plosives). Silva (2006) reported results consistent with older speakers having a Pattern I system, and younger speakers a Pattern II system: f0 differences on vowels after historically aspirated and lax plosives are large and consistent for speakers born between 1940 and 1980, while VOT differences between aspirated and lax plosives steadily decrease by birth year. The older speakers in his study therefore produce the contrast with both the original VOT difference and an exaggerated f0 difference, while the younger speakers have lost the original source of the effect (the VOT contrast) and now rely on f0 only. Kang and Han (2013) report similar results in a longitudinal study of a single male speaker of Seoul Korean who relied primarily on a VOT difference to cue the contrast between aspirated and lax plosives when originally recorded in 1935 at 11 years of age. In a second recording done 70 years later, the VOT difference between these two categories was still robustly present, and did not differ significantly from the earlier recording, but the f0 difference on post-plosive vowels was significantly larger than in the earlier recording. In 2005, this speaker therefore shows a pattern consistent with Pattern I, with both the VOT and f0 cues robustly present in his speech.

Other studies of Seoul Korean, however, have found evidence that the VOT contrast between lax and aspirated plosives is reduced in magnitude as the f0 contrast on vowels following these plosives increases, more in line with a cue reweighting situation in which enhancement of one cue is accompanied by de-emphasis of another (see Section 1.1). This is representative of a change in which exaggeration of a contextual property goes hand-in-hand with loss of the original source of the cue (along the lines of Pattern II). Kang (2014; see also Bang, Sonderegger, Kang, Clayards, & Yoon, 2015), for instance, conducted a corpus study of the speech of Korean speakers born between 1930 and 1980, and found that the VOT contrast between aspirated and lax plosives gradually decreases in magnitude with later birth year, while the f0 difference on vowels following these plosives gradually increases at the same time. (See also the computational model of the Seoul Korean change by Kirby (2013), who assumed a situation in which increase in the f0 contrast is accompanied by reduction in the VOT contrast.) The partially incompatible results for Seoul Korean reported in the literature may reflect a situation in which the change can happen by either of the two routes, with the production of different members of the speech community reflecting different routes.

Turning to the Afrikaans situation that is the focus of the current study, we have already indicated (anticipating the results) that younger speakers produce more devoiced realizations of historically voiced plosives than do older speakers. Thus, one possible pattern for f0 and VOT may be that especially large f0 differences for vowels following historically voiced and voiceless plosives go hand-in-hand with the loss of prevoicing (Pattern II). In this case, we should find that f0 is especially low after devoiced realizations of historically voiced plosives (see also Section 1.1) and therefore that the contrast in vocalic f0 associated with voiced and voiceless plosives is larger for younger than older speakers. On the other hand, if Afrikaans is more representative of Pattern I, the f0 contrast would have been exaggerated initially without simultaneous loss of the VOT contrast. In this case, f0 after voiced realizations of historically voiced plosives should not differ from f0 after voiceless realizations of such plosives. Similarly, the f0 difference between historically voiced and voiceless plosives

should not be larger for younger than older speakers. Both of these patterns have been described for Seoul Korean and both are in principle possible for Afrikaans as well. As will be shown in the results reported below, however, Afrikaans is more consistent with Pattern I, with the possibility that it is currently developing into a Pattern II type system.

1.3. The history of plosive voicing in Afrikaans

Except in marginal loan words, Afrikaans lacks /g/, so that plosives contrast historically for voicing only at the labial and alveolar places. Although, as shown below, historically voiced /b d/ are often realized as voiceless unaspirated [p t] in Afrikaans, especially by younger speakers, Afrikaans most likely had a robust contrast between prevoiced and voiceless unaspirated plosives until quite recently. Evidence for the existence of such a contrast in Afrikaans comes from two sources. Phonetic descriptions of Afrikaans dating back as far as 90 years (Le Roux & Pienaar, 1927) to more recent descriptions from approximately 40 years ago (van Wyk, 1977; Wissing, 1982) describe Afrikaans as having such a contrast. Dutch, the direct primary ancestor of Afrikaans, also realizes the phonological voicing contrast in plosives in this manner (Collins & Mees, 1982, 2003; Gussenhoven, 1999; Lisker & Abramson, 1964; Verhoeven, 2005). There is some evidence that Dutch may be in the early stages of undergoing devoicing with possible concomitant exaggeration of the f0 differences. Simon (2009:385), for instance, reported a devoicing rate of 7%, van Alphen (2004:19) of 14.5%, and van Alphen and Smits (2004:462) of 25%. This change is recent enough, however, that as recently as 2015 Pinget still called it an “incipient change” (Pinget, 2015:30). It therefore seems reasonable that the mid-seventeenth century Dutch that formed the input to what was to become Afrikaans over the ensuing three and half centuries would have had a robust contrast between prevoiced and voiceless unaspirated plosives.

No data are available about the influence of plosive voicing on the f0 of following vowels in earlier stages of Afrikaans history. However, based on the research reviewed in Section 1.2, it can be expected that Afrikaans would have had the same low-level phonetic perturbations in f0 caused by preceding consonant voicing as have been documented for many other languages with a plosive voicing contrast. It would also not be unreasonable to assume that, in Afrikaans, f0 at the onset of vowels after voiced plosives would have been slightly lower than after voiceless unaspirated plosives. The current situation in Dutch can again serve as corroborating evidence. Slis and Cohen (1969) reported f0 after voiceless plosives in Dutch to be 6 Hz higher than that after voiced plosives. More recently, van Alphen and Smits (2004) reported a difference of 16 Hz at vowel onset between post-voiced and post-voiceless consonant contexts.

1.4. Individual and community-level variation

If a specific contrast is marked by more than one phonetic property, it can be expected that there will be variation in the speech community in terms of the weight that individual members of the community will assign to the various properties, in both production and perception. That individuals differ in the weights they assign to the contribution of VOT and f0 to the production and perception of voicing contrasts is already well established (e.g., Kong & Edwards, 2016; Schertz et al., 2015; Shultz et al., 2012; Whalen et al., 1993). An important question for theories of sound change is whether individuals who produce what might be viewed as the innovative weighting of these properties—that is, who produce the innovative production norm—are similarly innovative perceivers. Although a link between the production and perception repertoires of individuals is assumed by many approaches to sound change (e.g., Beddor, 2009; Harrington et al., 2008; Lindblom, Guion, Hura, Moon, & Willerman, 1995; Ohala, 1981; Yu, 2013), the hypothesized link has been somewhat difficult to establish, including for the relative articulatory and perceptual weighting of VOT and f0 (Schertz et al., 2015; Shultz et al., 2012).

In the Afrikaans situation under investigation here, one possibility is that the innovative speakers—that is, speakers who produce high f0 and no prevoicing for historically voiced plosives—will rely extensively on f0 to perceptually differentiate historically voiced and voiceless plosives. Along similar lines, those who are more likely to produce historically voiced plosives with prevoicing may perceptually rely more on prevoicing. Depending on whether the f0 and VOT cues in Afrikaans are independent from each other or in a trading relation, the relation between perception and production—if it exists—may hold for f0 only, VOT only, or for both of these properties.

1.5. Research questions and hypotheses

This study investigates the realization of the historical plosive voicing contrast in the production and perception of speakers of contemporary Afrikaans. Given that Afrikaans has been described as contrasting prevoiced and voiceless unaspirated plosives, we expect to find that speakers will produce the historically voiced plosives at least sometimes with prevoicing. If Afrikaans is currently undergoing a change whereby the voicing contrast is being lost, we hypothesize that both older and younger speakers will sometimes produce /b d/ as voiceless unaspirated, although younger speakers should do so more often. Moreover, if post-plosive f0 effects have been phonologized, such that they now function as information for the historical voicing contrast, f0 should be low after both phonetically voiced and voiceless realizations of historically voiced /b d/. If phonologized, f0 differences should also be larger in Afrikaans than in languages such as English and Dutch where f0 serves only as secondary cue of the plosive voicing contrast.

In perception, we expect, in accordance with the literature reviewed in Sections 1.1 and 1.2, that all listeners will rely to some extent on both VOT and f0 as information for differentiating historically voiced and voiceless plosives. We predict as well a perceptual trading relation in which reliance on f0 increases as information for plosive voicing decreases. Also, if Afrikaans is currently undergoing a change whereby the voicing (VOT) contrast is being replaced by a tonal (f0) contrast, we would expect younger listeners to perceptually rely less on VOT and more on f0 than older listeners. In the next two sections, we present the results of production and perception experiments designed to investigate these hypotheses

2. Production experiment

The purpose of the production experiment is to determine, for word-initial plosives produced by older and younger speakers of Afrikaans, VOT and post-plosive f0 characteristics. Assuming the predicted pattern of greater devoicing of phonological /b d/ by younger speakers, f0 should be especially low after the devoiced realizations if younger speakers are maximizing the phonological voiced-voiceless contrast. Alternatively, if the situation in Afrikaans is similar to that found in some studies of Seoul Korean, in which the magnitude of the f0 effect is independent of plosive VOT, then post-plosive f0 should be similar for older and younger speakers' productions.

2.1. Methods

2.1.1. Participants

The participants were female native speakers of Afrikaans, recruited from among the faculty, staff and students of the North-West University in Potchefstroom, South Africa. Limiting participation to only female (or only male) speakers facilitates between-speaker comparison of f0. Additionally, given that male and female speakers often do not participate in an ongoing sound change at the same rate, potential complexity is avoided by focusing on female speakers only. A more complete understanding of plosive voicing system of Afrikaans, however, would require also investigating the production and perception patterns of male speakers. In order to minimize potential dialectal differences among the participants, participants were all of Caucasian descent and spoke the northern variety of so-called "Standard Afrikaans". Participants were recruited in two age groups. The younger group consisted of 10 speakers with ages ranging from 21 to 23 years. Due to technical problems, the recordings from two additional younger speakers were of too poor quality for analysis, and were therefore excluded from the study. Thirteen older speakers with an average age of 60 years were recruited (one speaker was 41, four were in their fifties, and the remaining eight were in their sixties). Participants all self-reported having normal speech and hearing and were paid for their participation.

2.1.2. Stimuli and recording procedures

Sixty target words beginning with an initial labial (/p b/) or alveolar (/t d/) plosive were selected. Within each place of articulation, 15 words started with a voiced plosive (/b/ or /d/) and 15 with a voiceless plosive (/p/ or /t/). The voiced and voiceless groups within each place of articulation were balanced in terms of the identity of the vowel following the initial consonant. In addition to the target words, 40 fillers were included, half of which started with a voiceless consonant (/f k/), and half with a voiced consonant (/m n v/). Although the nasal-initial words were originally intended as fillers, f0 following the nasals was also analyzed in order to assess the patterns observed for the plosives relative to the rest of the consonantal voicing system of Afrikaans. All words, tokens and fillers, were either monosyllabic or disyllabic with initial stress. A complete list is given in Appendix A.

Stimuli were presented as a list of words, and were differently randomized for each participant. The first participant read the full stimulus list 10 times, while all remaining participants read the list eight times. Recordings were made with an AKG C1000S microphone at a sampling rate of 44.1 kHz in a sound-attenuated room at the North-West University in Potchefstroom, South Africa, using *Praat* (Boersma & Weenink, 2013). A recording session lasted about 10 min.

2.1.3. Analyses

Of the eight to 10 repetitions of every word that were recorded, four were included in the analysis. Since the target stimuli consisted of 15 words each per relevant plosive category and 10 nasal-initial words, including four repetitions in the analysis allowed 40–60 tokens per category per speaker to be analyzed, providing sufficient power to detect any meaningful differences between token types. The total possible number of analyzed voiced and voiceless plosive tokens was therefore 5520 (60 words $\times$ 4 repetitions $\times$ 23 participants), and the total possible number of nasal tokens was 920 (10 words $\times$ 4 repetitions $\times$ 23 participants). To facilitate accurate measurement of fundamental frequency, the first four repetitions by each speaker produced in modal, non-creaky voice were selected. Furthermore, in order to ensure accurate measurement of f0, all token words were visually inspected, using *Praat*'s "Show pitch" function. Tokens for which *Praat*'s pitch tracker resulted in inaccurate measurements because of either pitch halving or pitch doubling were excluded. Due to speaker error and to the exclusion of tokens for creakiness or inaccurate pitch tracking, the number of analyzed plosive-initial tokens fell four short of the total possible number (two tokens short for one speaker, and one token each for two additional speakers).

Fig. 1 shows the intervals that were marked for plosive-initial tokens. Interval A is voice onset time. Fig. 1a is representative of productions with phonetically voiceless initial plosives (i.e., tokens that start with phonological /p t/ and voiceless productions of tokens that start with phonological /b d/). For such tokens, (positive) VOT was measured from onset of oral closure release to onset of voicing of the following vowel (identified as the first clear vocal pulse). This interval therefore encompasses, for example, both the "REL" and "ASP" intervals of Abramson and Whalen's (2017) labeling procedure.

For tokens realized with some voicing, as in Fig. 1b, (negative) VOT was measured from the onset of voicing during plosive closure to closure release, with voicing onset identified as the moment when the first sign of a voice pulse was observed in the waveform, associated with the low-frequency voicing bar in the spectrogram. Because stimuli were produced in utterance-initial position, it is not possible to determine whether voicing began at the onset of closure. As a result, it is not possible to determine whether these plosives fall, for example, in the category of Davidson's partial negative VOT pattern, in which voicing begins *after* closure onset (Davidson, 2016). However, once voicing did begin, it lasted up to the moment of release in the overwhelming majority of tokens (with Fig. 1b being a representative example). In a small number of tokens, voicing started early in the closure and then ceased before the closure was released. Such tokens therefore had both a voiced and voiceless closure interval, corresponding to the VDCLO and VLCLO intervals in the coding scheme of

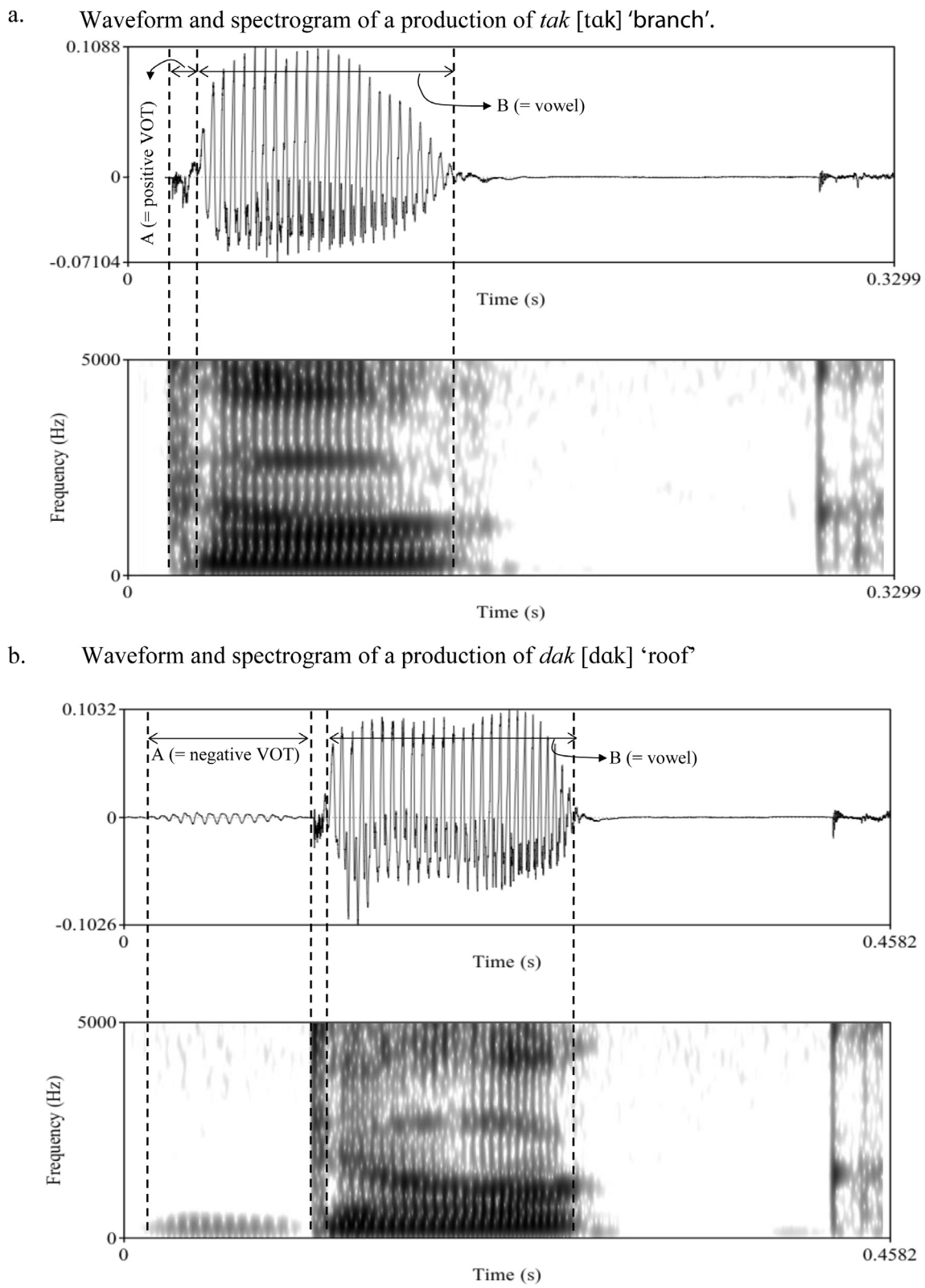


Fig. 1. Sample segmentation intervals for voiceless (a) and voiced (b) plosive onsets.

Abramson and Whalen (2017). Any tokens with closure voicing, whether full or partial, were therefore categorized as phonetically voiced.

Interval B in Fig. 1a and b is the vowel. For all plosive-initial words, the vowel was measured from the onset of the first clear voice pulse associated with the vowel to the end of the last vocalic pulse. Vowel offset was identified as the last voice pulse associated with the vowel. For the small number of tokens in which the vowel was followed by an /l/, identifying this final vocalic pulse was sometimes more difficult; here we

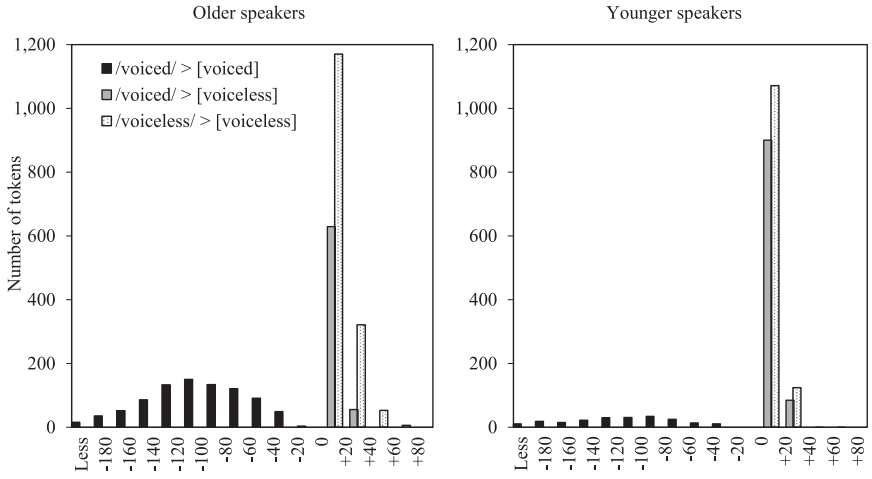


Fig. 2. Histograms representing the distribution of VOTs for productions of /b d p t/ for older (left) and younger (right) speakers. Each 20 ms bin includes all tokens that have a VOT value equal to or lower than the bin label (and higher than the label of the immediately lower bin). Black bars represent voiced productions of /b d/, gray bars devoiced productions of /b d/, and dotted white bars /p t/.

looked for a change in the shape of the waveform, associated with a change in the spectrogram, and also relied on auditory judgments. For the nasal-initial words, only the vowel portion was marked.

To measure vocalic f0, the vowel (Interval B) was divided into 20 intervals of equal duration, and f0 was measured at the boundary between each of these 20 intervals (i.e., at 19 locations spaced equally across the vowel). F0 was calculated using *Praat*'s built-in "To Pitch…" autocorrelation-based pitch analysis method (described in Boersma, 1993), constraining the search to f0 values in the 100–500 Hz range. All other algorithm-specific settings (e.g., silence and voicing thresholds, octave and jump costs) used the algorithm's standard settings (described in *Praat*'s "To Pitch…" help page). This method produces near-optimal results while maintaining congruence with similar analyses in the literature.

2.2. Results

2.2.1. Voice onset time of /b d p t/

Each realization of words starting on /b d/ was classified as phonetically [voiced] if there was any voicing during the closure phase of the plosive, and as [voiceless] otherwise. This binary classification was straightforward because, as shown in the histograms of VOT values for /b d/ and /p t/ productions in Fig. 2, tokens that were classified as [voiced] and [voiceless] are clearly separated from each other. Tokens classified as [voiced] have negative VOTs, with the region between −20 ms and 0 ms being virtually unpopulated by any tokens. Most tokens classified as [voiceless] have VOTs between 0 ms and +20 ms. Fig. 2 also shows that the range of VOT values for devoiced /b d/ tokens (gray bars) is a subset of the range of the VOT values for /p t/ (dotted white bars). For both the /p t/ and devoiced /b d/ categories, most VOT values fall in the +20 ms and +40 ms bins, with older speakers having a few productions of /p t/ with VOT in the +60 ms bin.[3]

[3] Voicing classification is based solely on VOT values. Our acoustic data do not address whether voiceless /p t/ and tokens we label as "devoiced" /b d/ might differ in laryngeal abduction despite similar VOTs.

In order to assess whether the devoiced productions of /b d/ differ from /p t/ productions in VOT, a linear mixed-effect model was fit to the VOT values of these tokens using the *lmer* function from the *lme4* package (Bates, Maechler, & Bolker, 2013) in R (R Core Team, 2013), with *p*-values estimated using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2016). The model included as fixed effects Age (Older vs. Younger), Phonological Voicing (/voiced/ vs. /voiceless/), Place of Articulation (Labial vs. Alveolar), and all interactions between these factors. The random effect structure included random intercepts for Speaker and Word, and random slopes for Speaker by each of Phonological Voicing and Place of Articulation. The reference values of the fixed effects were "/voiced/" for Phonological Voicing, "Older" for Age, and "Alveolar" for Place of Articulation. The results of this model are reported in Table 2. Only the interaction term that was significant is included in the table.

The model showed a significant effect of Phonological Voicing, with /voiceless/ productions having, on average, slightly higher VOT values than devoiced /voiced/ productions. Inspection of Fig. 2 shows that the reason for this is not that /voiced/ and /voiceless/ tokens occupy a different region of the VOT range. Rather, the VOT range of the /voiced/ tokens overlaps with that of the /voiceless/ tokens. There are, however, some /voiceless/ tokens that have higher VOT values than any /voiced/ tokens, especially for older speakers' productions. To investigate the interaction between Age and Phonological Voicing, mean VOT values were calculated for the voicing conditions separately for the two age groups, showing that older speakers on average maintained a difference between /voiced/ and /voiceless/ tokens (12.0 ms vs. 16.2 ms), while younger speakers did not (12.2 ms vs. 12.6 ms).

2.2.2. Prevalence of /b d/ devoicing

Younger speakers overall had a higher devoicing rate than older speakers: 83% and 82% for /b/ and /d/, respectively, for younger speakers compared to 44% for both /b/ and /d/ for older speakers. There was also virtually no difference in devoicing rates between /b/ and /d/ for either age group. In

Table 2
Results of model testing for the effect of phonological voicing on voice onset time.

Predictor	β	SE(β)	*df*	*t*	*p*
Intercept	11.77	1.32	73	8.91	<0.001
Age	1.12	1.26	21	0.89	0.38
Phonological Voicing	5.59	1.67	78	3.36	0.001
Place of Articulation	−0.43	1.68	82	−0.25	0.80
Age * Phonological Voicing	−4.44	1.21	23	−3.69	0.001

Table 3
Results of logistic model fit to the devoicing data.

Predictor	β	SE(β)	*z*	*p*
Intercept	−0.28	0.41	−0.68	0.50
Age	2.33	0.57	4.12	<0.001
Place of Articulation	−0.01	0.33	−0.02	0.98
Age * Place of Articulation	0.43	0.38	1.15	0.25

order to evaluate the different contributions of age and place of articulation to the rate of devoicing, a generalized mixed-effect model with a logistic link function was fitted to the devoicing data, using the *glmer* function from the *lme4* package (Bates et al., 2013) in R (R Core & Team, 2013). The dependent variable was Devoicing, with devoiced productions coded as "1" and productions with voicing as "0". The fixed effects in the model included Age (Older vs. Younger), Place of Articulation (Labial vs. Alveolar) and their interaction. The model also contained random intercepts for Word and Speaker, as well as a random slope for Speaker by Place of Articulation. The results of the model are summarized in Table 3. The reference level for the effect of Age was set to "Older" and for Place of Articulation to "Labial". The only significant effect that was found is that of Age, confirming that younger speakers devoiced more than older speakers.

2.2.3. F0 profiles of post-plosive vowels

Fig. 3 represents the f0 of vowels following /b d p t/, plotted separately for each of the three realization conditions, age of the speaker and place of articulation. These plots show a clear f0 difference between vowels after phonologically voiceless (/voiceless/ → [voiceless]) and phonologically voiced consonants (/voiced/ → [voiced] and /voiced/ → [voiceless]). However, f0 of vowels after phonologically voiced consonants does not appear to differ—at least not substantially so—depending on whether the consonant is realized with VOTs characteristic of [voiced] or [voiceless] plosives. Although Fig. 3 shows that younger speakers have an f0 range that is about 25 Hz higher than that of older speakers, the shape of the f0

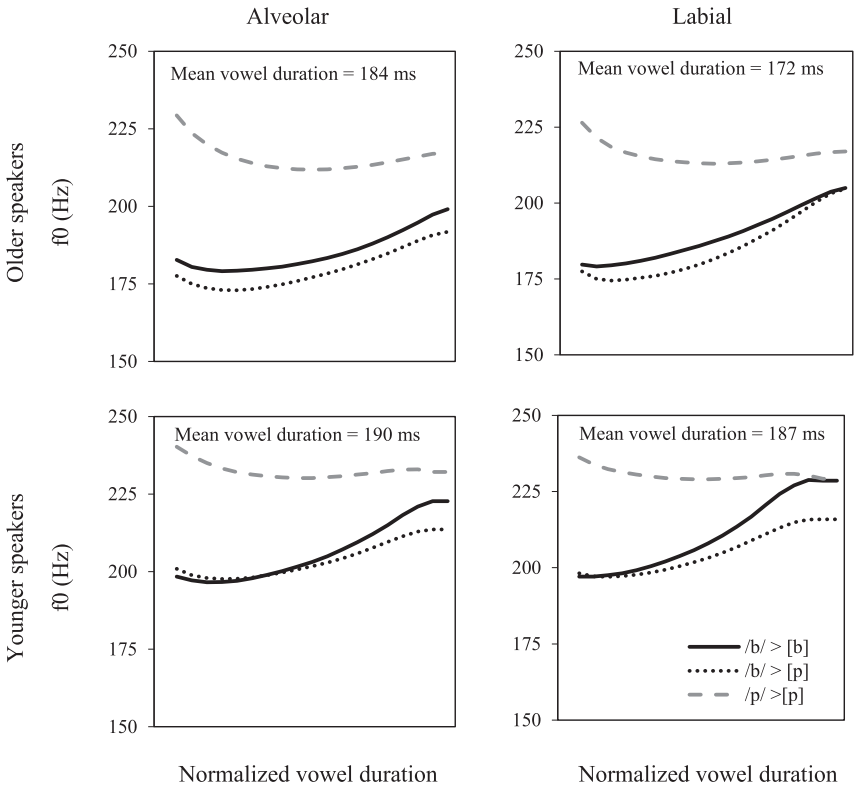


Fig. 3. F0 profiles of vowels after /b d p t/ based on 19 f0 measures taken at equally spaced intervals across the vowel. (*x*-axis represents the time-normalized duration of the vowel.) Left and right panels: post-alveolar and post-labial f0 profiles, respectively. Top and bottom panels: older and younger speakers' productions. Average vowel duration is given in each panel.

profiles is similar for the two age groups. It also appears that, for the younger speakers, f0 after phonetically voiced plosives (/voiced/ → [voiced]) rises up to the level typically associated with historically voiceless plosives (/voiceless/ → [voiceless]). (We return to this issue later in this section.) Fig. 3 represents the f0 profiles in normalized time (on the *x*-axis). Although it is therefore not possible to determine from this figure the absolute temporal characteristics of f0 profiles, to facilitate this assessment, each panel in Fig. 3 specifies the average duration of the vowel in that condition. As is clear from these values, the f0 effect lasts throughout the vowel (with one exception); moreover, the absolute duration of the effect appears to be longer than the durations reported by Dmitrieva et al. (2015:89) for Spanish (53 ms) and English (75 ms), for instance.

Because mixed-effects regression models of these data failed to converge, and to adequately model the non-linear f0 curves, we opted to analyze the data using a method based on Generalized Estimating Equations (GEE). All statistical calculations in this and the next section were done using Python 3.4.2, NumPy 1.9.1, and Statsmodels 0.6. In order to account for non-linguistically relevant between-speaker differences in f0, f0 effects were allowed to vary overall by speaker, between utterances of the same word for the same speaker, and between utterances of different words for the same speaker. Each of these categories has its own pattern of within-speaker correlation, and represents a more flexible way of handling within-speaker correlations compared to data normalization, or the use of random intercepts in a mixed model. We first modeled the mean structure of the data using an Ordinary Least Squares (OLS) approach, specified in terms of age, condition, and time, fitting separate models for the labial and alveolar conditions. To capture the non-linear effects of time on these factors, all means were modeled using Cubic B-splines (de Boor, 1978:87–106) with four degrees of freedom. The mean structure included all interactions up to the third order between Age (Older vs. Younger), Condition (/voiced/ → [voiced] vs. /voiced/ → [voiceless] vs. /voiceless/ → [voiceless]), as well as the basis functions which specify the curvilinear effects of time, resulting in 30 mean structure parameters in the final model.

To make inferences based on this mean structure, we then generated confidence intervals using GEE. To do this, the covariance structure was estimated from the OLS model residuals and used to define a sampling covariance matrix for the regression parameters that is robust to heteroscedasticity and correlation (Amemiya, 1985; Green & Silverman, 1994). The covariance matrix included both within-subject and between-subject correlations. Within-subject correlations were estimated for each pair of time points (s, t) (where s and t are any two f0 measures taken at one of the 19 time points across a vowel), separately for three settings: (i) s, t are in the same utterance, (ii) s, t are in distinct utterances of the same word, and (iii) s, t are in different words. In addition, between-subject covariances were estimated for all time points s, t in the same word produced by different speakers. Sparse matrix methods were used to make the calculations feasible. From this covariance matrix, we can obtain simultaneous confidence intervals for comparisons between the three Conditions, as presented in the results below.

In order to investigate the difference in the f0 profiles for the different Conditions, we calculated for each age group and each place of articulation three sets of difference scores (relying on the covariance matrix described above): (i) (f0 of /voiced/ → [voiceless]) minus (f0 of /voiced/ → [voiced]) to compare the [voiced] and [voiceless] realizations of /b d/ (the solid and stippled black lines in each of the images in Fig. 3); (ii) (f0 of /voiceless/ → [voiceless]) minus (f0 of /voice d/ → [voiced]) to compare the [voiced] realizations of /b d/ with the realizations of /p t/ (the solid black and broken gray lines in Fig. 3); (iii) (f0 of /voiceless/ → [voiceless]) minus (f0 of /voiced/ → [voiceless]) to compare the [voiceless] realizations of /b d/ with the realizations of /p t/ (the stippled black and broken gray lines in Fig. 3). For each of these difference scores, a 95% simultaneous confidence interval was obtained at each of the 19 time points across the vowel (again using the covariance matrix described above). The results of these comparisons are represented in Fig. 4 for alveolar-initial words and in Fig. 5 for labial-initial words. We can infer that the two conditions being compared are significantly different from each other at any time point where the confidence interval excludes zero.

Inspection of Figs. 4 and 5 shows that, for both age groups and both places of articulation, the confidence intervals for comparison between [voiced] and [voiceless] productions of phonologically voiced plosives (leftmost panels in the two figures) include the zero line for the complete duration of the vowel. There is therefore no evidence that the different phonetic (VOT) realizations of phonologically voiced plosives correspond to differences in f0 of the following vowel. In comparison, for the other four panels in each figure, the confidence interval excludes the zero line for either the entire duration or the majority of the vowel.

Although [voiceless] productions of phonologically voiced plosives cannot be differentiated based on VOT from the productions of phonologically voiceless plosives (see Section 2.2.1),[4] the contrast between these two kinds of plosives is maintained in the f0 of the following vowel (middle panels of Figs. 4 and 5). As expected, the [voiced] productions of phonologically voiced plosives and the productions of phonologically voiceless plosives differ in both f0 (rightmost panels) and VOT (Section 2.2.1). For the younger speakers, the significant f0 difference between these categories collapses toward the end of the vowel. There are two possible explanations for this effect. First, given the high rate of devoicing for the younger speakers, the /voiced/ → [voiced] category for these speakers is represented by fewer tokens so that the f0 estimates for this category may be less accurate for the younger speakers. Secondly, this may be evidence of a trading relation in the speech of the younger speakers. When the phonologically voiced plosives are produced with closure voicing, the contrast between these plosives and the phonologically voiceless plosives is unambiguously marked on the plosive itself such that there is less need to also cue the contrast with f0.

[4] As reported in Section 2.2.1, for older speakers the average VOT of /voiceless/ plosives was higher than that of [voiceless] productions of /voiced/ plosives. Although those few tokens with the longest VOTs (60 or 80 ms) for these speakers are only /voiceless/, for the majority of tokens produced as [voiceless] the VOT ranges of /voiced/ and /voiceless/ tokens overlap so that these two categories cannot be differentiated from each other in terms of VOT.

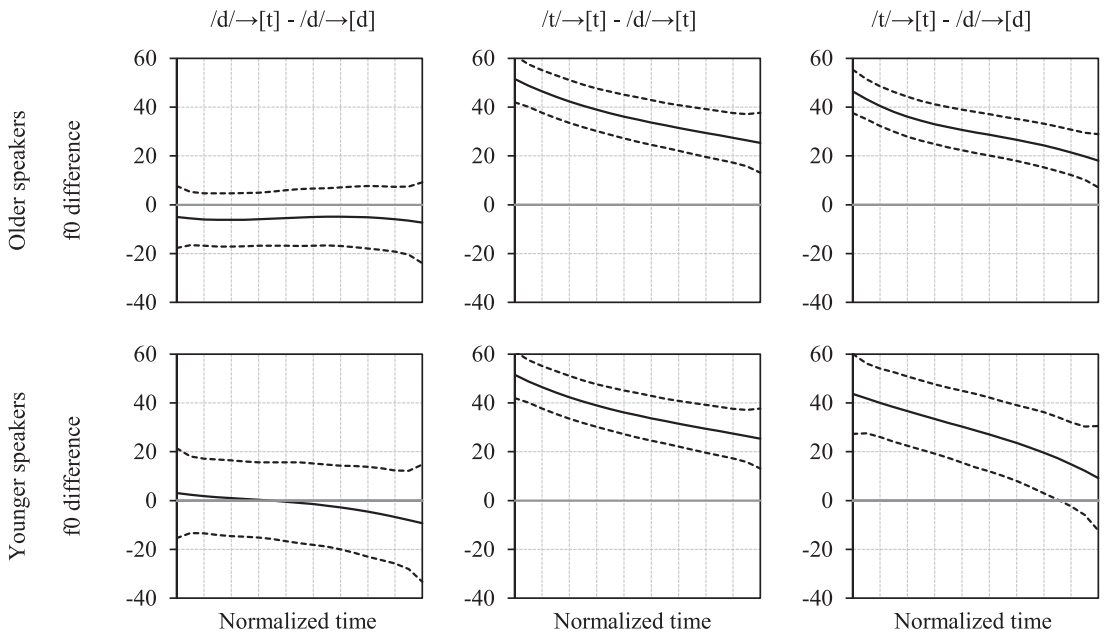


Fig. 4. Comparisons of f0 profiles of the three Conditions for /t d/-initial words. Solid black line: estimated mean f0 difference; broken lines: 95% confidence interval around this mean. *x*-axis: 19 time points across the vowel where f0 was measured; *y*-axis: difference in model predictions for f0.

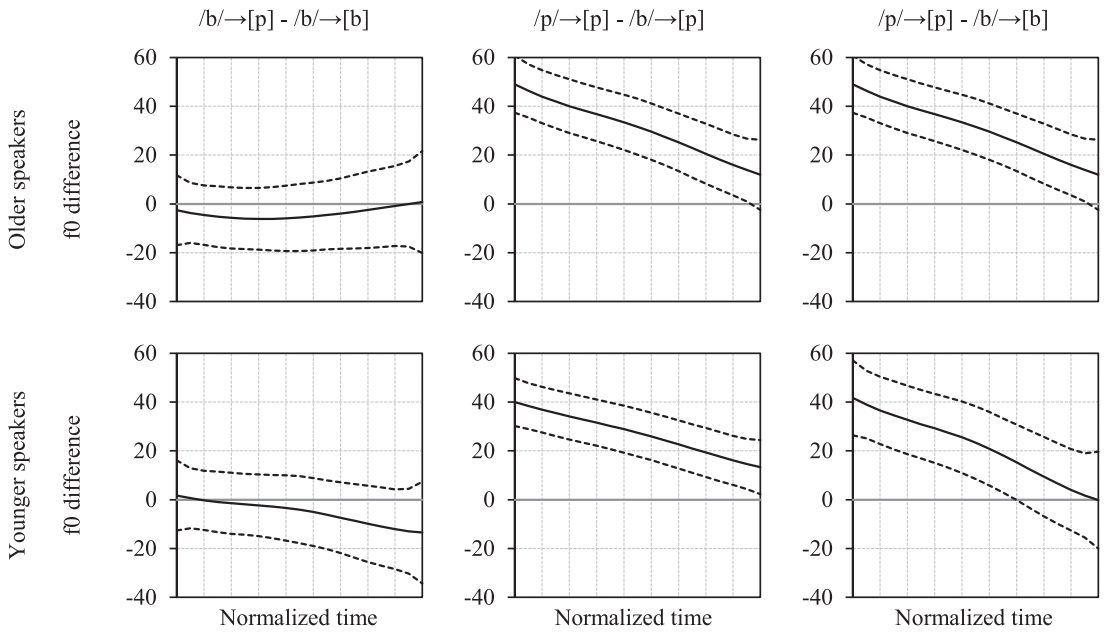


Fig. 5. Comparisons of f0 profiles of the three Conditions for /p b/-initial words. See caption for Fig. 4 for an explanation of the figure.

2.2.4. F0 profiles of vowels after nasals

Hanson (2009) and Kirby and Ladd (2016) argued for the use of sonorant consonants as a baseline comparison for post-plosive f0 because sonorants, which do not participate in a voicing contrast, would not require f0 as a voicing cue. Fig. 6 shows the average f0 profiles of vowels after nasals and phonologically voiced plosives produced by older and younger Afrikaans speakers. Since the f0 profiles after the two possible surface realizations (voiced and voiceless) of phonologically voiced plosives were found not to differ significantly (Section 2.2.3), all phonologically voiced plosives are combined here.

As inspection of Fig. 6 shows, although f0 starts as low after nasals as it does after phonologically voiced plosives, it rises rapidly throughout the vowel. To further investigate the difference in f0 curves after nasals and voiced plosives, the differences between these curves were modeled using the GEE approach described in Section 2.2.3 to compare the f0 profiles after different plosive realizations. A difference score was calculated by subtracting the f0 of vowels following phonologically voiced plosives from the f0 of vowels following nasals. A positive difference indicates higher f0 after nasals than after phonologically voiced plosives. A 95% simultaneous confidence interval was obtained at each of the 19 time points

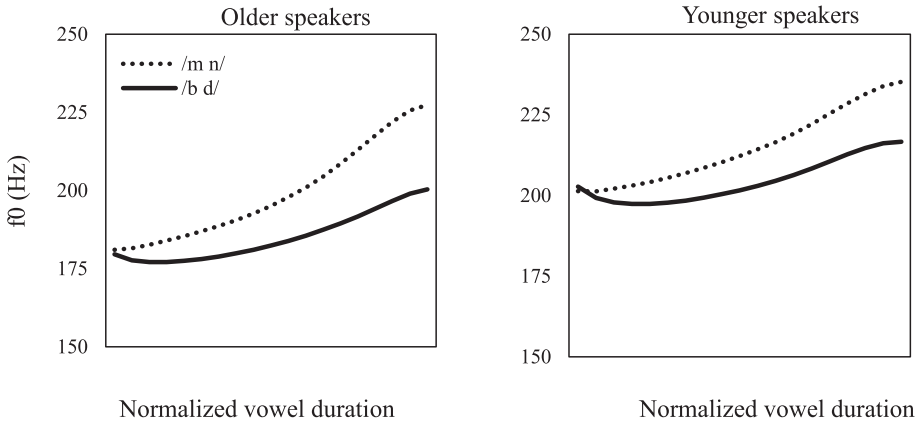


Fig. 6. F0 profiles of vowels after phonologically voiced plosives /b d/ and nasals /m n/ based on the 19 f0 measures taken at equally spaced intervals across the vowel for older (left) and younger (right) speakers' productions. (*x*-axis represents time-normalized vowel duration.)

across the vowel for this difference score. The results of this comparison are represented in Fig. 7. As inspection of this figure shows, the confidence intervals for the difference in f0 profiles includes zero only for approximately the first 10% of the vowel, indicating that vocalic f0 is significantly higher after nasals than phonologically voiced plosives for the majority of the vowel. The plosive-nasal difference over time suggests that speakers are actively maintaining the f0 lowering associated with voiced plosives, which may be indicative of enhancement of the f0 effect in the plosive context in which f0 serves as information for the phonological voicing contrast. This differs from the situation in French and Italian, for which Kirby and Ladd (2016) reported minimal to no difference in the f0 profiles in vowels after nasals and phonologically voiced plosives, and also in English, for which Hanson (2009) reported results similar to those of Kirby and Ladd. Arguably, the different f0 patterns emerge for Afrikaans because the Afrikaans voicing contrast—unlike that of English, French, or Italian—is not consistently indicated by VOT.

3. Perception experiment

The results of the production experiment showed that, although the older and younger speakers did not differ meaningfully in post-plosive f0 patterns (with both groups showing large f0 differences between phonological /b d/ and /p t/), younger speakers were significantly more likely than older speakers to produce devoiced variants of /b d/. The perception experiment assesses these same participants' relative weighting of the acoustic properties of VOT and f0 as information about the contrast between /b d/ and /p t/. If perception parallels production at the age-group level, we expect that the two groups of participants would attend similarly to f0 variation in the stimuli, but that the older participants would attend to voicing information more than the younger participants.

3.1. Methods

3.1.1. Participants

The same 23 Afrikaans speakers from the production experiment also participated in the perception experiment.

3.1.2. Stimuli

Stimuli were continua varying in the VOT and f0 characteristics of labial- and alveolar-initial words. Continua were created by modifying naturally produced tokens of Afrikaans words. Continuum endpoints were members of minimal pairs differing in the phonological voicing of the initial plosive: alveolar *doer* /duːr/ 'over there' and *toer* /tuːr/ 'tour', and labial *bas* /bɑs/ 'tree bark/buttocks' and *pas* /pɑs/ 'fit'. Because the labial- and alveolar-initial continua were created in the same manner, we discuss only the labial continua as an example here.

A female speaker of the same variety of Afrikaans as the participants was recorded reading an Afrikaans word list similar to that used in the production experiment, using the same AKG C1000S microphone. The speaker was of the same age range as the younger participants in the current study, was not a trained phonetician, and was uninformed as to the nature of the research. Although the speaker had a devoicing rate lower that the average for the younger speakers in our study, producing about 20% phonologically voiced plosives as voiceless, that rate is not out of the range observed in the speech community. She produced the post-plosive f0 patterns that were characteristic of those produced by the older and younger participants in the production experiment. Recordings were made with *Praat* (Boersma & Weenink, 2013) in a sound-attenuated room at the North-West University in South Africa at a sampling rate of 44.1 kHz. Several repetitions of the endpoints *bas* and *pas* were recorded and, for each word, a token produced with modal, non-creaky phonation was chosen. Spectrograms of the selected endpoint tokens are included in Appendix B.

The selected *bas* token was produced without closure voicing (i.e., /b/ was produced as [p]). The f0 values of the vowels in the selected *bas* and *pas* tokens were typical of the values observed in the production experiment: at vowel onset, f0 of *bas* was about 50 Hz lower than that of *pas*. The recordings of the two endpoints were scaled to have an average intensity of 70 dB using the *Scale intensity*... command in *Praat*. The duration of the vowels in *bas* and *pas* was also scaled to be equal to the average duration of the two vowels using the Pitch Synchronous Overlap and Add (PSOLA) algorithm in *Praat*. The f0 values of the vowels from *bas* and *pas* were then extracted at 20 equally spaced intervals across the vowels.

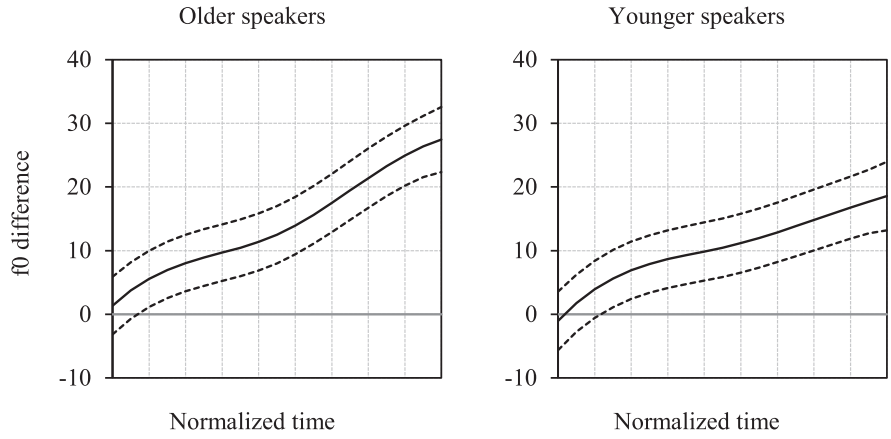


Fig. 7. Comparisons of f0 profiles of vowels after phonologically voiced plosives /b d/ and nasals /m n/ produced by older (left) and younger (right) speakers. Solid black line: estimated mean f0 difference; broken lines: 95% confidence interval around this mean. *x*-axis:19 time points across the vowel where f0 was measured; *y*-axis: difference in model predictions for f0.

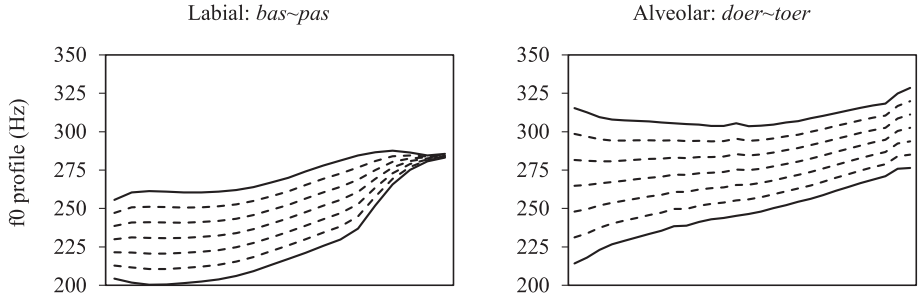


Fig. 8. F0 profiles of the stimuli in the seven-step continua for the perception experiment. In each panel, top solid line: f0 of original voiceless endpoint (*pas* or *toer*); bottom solid line: f0 of original voiced endpoint (*bas* or *doer*). Broken lines: f0 of the five intermediate steps on each continuum.

At each of these 20 time points, five values equally spaced between the value from *bas* and *pas* were calculated. Connecting the 20 points for each of these intermediate values resulted in five f0 profiles intermediate between the original f0 profiles of *bas* and *pas*. These values were used to create a seven-step f0 continuum ranging from the f0 profile of the original *bas* to that of the original *pas*. Stimuli on the continuum were synthesized using the PSOLA algorithm in *Praat.* One set of stimuli was synthesized using the original *bas* as the basis, and a second set using the original *pas*. These two continua therefore had identical f0 profiles, intensity and vowel duration. However, since the other acoustic properties of the original *bas* and *pas* were not affected by the synthesis procedure, the tokens on each of the two continua differed along the same acoustic dimensions as the original *bas* and *pas* recordings. Fig. 8 gives the f0 profiles of the stimuli for the alveolar and labial continua. The line corresponding to the highest f0 is identical to the actual f0 profile of the original *pas* token, and the line corresponding to the lowest f0 is identical to the f0 profile of the original *bas* token.

A token of a different /b/-initial word produced by the same speaker with closure voicing (i.e., /b/ produced as [b]) was then selected, and the closure voicing was spliced from this word. Two versions of this closure were created. The "full voicing" version of the closure had features typical of voiced closures for this speaker. The duration of the full voicing closure was scaled using *Praat*'s PSOLA algorithm to 130 ms (the average duration of closure voicing for all tokens produced with voicing by this speaker). The intensity of the full voicing closure was scaled using the *Scale intensity...* function in *Praat* to be 75% of that of the following vowel (the average intensity ratio of voiced closures to the following vowel for this speaker). A "reduced voicing" version of the closure was also created with a duration of 68 ms and intensity 65% of the following vowel (values similar to the most reduced voiced closures produced by this speaker).

The two versions of the voiced closure were spliced onto the f0 continua, resulting in three different f0 continua: a continuum with no closure voicing, one with reduced closure voicing and one with full closure voicing. In total there were therefore 12 (6 labial, 6 alveolar) different continua: 4 created from different original tokens (*bas, pas, doer, toer*) x 3 voicing values.

Because the stimuli for the alveolar and labial conditions were synthesized from naturally produced tokens of the voiced and voiceless endpoints at each of the two places of articulation (*bas* and *pas*, *doer* and *toer*), direct comparison of responses to these conditions is not possible. As Fig. 8 shows, across the seven-step continua, the alveolar stimuli had a larger f0 range at vowel onset than the labial stimuli and, unlike the labial stimuli, maintained (much of) this range throughout the vowel. Thus, f0 contributed more to the alveolar continua, which influenced the results in ways that become apparent in Section 3.2. Additionally, in the words used to simulate the alveolar condition (*doer* and *toer*), the vowel is followed by

/r/, a context in which Afrikaans vowels are allophonically lengthened. As a consequence, the vowel in the alveolar condition was longer than that in the labial condition (283 ms vs. 202 ms). For these reasons, the results for the labial and alveolar conditions were analyzed separately.

3.1.3. Procedure

Data collection took place in the same session as the production experiment, after the production recordings. Stimuli were presented over headphones at a comfortable listening level; stimulus presentation was controlled with Superlab stimulus presentation software. Responses were collected through a response box connected to a MacBook laptop computer. Stimuli were blocked by place of articulation, and the order of block presentation was counterbalanced across participants. Stimuli were differently randomized for each participant. Participants were told that they would hear a single Afrikaans word at a time, and that their task was to determine what the word was (either *doer*/*toer* or *bas*/*pas* depending on experimental block). Participants were encouraged to respond quickly, but without sacrificing accuracy, in order to prevent them from reflecting too much on the task. Because the task was fairly easy, no practice trials were given. Each stimulus was presented 10 times to the first participant. In order to keep the duration of the experiment more reasonable, subsequent participants heard each stimulus only eight times. Completion of the task took between 30 and 40 min.

3.2. Results

The responses to the labial continua are shown in Fig. 9, and those to the alveolar continua in Fig. 10. More voiced responses (*bas*/*doer*) were observed for continua that were synthesized from the original voiced than the original voiceless endpoints (top vs. bottom panels). There is also a clear difference in responses between the two continua that had closure voicing compared to the continuum without closure voicing, with more voiced responses given to stimuli that had voicing during closure. Thus, listeners of both age groups relied on voicing during closure as information about the initial consonant. The voiced responses declined with rising f0 for all three closure voicing conditions (although the size of the effect is smaller in the full and reduced voicing conditions; black and gray vs. white circles), showing that older and younger listeners also relied on f0 as information about the phonological voicing of the initial consonant. The trade-off between closure voicing and f0 is visible at the high f0 endpoint for continua that had voicing during closure. For these stimuli, closure voicing and f0 give conflicting evidence about the phonological voicing of the initial plosive. For listeners from both age groups, averaged responses to these stimuli were intermediate between voiced and voiceless. Also for these stimuli with conflicting evidence about voicing, differences between the age groups are visible (left vs. right panels). Overall, older listeners appear to be more likely than younger listeners to identify these stimuli in accordance with closure voicing (as voiced) and in opposition to the high f0 (indicating voicelessness), suggesting that older listeners may rely less on f0, at least when closure voicing and f0 are in conflict. As a measure of this age group difference, we calculated the difference in percent voiced responses between older and younger listeners for the three stimuli with the highest f0 in each of the two conditions with voicing (full or reduced). A positive difference indicates that the older listeners were more likely to identify these stimuli as voiced than the younger listeners. This difference was found to be 24% for the labial continua and 7% for the alveolar continua, indicating that older listeners were more likely than younger listeners to rely on closure voicing, although the age difference was negligible for the alveolar continua.

The responses of the older and younger listeners are also suggestive of possible age-related differences in the categoricity of judgments. For both older and younger listeners, f0 functions gradiently, as is most clearly illustrated in the relatively shallow *s*-shaped curves for responses to tokens without prevoicing. Prevoicing for the labial conditions, though, appears to function less gradiently for the older than the younger listeners: older listeners identified the majority of labial stimuli with prevoicing as voiced, even when these tokens had high f0 values.[5] In the alveolar conditions, however, prevoicing, like f0, functions gradiently for both older and younger listeners.

Because mixed-effect regression models of these data failed to converge, we again modeled the mean structure with marginal logistic regression models, and then used Generalized Estimating Equations to account for within-subject dependence and to generate confidence intervals using this information. The mean structure was modeled in terms of Age (Older vs. Younger), Closure Voicing (full vs. reduced vs. none), Base Voicing (based on original voiced or voiceless endpoints), and f0 (7 steps) using all interactions up to four-way interactions, giving a mean structure with 24 parameters (coding f0 quantitatively and all other variables nominally). Separate analyses were performed for the alveolar and labial data. Pairs of groups defined in terms of Age, Closure Voicing and Base Voicing levels were compared as functions of f0 using simultaneous confidence bands for the difference of fitted linear predictor values (Sun, Loader, & McCormick, 2000).

3.2.1. Influence of f0

In order to determine whether f0 had an effect on the response patterns, we inspected the values predicted by the fitted models together with their 95% confidence bands for each of the response curves from Figs. 9 and 10. If f0 has no effect on response patterns, we would expect the responses to remain flat regardless of changes in f0. Thus, if it is not possible to draw a horizontal line that is contained completely within the 95% confidence band surrounding the estimated mean values for a curve, a significant contribution of f0 to the response pattern can be confirmed. Examples are given in Fig. 11 of the estimated response curves for younger listeners' responses to labial continua based on the voiced endpoint *bas*. Since it is not possible to draw a horizontal line contained within the 95% confidence band for any of these curves, it follows that f0 significantly influenced the younger listeners' responses for all three of these curves. All 24 response curves (3 voicing conditions $\times$ 4 continua $\times$ 2 age groups) were inspected in this manner, and it was found that f0 significantly influenced the response patterns in all 24 curves. The

[5] Even here, though, prevoicing did not categorically override f0 for older listeners. The labial token with full voicing (based on original *bas*) but high f0 was still identified as voiceless 18% of the time.

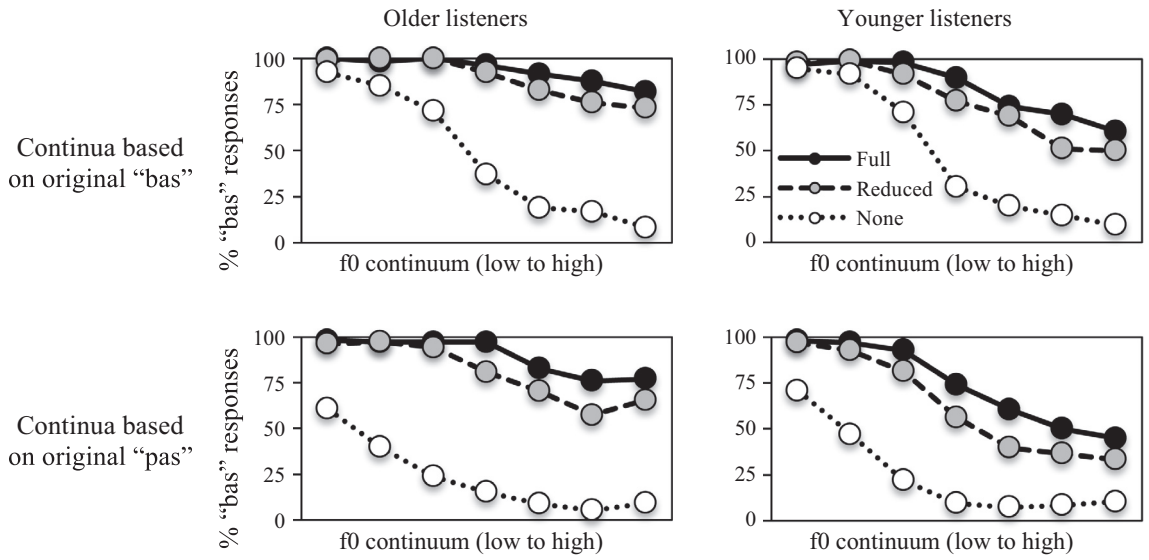


Fig. 9. Percent voiced responses to the labial continua of older (left panels) and younger (right) listeners. The *x*-axis represents the 7 steps of the f0 continuum from lowest to highest f0.

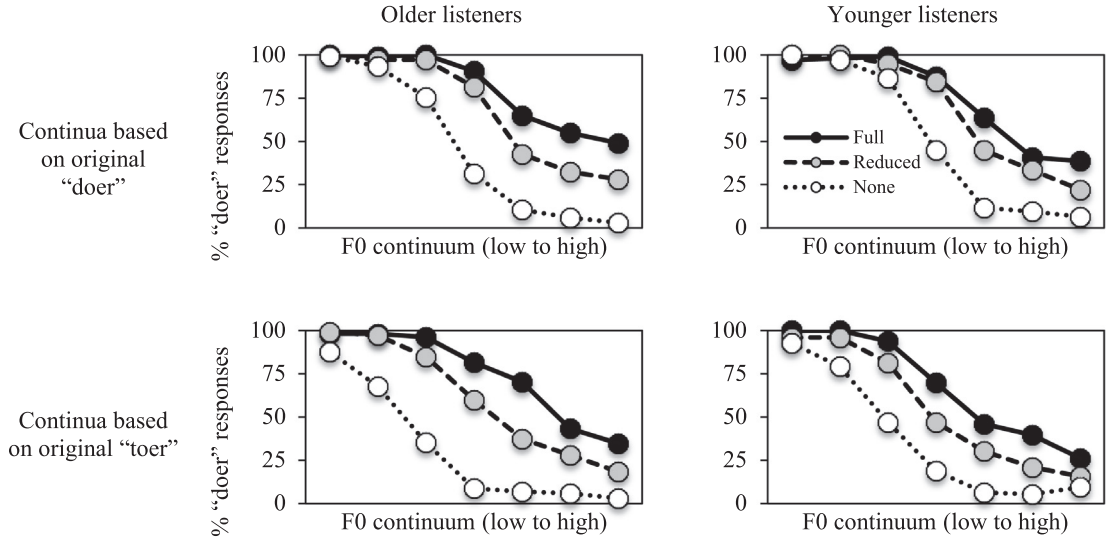


Fig. 10. Percent voiced responses to the alveolar continua. The *x*-axis represents the 7 steps of the f0 continuum from lowest to highest f0.

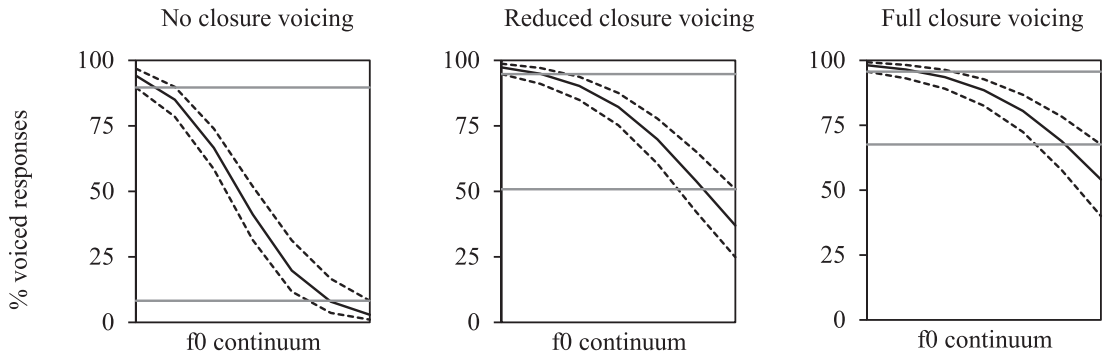


Fig. 11. Estimated response curves for younger listeners to labial stimuli synthesized from the original voiced endpoint *bas*. The solid black line marks the estimated mean response, and the broken lines the upper and lower 95% confidence bands. The solid horizontal lines show that it is not possible to draw a horizontal line that is fully included within the confidence band.

full set of curves is given in Appendix C. Both older and younger listeners therefore relied on f0, and were more likely to give a voiced response to stimuli with lower f0 values.

3.2.2. Influence of closure voicing

To investigate the extent to which closure voicing influenced the response patterns, difference scores were calculated from the fitted linear predictor values for each Closure Voicing pair (full vs. reduced, full vs. none, reduced vs. none), again using simultaneous confidence bands for comparing differences of the fitted values (Sun et al., 2000). If the confidence band for a difference score excludes zero at any point along the f0 continuum, it indicates that the two Closure Voicing conditions being compared are significantly different from each other at that f0 value. Fig. 12 gives examples of the comparisons for older listeners' responses to labial continua based on the voiceless endpoint *pas*. For each of the three comparisons, the confidence band excludes zero, indicating that closure voicing influenced the responses in all closure voicing conditions, with more voiced responses given to full than reduced or no voicing conditions, and more voiced responses to reduced than to no voicing conditions. Similar analyses performed for all other comparisons found that the confidence intervals for all but one of the 24 possible comparisons exclude zero for at least part of the f0 continuum, indicating that Closure Voicing significantly influenced response patterns to labial and alveolar stimuli for both older and younger listeners, and both for stimuli based on the original voiced endpoints (*bas*, *doer*) and those based on the original voiceless endpoints (*pas*, *toer*). Both age groups therefore relied on Closure Voicing as information about the phonological voicing of the initial plosives. Plots of all 24 comparisons are included in Appendix D.

3.2.3. Differences between older and younger listeners

The differences in the response patterns between older and younger listeners were assessed in the same manner by calculating difference scores from the fitted linear predictor values for the older and younger listeners on each of the 12 continua (3 closure voicing conditions for each continuum based on *bas*, *pas*, *doer*, *toer*), and using simultaneous confidence bands for the differences of the fitted values (Sun et al., 2000). The scores for younger listeners were subtracted from those for older listeners so that a positive difference score indicates that older listeners gave more "voiced" (i.e., *bas* or *doer*) responses. As before, if the confidence band excludes zero at any point along the f0 continuum, it indicates that the two age groups differ significantly from each other at that point on the f0 continuum. Inspection of the difference scores for the alveolar continua (based on *doer* and *toer*) revealed differences only for the continua based on the original *toer*, and even then these differences were small and observed only for a small portion of the f0 continuum. For the labial continua, differences were found for stimuli based on both original *bas* and *pas*, and across all or most of the f0 continuum. Fig. 13 represents only the difference scores for the six labial continua. The figures corresponding to the alveolar continua are included in Appendix E. (For discussion of the difference in responses to the labial and alveolar continua, see Section 4.)

As inspection of Fig. 13 confirms, older and younger participants do not differ from each other in their responses to the stimuli without closure voicing (left panels in Fig. 13): when the only cue to the phonological voicing of the initial consonant is f0 of the following vowel, both older and younger listeners relied on f0 and they did so to the same extent. Differences between the older and younger participants' responses arise for the continua with some closure voicing (two right-most panels). For these continua, older listeners were, on average, more likely than younger listeners to identify the initial consonant as voiced /b/. When the older participants' responses in Fig. 9 are assessed in light of the results in Fig. 13, the outcome is that, when there is voicing during the labial closure, older listeners are more likely than younger listeners to respond based on this closure voicing rather than on f0 of the following vowel.

3.2.4. Relation between production and perception patterns

As was shown in Section 2.2.2, older speakers are more likely than younger speakers to produce phonologically voiced /b d/ with phonetic voicing ([b d]). Similarly, as shown in Section 3.2.3, older listeners are more likely than younger listeners to identify stimuli based on closure voicing than on f0 when closure voicing and f0 give conflicting information about the voicing status of the stimulus, at least in the labial condition. Thus, as expected, at the group level there is evidence for a systematic relation between production and perception patterns: older speakers produce more tokens with closure voicing and rely more on closure voicing in perception, particularly in the labial condition.

Although we did therefore find evidence for a relation between production and perception repertoires at the level of the age cohort, these analyses do not directly investigate the relation between perception and production at the level of the individual, and in particular do not identify the ways in which individuals may diverge from the group level patterns. In order to visualize the relation between perception and production at the individual level, we calculated a single production value and a single perception value for each participant in each of the labial and alveolar conditions. The value representing the production pattern was the mean VOT for that speaker for all phonologically voiced tokens (i.e., one value for each speaker for all /b/-initial and one for all /d/-initial words). A speaker who produces more voiced variants of phonologically voiced plosives will therefore have a lower mean VOT value than a speaker who produces more voiceless variants of these plosives. The value representing the perception pattern was each listener's total percent "voiced" responses to the three stimuli with the highest f0 on the full and reduced voicing continua. We chose these stimuli because they elicited the most variable responses across groups and across individuals. A higher score would therefore correspond to listeners who weight closure voicing more heavily in perception when closure voicing and f0 conflict. These values are represented as scatter plots in Fig. 14. A (nearly) perfect correlation (as opposed to the weak correlations found here)[6] between perception and

[6] Simple linear regressions performed separately by place of articulation found a weak but significant negative correlation in the labial condition ($r^2 = 0.22$, $p < 0.03$) and a yet weaker negative correlation that only tends towards significance in the alveolar condition ($r^2 = 0.15$, $p < 0.07$). Given the small number of data points per age cohort and the more limited variation within each age cohort, correlations run separately for age cohorts are not informative.

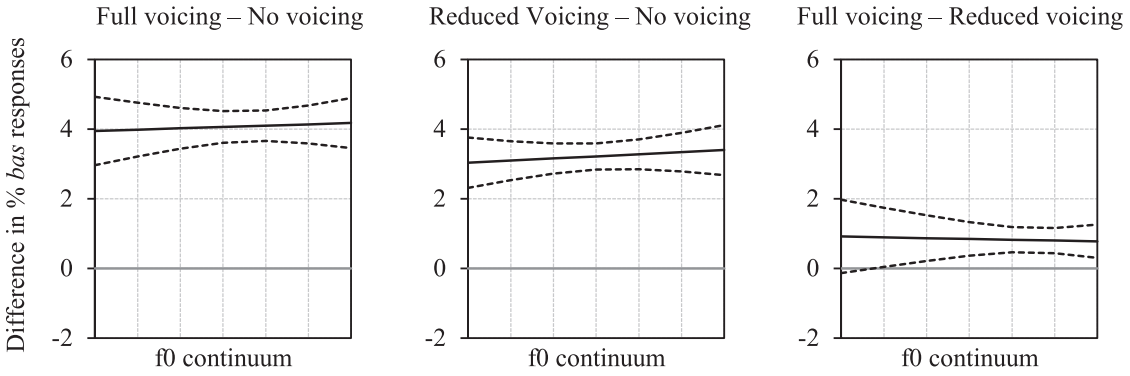


Fig. 12. Comparisons of the response curves of older listeners to continua synthesized from *pas*. See text for details about how these comparisons were implemented. Solid black line: estimated mean difference between response curves; broken lines: 95% confidence interval around this mean. *x*-axis: 7-step f0 continuum; *y*-axis: difference in model predictions for percent voiced responses.

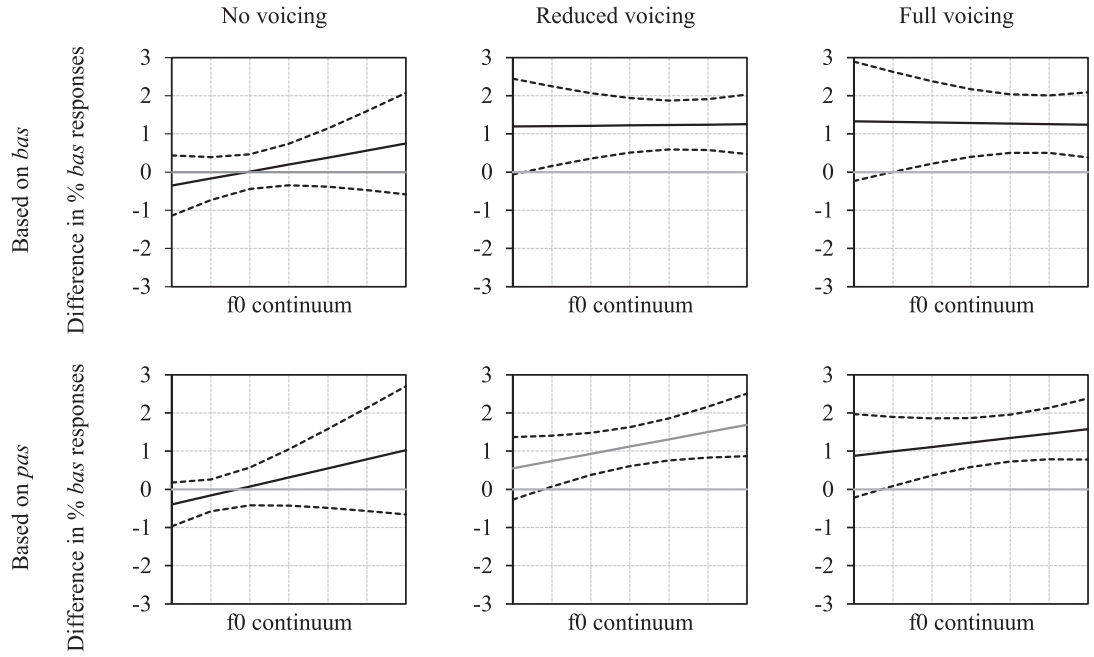


Fig. 13. Comparison of the response curves of older and younger listeners to continua synthesized from *bas* (upper panels) and *pas* (lower). Solid line: estimated mean difference between response curves (Older – Younger); broken lines: 95% confidence interval around this mean. *x*-axis: 7-step f0 continuum; *y*-axis: difference in model predictions for percent voiced responses.

production repertoires at the individual level would have all values clustered along the diagonal from the upper left to lower right quadrants, implying that speakers who produce more phonologically voiced plosives as phonetically voiced (and are therefore lower on the *x*-axis) are also more likely to perceptually identify tokens based on voicing rather than on f0 (and are therefore higher on the *y*-axis).

In broad terms, these plots agree with the results reported for the production and perception experiments above. The results of the younger participants tend to cluster towards the lower right quadrant, indicating that these participants were more likely to produce phonologically voiced plosives as voiceless and also to identify tokens based on f0 rather than voicing. Older participants' results tend to cluster more towards the upper left quadrant, corresponding to a tendency to produce historically voiced plosives as voiced and to perceptually identify tokens with some voicing based on that voicing. The new information provided by this inspection of individuals' responses is that individuals who do not follow these general trends tend to populate the upper right quadrant of the plots, especially in the labial condition. When individuals do not have perceptual and production repertoires that are aligned, they are more likely to rely on voicing in perception than in production. To investigate this apparent mismatch between production and perception for some individuals, we calculated a different perception and production index for every participant. The production value is the percent phonological /b d/ produced as voiced [b d], and therefore indexes the extent to which a speaker uses VOT (independent from f0) to convey the contrast between phonologically voiced and voiceless

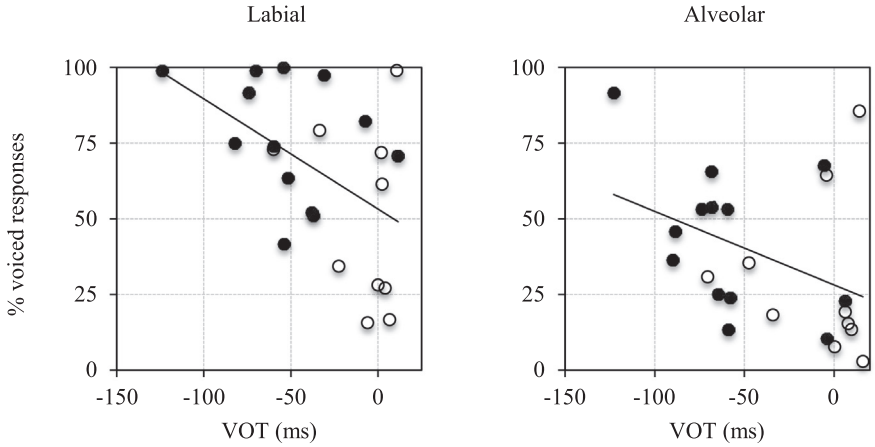


Fig. 14. Scatterplots showing, for each participant, the relation between mean produced VOT for phonologically voiced plosives, and mean percent "voiced" responses to stimuli with the highest three values on the f0 continua with full and reduced voicing. The solid lines represent the linear regression between the production and perception values. Left panel: labial conditions; right: alveolar conditions. Solid circles: older participants; unfilled circles: younger.

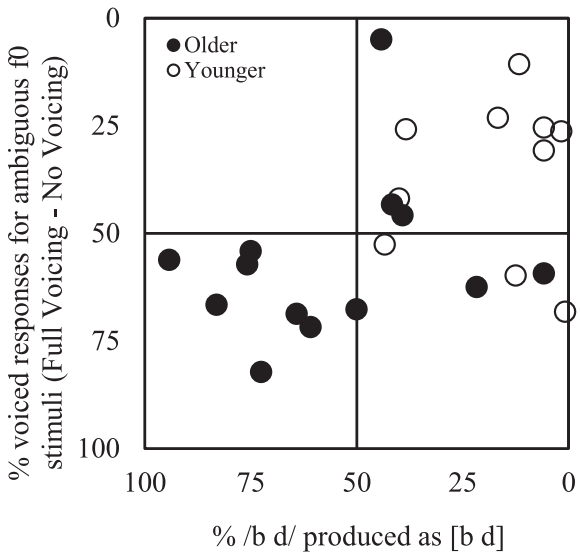


Fig. 15. Reliance on voicing in production (*x* axis) and perception (*y* axis). See text for explanation of measures.

categories. To arrive at a similar index of a listener's reliance on VOT independent from f0, we considered only stimuli with a relatively neutral or ambiguous f0, stimuli 3–5 on the 7-step f0 continuum, and calculated the mean percent voiced response difference: (stimuli with full voicing) – (stimuli with no voicing).[7] (The larger the difference, the more that listener attends to voicing.) Fig. 15 plots this relation, with the axes oriented so that the more conservative (mostly older) participants in terms of the traditional voicing property are closer to the origin in the lower left quadrant and the more innovative (mostly younger) participants are in the upper right quadrant. If production and perception were aligned, we would expect participants to fall roughly along the diagonal. And, indeed, most participants do (and therefore occupy the lower left and upper right quadrants). Participants for whom perception and production are not aligned are, again, those who rely more heavily on VOT in perception than production (lower right quadrant). We return to the interpretation of these patterns in the discussion section.

4. Discussion

As documented in Section 2, the contrast between phonologically voiced and voiceless plosives is realized in Afrikaans in terms of both a VOT difference in the plosives and a difference in f0 of the post-plosive vowel. However, although the f0 difference is consistently present, the VOT difference is only sometimes realized, with variation both within and between individuals. In particular, older speakers tend to realize the voicing contrast with both f0 and VOT differences, while younger speakers are more likely to use only f0. These production patterns are mirrored in perception. Although all listeners rely on both f0 and VOT to differentiate phonologically voiced and voiceless plosives, older listeners tend to rely on voicing more than younger listeners, particularly in the labial condition. In the rest of this section, we explore the implications of these production and perception patterns in terms of their relevance for theories of differential cue weighting and sound change.

The differential use of f0 and VOT in perception was most robustly observed in the labial condition (see in particular Sections 3.2.3 and 3.2.4), and the discussion that follows will therefore focus mostly on this condition. The difference in response patterns between the labial and alveolar conditions most likely does not reflect different perceptual strategies for labial and alveolar plosives in the speech community, but rather results from idiosyncratic properties of the labial and alveolar stimuli used in the perception experiment. Specifically, as discussed in Section 3.1.2, the f0 difference between the voiced and voiceless endpoints was larger in the alveolar continua than in the labial continua, both temporally and spectrally. The weaker influence of voicing on perception in the alveolar condition most likely reflects this difference in the stimuli.

[7] A potentially better measure would be to identify for every listener those three stimuli on the f0 continua that flank the crossover point from voiced to voiceless responses for that individual. However, many listeners (especially for the continua that had closure voicing) identified even the stimulus with the highest f0 value as voiced more than 50% of the time, eliminating this as a possible approach. We therefore used the middle steps on the continuum, under the assumption that these steps at least represent what would be the ambiguous portion of the f0 range for the speaker on whose speech the stimuli were modeled.

4.1. An enhanced f0 effect?

As documented in Section 1.2, languages with consonantal voicing contrasts have typically been found to show small local f0 perturbations on following vowels that are spectrally and temporally less extensive than the f0 differences between tones in languages with phonemic tonal contrasts. If a language were to develop a phonemic tonal contrast from consonantally induced f0 perturbations, it would therefore be expected that the f0 differences in such a language would increase in both their absolute size and their temporal reach into the following vowel. We have speculated that Afrikaans historically had a robust plosive voicing contrast and had only the small, localized f0 perturbations on following vowels typically associated with such a contrast (see Section 1.3). The production data reported here, however, show that the f0 effect in contemporary Afrikaans is comparable in size and extent to that found for languages with phonemic tonal contrasts. For Afrikaans, the f0 difference for phonological voicing is, on average, 44 Hz at vowel onset (Fig. 3), which is three to four times larger than the differences typical for languages without tonal contrasts (Table 1). The f0 difference in Afrikaans also extends throughout the vowel (middle and right panels of Figs. 4 and 5). Moreover, this large and temporally extensive f0 difference depends on the historical (and presumed phonological) rather than phonetic voicing of the preceding consonant (Figs. 4 and 5). That is, the f0 profiles of vowels after voiceless [p t] productions of phonologically voiced /b d/ plosives are statistically indistinguishable from those after voiced [b d] productions.

Further evidence for the exaggeration of the post-plosive f0 effect in Afrikaans comes from comparison of the f0 profiles of vowels after phonologically voiced plosives with those after nasals (Section 2.2.4). Since nasals do not have to maintain a voicing contrast with other sounds, f0 profiles after nasals are often used as a point of reference. In English, French, and Italian, no appreciable difference was found in the f0 profiles after nasals and phonologically voiced plosives, consistent with speakers of these languages not depressing f0 after voiced plosives (Hanson, 2009; Kirby & Ladd, 2015, 2016). In Afrikaans, however, although f0 contours of vowels after nasals and phonologically voiced plosives start out equally low, f0 remains low throughout the vowel after phonologically voiced plosives, while it rises rapidly after nasals (Fig. 6). This pattern suggests that Afrikaans speakers may be actively depressing f0 after phonologically voiced plosives beyond the potential localized influence of the plosive.

An important difference, though, between Afrikaans and other languages in which historical contrasts between plosive consonants have given rise to tonal contrasts on following vowels is that, in some of those languages, the increase in the size of the f0 contrasts has been documented. For example, younger speakers of Seoul Korean have been found to have larger f0 effects than older speakers (e.g., Kang, 2014; see also Bang et al., 2015). In the Afrikaans community that we study here, however, such a difference between older and younger speakers was not found. We hypothesize that this may be because, at the stage in the development of the system when our data were collected, the exaggeration of f0 has been completed and stabilized for all members of the speech community.[8]

Assuming that our interpretation is correct and that the current large f0 effect in Afrikaans reflects a change from an earlier stage when this effect was both smaller and more localized, Afrikaans would be a language in which the consonantally induced f0 perturbations have been exaggerated via one of the mechanisms reviewed in Section 1.2, and may therefore represent a case in which a historical voicing contrast is developing into a tonal contrast. Given that the assumed source of the exaggerated f0 effect (prevoicing in historically voiced plosives) is still robustly present for some (especially older) speakers, Afrikaans would be representative of a Pattern I language in which both the exaggerated contextual effect and its source are present in the system (see Section 1.2). At the same time, the source of the effect (prevoicing) is often absent, especially for younger speakers who produce only 17% of their historically voiced plosives with prevoicing (see Section 2.2.2), indicating that Afrikaans may be developing into a Pattern II system in which the original source of the effect has been lost.

However, more generally determining the phonemic status of f0 in Afrikaans requires further research. The data reported here are for utterance-initial plosives only, and it remains to be determined whether these effects generalize to plosives in other contexts. The Seoul Korean tonogenesis pattern, for example, has been shown to be limited to the initial position in an intonational phrase (Jun, 1996; Kim, 2000). Also in Korean, exaggerated f0 is not limited to aspirated plosive contexts, but has also been found after /h/, the other Korean consonant produced with a spread glottis (Kang, 2014). For Afrikaans, study of f0 patterns after consonants other than the historically voiced and voiceless plosives is in progress.[9]

4.2. Differential cue weighting or sound change?

As reviewed in the previous section, there is evidence for enhancement of the consonantally induced f0 effect in Afrikaans. Such enhancement could be representative of a fairly stable situation of cue trading in the speech community (Section 1.1) or of changing production and perception norms (Section 1.2). As we will show in this section, the data from our study are consistent with both scenarios, though possibly more so with the latter.

Some studies of Korean tonogenesis document a trading relation between VOT and f0 in production (see Section 1.2) such that the f0 difference on post-plosive vowels tends to be larger when the VOT difference is smaller. If a similar pattern were present in Afrikaans production, it should be most

[8] Confirmation of this hypothesis would require gaining access to recordings of speech from speakers one or two generations before the older speakers in our study, a possibility that we are currently exploring.

[9] In this ongoing follow-up work, we have found that the Afrikaans voicing contrast between phonologically voiced and voiceless plosives are maintained in word-medial, intervocalic position, but that the post-consonantal f0 effects after these word-medial plosives are equal in size to those observed in word-initial position. In these regards, Afrikaans appears to be similar to the Tyrolean variety of German, in which the plosive VOT contrast is being lost word-initially but maintained intervocalically (Vietti, Alber, & Vogt, to appear), but different from Korean, in which the f0-effects are only found in phrase-initial position.

clear in comparison of the f0 profiles of vowels after voiced [b d] and voiceless [p t] realizations of /b d/, that is, when VOT information for the latter's contrast with /p t/ is absent. As shown in Section 2.2.3 (Figs. 4 and 5, left panels), the f0 profiles after voiced and voiceless realizations of /b d/ are statistically indistinguishable. There is thus no evidence that Afrikaans speakers exaggerate f0 lowering on vowels after voiceless productions of historically voiced plosives to compensate for the lack of VOT information. This agrees with the results reported by Dmitrieva et al. (2015) for Spanish, who also found no correlation between the amount of prevoicing and f0 of the following vowel. It differs, though, from the findings of Kirby and Ladd (2016) for French and Italian which showed, depending on the sentential prominence position, either a trading relationship between f0 and prevoicing or the opposite relation (i.e., lower f0 associated with more prevoicing).

Given that younger Afrikaans speakers are significantly more likely than older speakers to realize phonologically voiced /b d/ as voiceless [p t] (see Section 2.2.2), it may also be expected that, were the production patterns characterized by a cue trading relation between f0 and prevoicing, younger speakers should produce larger f0 differences to convey the difference between historically voiceless /p t/ and phonetically voiceless [p t] productions of historically voiced /b d/ plosives. As shown in the middle panels of Figs. 4 and 5, however, older and younger speakers alike produce large f0 differences between these realizations. Thus, there is no convincing evidence for a trading relation between f0 and prevoicing in the Afrikaans production data.

Unlike in production, there is evidence for trading between VOT and f0 in the perceptual differentiation of phonologically voiced /b d/ and voiceless /p t/ in Afrikaans. For the stimuli used in the perception study, cue trading would be evident if the crossover from voiced to voiceless judgments along the f0 continuum were to shift as a function of voicing information (full, reduced, none). This pattern is robustly present in the judgments of older and younger listeners for both the labial and alveolar continua (Figs. 9 and 10). Although f0 variation is sufficient for listeners to shift from voiced to voiceless responses, the proportion voiceless responses predictably decreases as prevoicing increases.

For the labial stimuli with prevoicing, there is also evidence of differential cue weighting by older and younger listeners. Specifically, younger listeners were significantly more likely than older listeners to identify stimuli with prevoicing but high f0 as voiceless (middle and left panels in Fig. 13 and Appendix E). Llanos et al. (2013) investigated a similar scenario with Spanish-speaking listeners and found, unlike the current results, that listeners identified virtually all tokens with prevoicing as voiced, even if those tokens had fairly high f0 (i.e., participants in their study did not perceptually trade f0 and prevoicing). This difference between Spanish and Afrikaans may reflect the different function of prevoicing in the plosive systems of these two languages: phonologically voiced plosives are virtually always realized with prevoicing in Spanish, but not Afrikaans.

The results from our study therefore show that the VOT information for the contrast between historically voiced and voiceless plosives is used differently by different members of the speech community. Differential cue use is expected for contrasts that are reliably marked by more than one acoustic property. Shultz et al. (2012; see also Massaro & Cohen, 1976, 1977), for instance, reported individual variation in cue weighting for American English speakers' production and perception of VOT and f0 as information for voicing contrasts. However, unlike the situation documented by Shultz et al. the variation in terms of the relative importance assigned to VOT in the Afrikaans speech community is generationally structured. Younger speakers are more likely not to use VOT to differentiate voicing contrasts in production and, similarly, younger listeners are more likely to ignore VOT and rely on f0 perceptually.

The age-based structure of the differential cue weighting in the Afrikaans speech community could be interpreted as evidence that Afrikaans is in the process of replacing a voicing contrast with a tonal contrast. This interpretation, though, depends on the assumption that "apparent time" (differences between speakers from different contemporaneous generations) can be used as evidence of change, an idea promoted by Labov based on his classic Martha's Vineyard and New York City studies (Labov, 1963, 1965, 1966). However, as has been pointed out by Labov himself (e.g., Labov, 1972:24) and many others (Bailey, 2002; Sankoff, 2006; Tagliamonte, 2012; Wagner, 2012), apparent time evidence does not provide definitive evidence of change—apparent time data are also consistent with "age-grading", i.e., a scenario in which linguistic patterns change over the lifetime of an individual. It is possible that the younger participants, as they age, may develop repertoires similar to those of the older participants. To differentiate between these two possibilities, it would be necessary to evaluate data with more time depth than is currently available for Afrikaans.

4.3. Individual and community norms

In our description of the patterns observed in the Afrikaans speech community we have mainly focused on patterns present at the level of the whole community, or the generational level. However, unsurprisingly, some individuals diverge from community norms, and the ways in which individuals do, and do not, differ from the majority pattern can be informative about how variation is structured at the level of the community, and about how change may spread through a community.

The relation between individuals and the community with regard to perception and production norms was visualized in Fig. 15. That figure shows that, in general, community members have perception and production repertoires that are aligned. Speakers who regularly produce prevoicing tend to attend more to prevoicing than f0 in perception and those who tend not to prevoice rely less on prevoicing perceptually (i.e., the older compared to the mostly younger participants in the lower left and upper right quadrants of Fig. 15, respectively). However, for two younger participants

and two older participants (Fig. 15, lower right quadrant, distant from the diagonal), the use of prevoicing in perception and production diverges: they rely on prevoicing perceptually even though they do not often produce it. Conspicuously absent from the data are individuals who rely on prevoicing in production but not in perception (Fig. 15, upper left quadrant). Although data from more participants might disconfirm this pattern, the current findings are consistent with a sound change in which production is in the lead: the individuals in the lower right quadrant have adopted the innovative (voiceless) production norm, but still rely on the older conservative perception norm. This outcome is different from that documented by Harrington et al. (2008) for /u/-fronting in Southern British English, where perceptual norms changed before production norms, but is consistent with Pinget's (2015) finding for obstruent devoicing in Dutch where, in the later stages of the change, perception lags behind production in that participants who devoice nonetheless categorize stimuli differing in voicing reasonably consistently.

4.4. Conclusion

This study documents how the historical plosive voicing contrast of Afrikaans is realized in the contemporary variety of this language. The acoustic evidence shows that both older and younger speakers produce many historically voiced plosives as voiceless unaspirated. The devoicing rate of 44% for older and 83% for younger speakers is higher than that observed in Dutch, where the highest reported devoicing rate is only 25% (van Alphen & Smits, 2004:462). Assuming that the difference between the age groups in the Afrikaans community is not due to age grading in the speech community, this result is suggestive of the word-initial voicing contrast being in the process of disappearing in Afrikaans. The acoustic data also show large and temporally extensive differences in f0 after historically voiced and voiceless plosives for both younger and older speakers of Afrikaans—differences that exceed the typical consonantally induced f0 perturbations present in languages like English and Dutch. This f0 effect depends on the historical/phonological voicing of the plosive rather than the actual phonetic (VOT) realization of the plosive (i.e., f0 is low even after voiceless productions of historically voiced plosives). Correspondingly, f0 is perceptually important as well. Afrikaans listeners systematically use f0 to differentiate phonologically voiced and voiceless plosives; and especially for some younger listeners, f0 cues can sometimes override voicing information.

In these regards, f0 patterns in Afrikaans like f0 in a language in which the f0 differences appear to have been phonologized (e.g., Korean), so that Afrikaans could be interpreted as an example of a language in which a contrast that used to be conveyed primarily by voicing differences between plosives is now primarily conveyed by f0 differences of vowels following the plosives. Given that the presumed source of the f0 effect (consonantal voicing) is still present in the speech of many Afrikaans speakers and also given that many Afrikaans-speaking listeners still rely on this information, we cannot conclude that Afrikaans has undergone tonogenesis. However, if the trend for lesser reliance on prevoicing that is observed in particular for younger speakers were to continue, it seems not unlikely that this process could go to completion.

Small consonantally induced f0 perturbations are most likely present in all languages that have a phonological voicing contrast (see Section 1.2). Why then do these f0 effects become exaggerated and even develop into independent phonemic contrasts in a small number of languages (as may be happening in Afrikaans), while they remain small and stable in most languages? An intriguing possibility for the Afrikaans development is language contact. What differentiates the Afrikaans situation from that of many other languages (including Dutch as Afrikaans's primary linguistic ancestor) is that Afrikaans was in close contact with tonal Khoisan languages during the early years of its formation, and has been in prolonged contact with tonal Bantu languages since. It is possible that this contact situation could have contributed to greater perceptual sensitivity to f0 differences in the Afrikaans speech community, and that this awareness could have contributed to exaggeration of the f0 differences in Afrikaans over successive generations. Thurgood (1996), for instance, argued that it was through contact with tonal Vietnamese that Eastern Cham developed a tonal split based on consonantal voicing (though see Brunelle, 2009 for potential evidence to the contrary). However, although it is possible that contact with tonal languages may have played a role in the origin of the current Afrikaans consonantal voicing system, the dynamics of South African society make this unlikely. As the economically and politically dominant language, the common situation was for speakers of Bantu and Khoisan languages to learn Afrikaans in order to facilitate communication with Afrikaans speakers. Throughout the history of Afrikaans, the percentage of speakers of White Afrikaans (the variety of Afrikaans that is the focus of our study) who would have known more than the most basic vocabulary of a Bantu or Khoisan language would have been negligible.

Acknowledgements

This work is supported by NSF Grant BCS-1348150 to Patrice Speeter Beddor and Andries W. Coetzee. Earlier versions of this work were presented at the 3rd Biennial Workshop on Sound Change (2014, Berkeley, CA), the 167th Meeting of the Acoustical Society of America (2014, Providence, RI) and the Annual Meeting of the Linguistic Society of Southern Africa (2015, Potchefstroom, South Africa). We also thank the audiences at several universities where we have discussed this research, and in particular James Kirby, Bob Ladd and Phil Hoole for valuable feedback on this project.[10]

[10] This footnote pertains to Appendix A. The phonetic transcriptions given here represent the conservative pronunciation of the northern variety of so-called "Standard" Afrikaans. All historically voiced plosives are transcribed here as also phonetically voiced, although such plosives are actually often produced as voiceless. In classifying tokens into the five different vowel groups, we do not take into account either phonemic or allophonic vowel length. In the variety of Afrikaans represented here, [ɛ] and [æ] are allophones of /ɛ/, with [æ] found before [r], [l] and velars. Afrikaans is currently undergoing a change in which stressed [ɑː] is raising and rounding to [ɔː]. The effects of this change are not represented in the transcriptions given here.

Appendix A

Token and filler words used in the production experiment

/b p/-initial words						
Vowel	/b/-initial			/p/-initial		
[ɑ]	bak	[bɑk]	‘bake’	pak	[pɑk]	‘pack’
	bars	[bɑrʃ]	‘burst’	pars	[pɑrʃ]	‘press’
	bas	[bɑs]	‘(tree) bark’	pas	[pɑs]	‘just now’
	baba	[bɑːbɑ]	‘baby’	padda	[pɑdɑ]	‘frog’
	bate	[bɑːtə]	‘asset’	papa	[pɑpɑ]	‘daddy’
[ɛ æ]	Ben	[bɛn]	‘Ben’	pen	[pɛn]	‘pen’
	berg	[bærx]	‘mountain’	pers	[pæːrʃ]	‘purple’
	bêre	[bæːrə]	‘put away’	perd	[pæːrt]	‘horse’
[ɔ]	bot	[bɔt]	‘bud’	pot	[pɔt]	‘pot’
	bont	[bɔnt]	‘variegated’	pond	[pɔnt]	‘pound’
	bons	[bɔns]	‘bounce’	pons	[pɔns]	‘punch’
[i]	bied	[bit]	‘offer (to)’	Piet	[pit]	‘Pete’
	bieg	[bix]	‘confess’	piek	[pik]	‘peak’
[u]	boet	[but]	‘brother’	poets	[puts]	‘polish’
	boer	[buːr]	‘farmer’	poel	[pul]	‘pool’

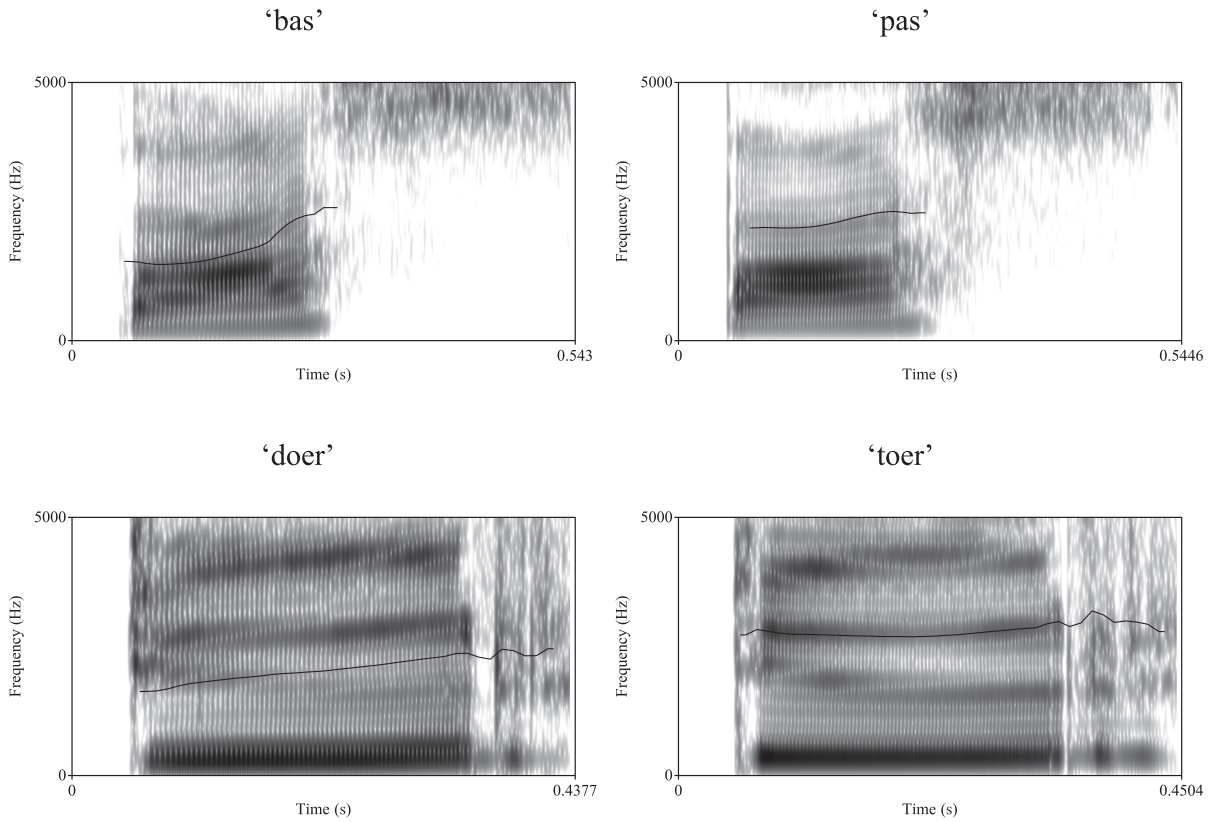
/d t/-initial words						
Vowel	/d/-initial			/t/-initial		
[ɑ]	dak	[dɑk]	‘roof’	tak	[tɑk]	‘branch’
	das	[dɑs]	‘neck tie’	tas	[tɑs]	‘suitcase’
	dan	[dɑn]	‘then’	tand	[tɑnt]	‘tooth’
	darem	[dɑrəm]	‘at least’	tannie	[tɑni]	‘aunt’
	dadel	[dɑːdəl]	‘date’	takel	[tɑːkəl]	‘tackle’
[ɛ æ]	den	[dɛn]	‘den’	tent	[tɛnt]	‘tent’
	derm	[dærəm]	‘intestines’	term	[tærəm]	‘term’
	derde	[dæːrdə]	‘third’	terte	[tæːrtə]	‘tarts’
[ɔ]	dor	[dɔr]	‘parched’	tor	[tɔr]	‘beetle’
	dons	[dɔns]	‘feather down’	ton	[tɔn]	‘ton’
	dokter	[dɔktər]	‘doctor’	tokkel	[tɔkəl]	‘strum’
[i]	dien	[din]	‘serve’	tien	[tin]	‘ten’
	dier	[diːr]	‘animal’	tier	[tiːr]	‘tiger’
[u]	doer	[duːr]	‘over there’	toer	[tuːr]	‘tour’
	doen	[dun]	‘do’	toets	[tuts]	‘test’

/v f/-initial words						
Vowel	/v/-initial			/f/-initial		
[ɑ]	was	[vɑs]	‘wash’	vas	[fɑs]	‘attached’
	wat	[vɑt]	‘what’	vat	[fɑt]	‘take’
[ɛ æ]	weg	[væx]	‘missing’	veg	[fæx]	‘fight’
	wet	[vɛt]	‘law’	vet	[fɛt]	‘fat’
[ɔ]	wond	[vɔnt]	‘wound’	font	[fɔnt]	‘font’
	word	[vɔːrt]	‘become’	fort	[fɔrt]	‘fort’
[i]	wiel	[vil]	‘wheel’	vies	[fis]	‘upset’
	wier	[viːr]	‘seaweed’	vier	[viːr]	‘four’
[u]	woed	[vut]	‘raging’	voet	[fut]	‘foot’
	woer	[vur]	‘whirr’	voer	[fuːr]	‘feed’

/m n k/-initial words						
Vowel	/m n/-initial			/k/-initial		
[ɑ]	maak	[mɑːk]	‘make’	kaas	[kɑːs]	‘cheese’
	nat	[nɑt]	‘wet’	kat	[kɑt]	‘cat’
[ɛ æ]	met	[mɛt]	‘with’	ken	[kɛn]	‘chin’
	nek	[næk]	‘neck’	kerk	[kærk]	‘church’
[ɔ]	mos	[mɔs]	‘after all’	kos	[kɔs]	‘food’
	nog	[nɔx]	‘additional’	kop	[kɔp]	‘head’
[i]	mier	[miːr]	‘ant’	kiem	[kim]	‘germ’
	nies	[nis]	‘sneeze’	kies	[kis]	‘choose’
[u]	moet	[mut]	‘must’	koek	[kuk]	‘cake’
	noem	[num]	‘name’	koets	[kuts]	‘charriot’

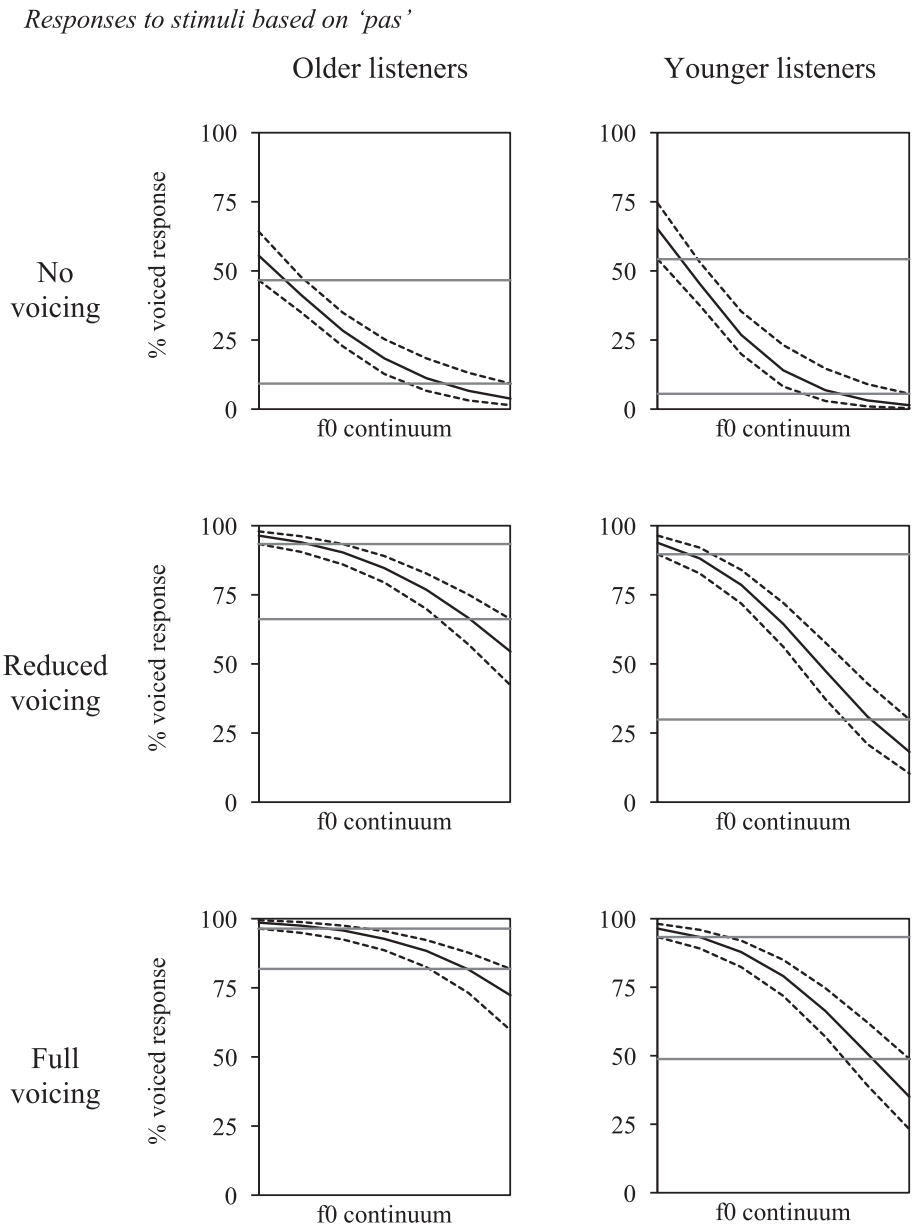
Appendix B

Spectrograms corresponding to the four endpoint tokens that were used for the simulation of the stimuli in the perception experiment. The solid black line on the spectrogram marks the f0 contour as estimated by *Praat* (on a scale from 75 to 500 Hz).

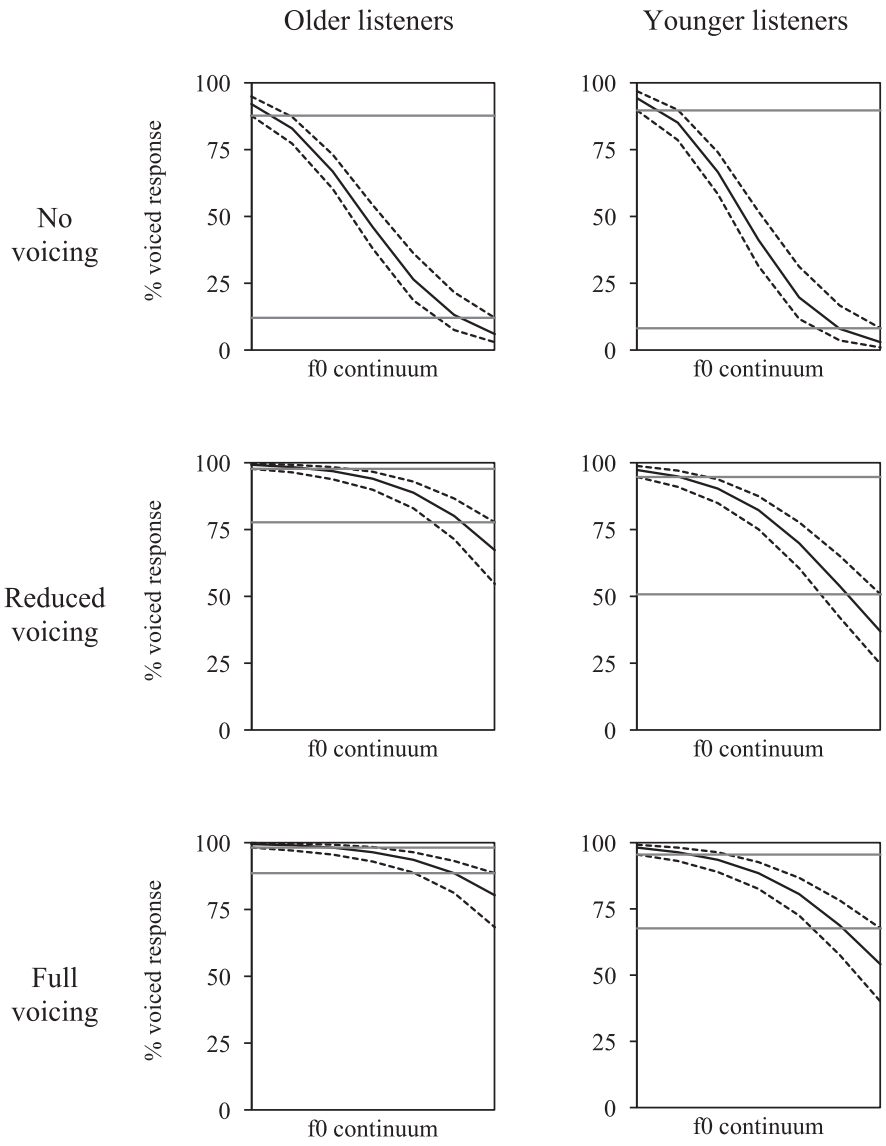


Appendix C

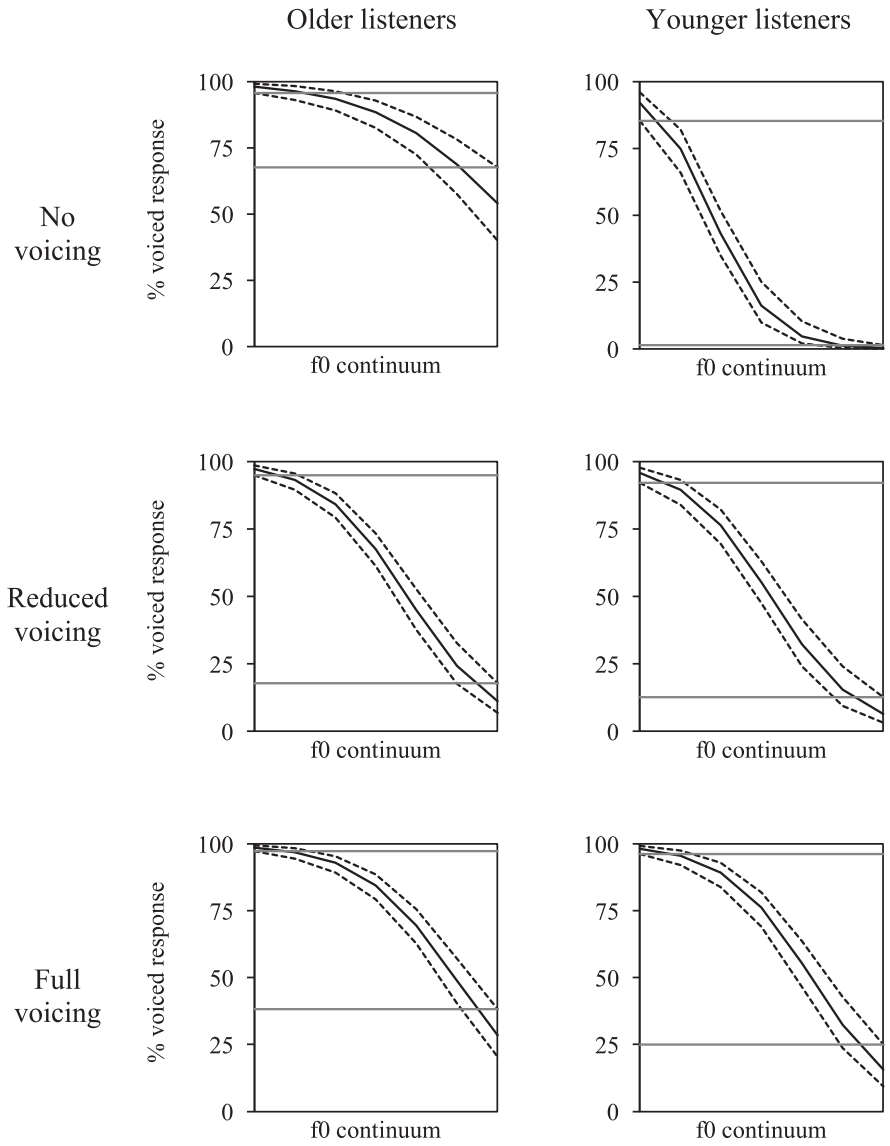
Modeled response curves for older and younger listeners for each of the 12 perception continua. In each graph, the solid black line marks the estimated mean response, and the broken lines the upper and lower 95% confidence bands. The solid gray lines show that it is not possible to draw a horizontal line that is fully included within the confidence band. Left graphs: older listeners; right: younger listeners. See Section 3.2.1 for more information.

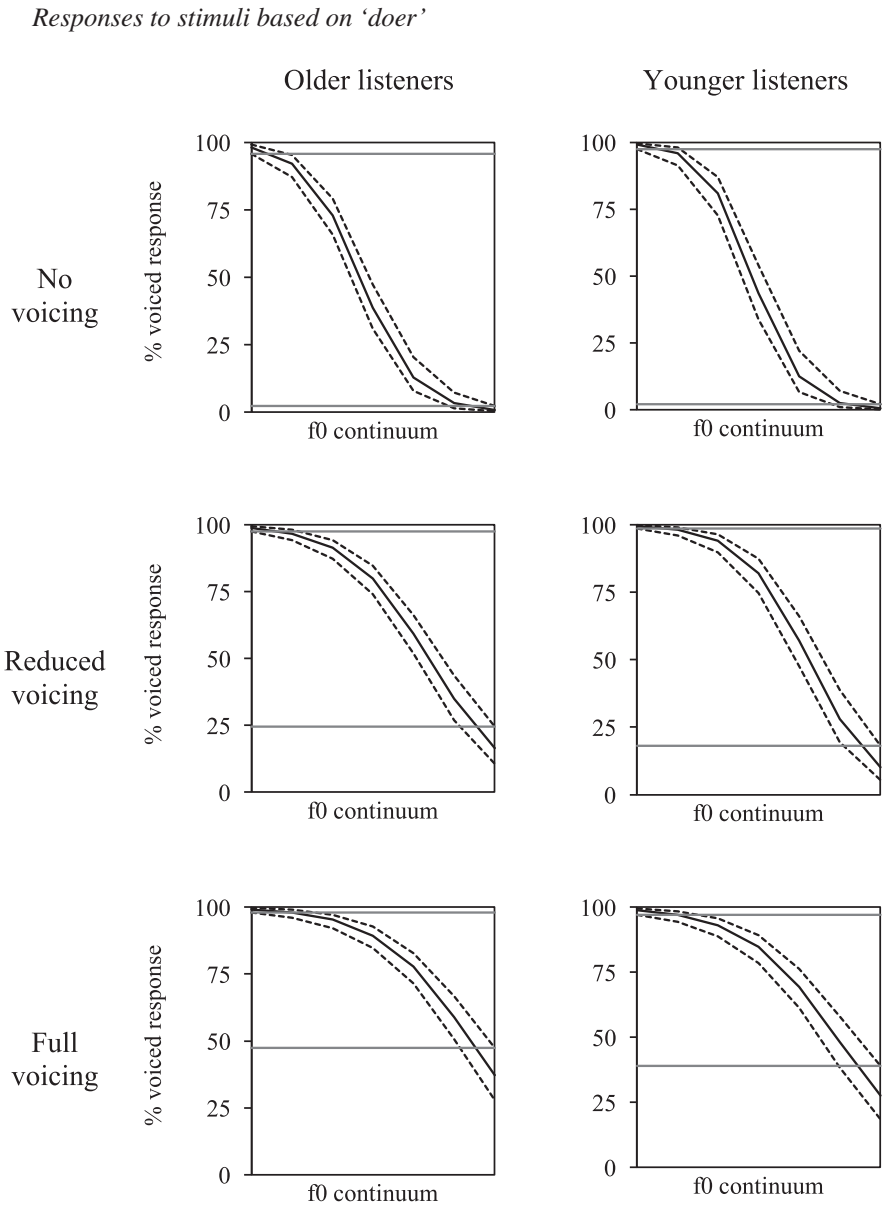


Responses to stimuli based on 'bas'



Responses to stimuli based on 'toer'

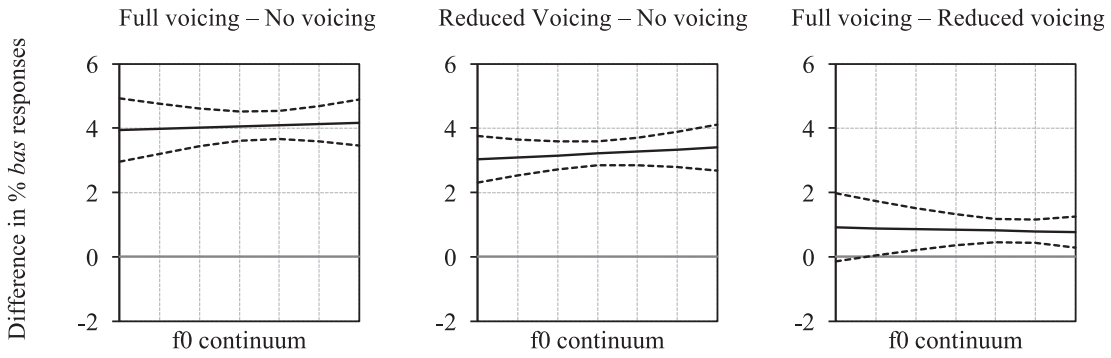




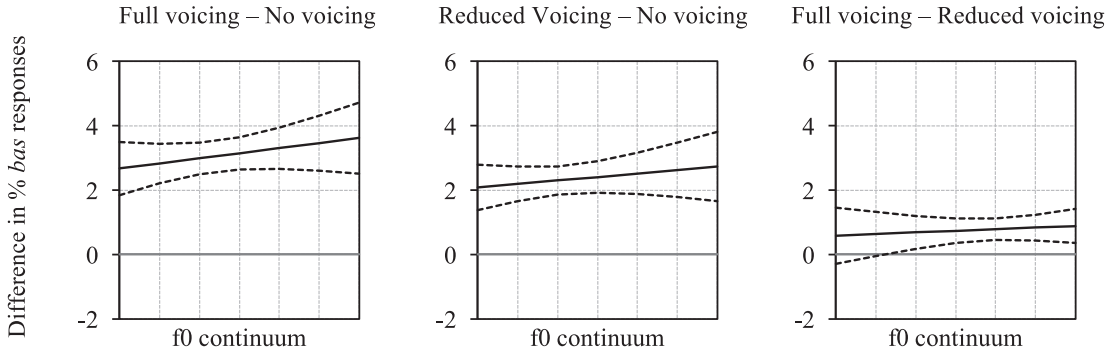
Appendix D

Difference in the modeled response curves for older and younger listeners between continua with different levels of closure voicing (no voicing, reduced voicing, full voicing). Solid lines mark estimated mean difference between response curves, and broken lines the 95% confidence interval around this mean. The *x*-axis represents the 7-step f0 continuum, and the *y*-axis the difference in model predictions for percent voiced responses. See Section 3.2.2 for more information.

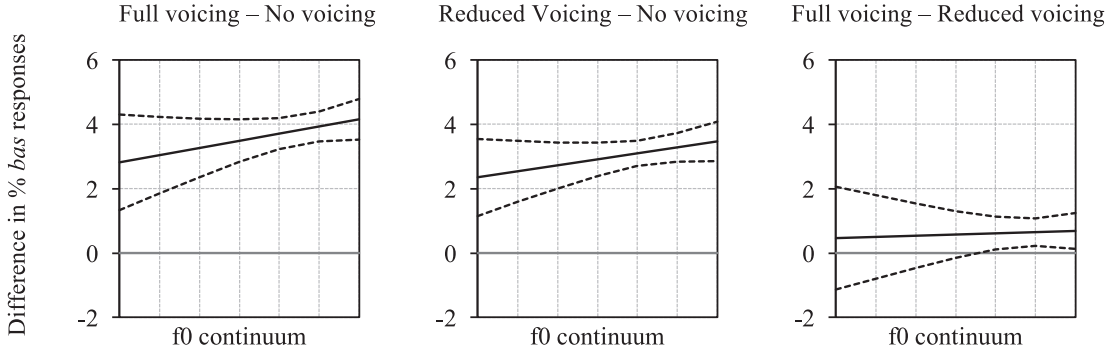
Responses to stimuli based on 'pas', older listeners



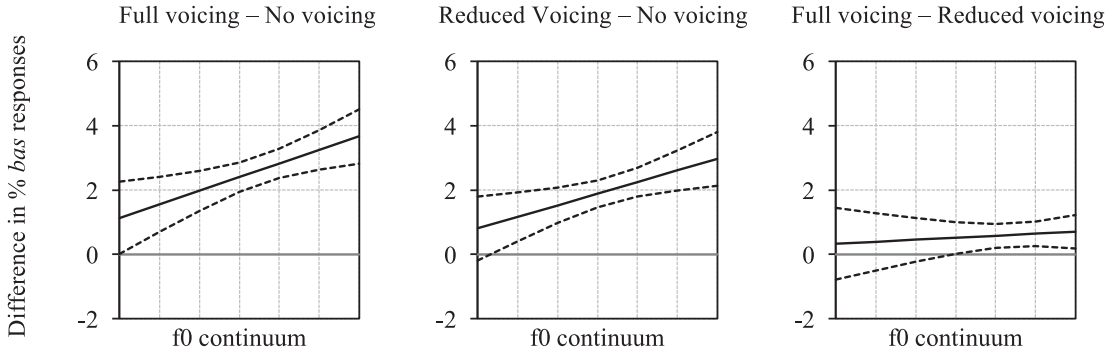
Responses to stimuli based on 'pas', younger listeners



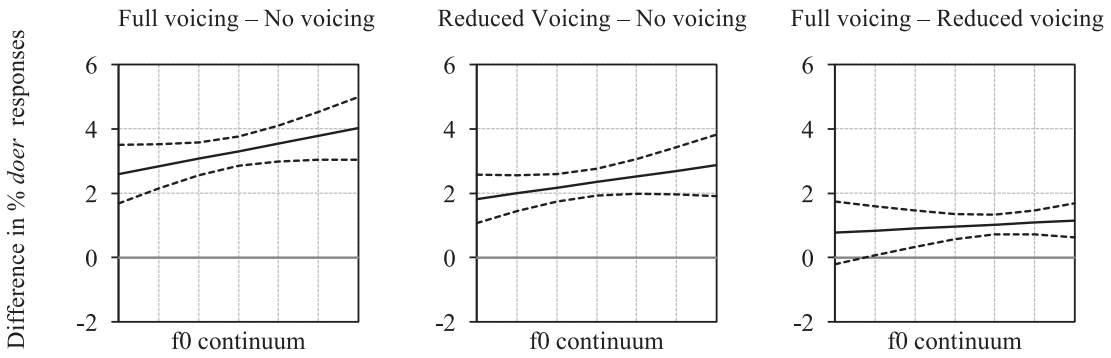
Responses to stimuli based on 'bas', older listeners



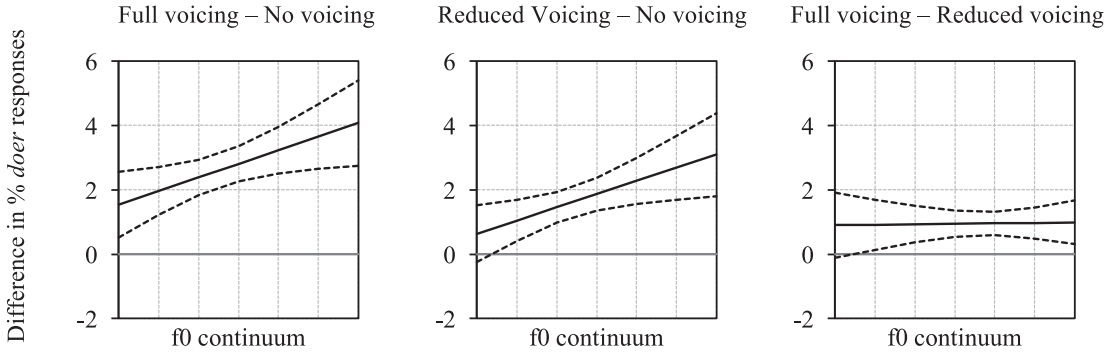
Responses to stimuli based on 'bas', younger listeners



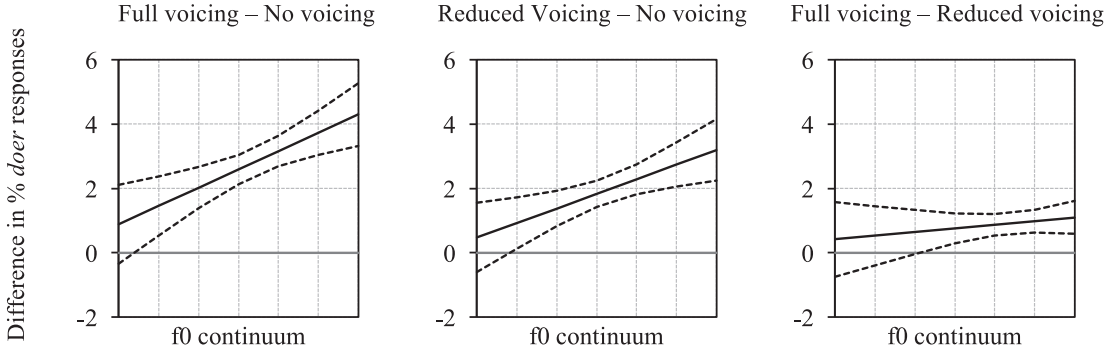
Responses to stimuli based on 'toer', older listeners



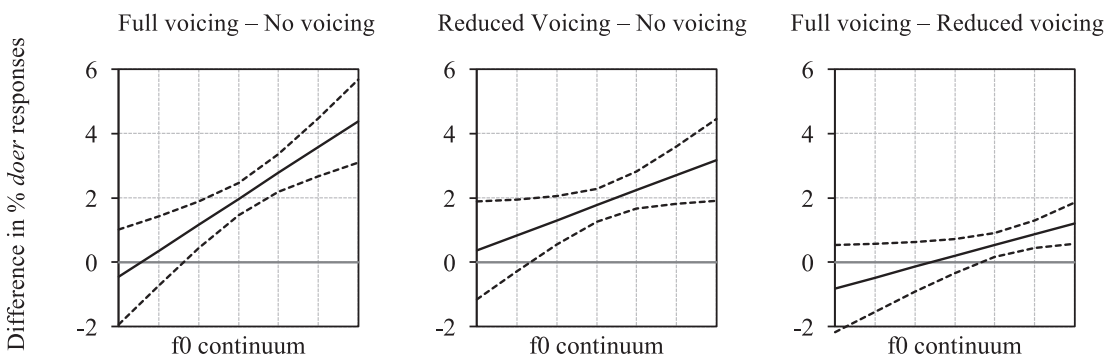
Responses to stimuli based on 'toer', younger listeners



Responses to stimuli based on 'doer', older listeners



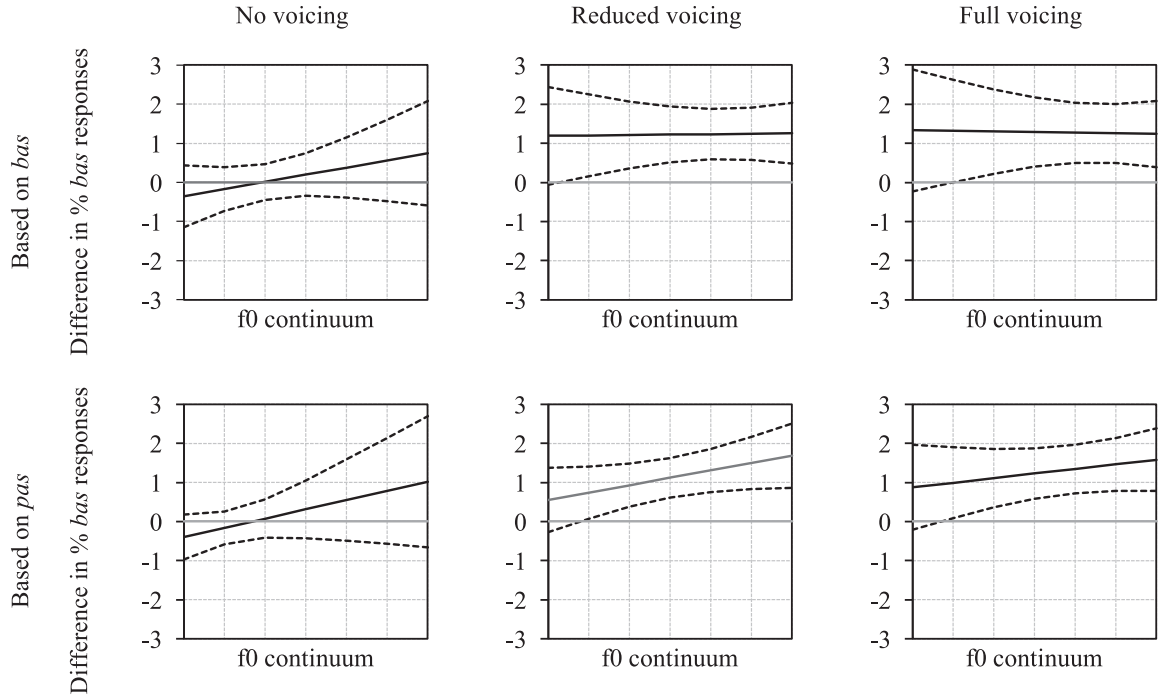
Responses to stimuli based on 'doer', younger listeners



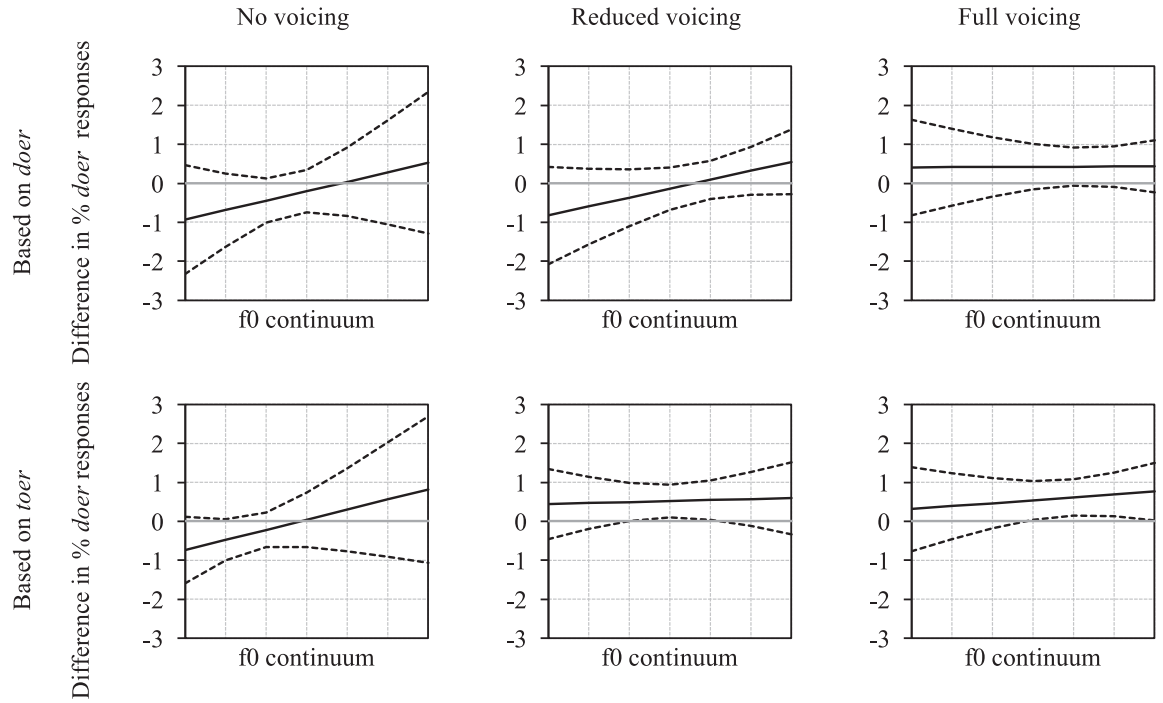
Appendix E

Modeled differences between older and younger listeners to each of the 12 perception continua. In all graphs, the solid black line marks the estimated mean difference between response curves (Older – Younger), and broken black lines the 95% confidence interval around this mean. The *x*-axis represents the 7-step f0 continuum, and the *y*-axis the difference in model predictions for percent voiced responses. See Section 3.2.3 for more information.

Labial continua



Alveolar continua



References

Abramson, A. S. (1975). Pitch in the perception of voicing states in Thai: Diachronic implications. *Status Report, Haskins Laboratories, 41*, 165–174.

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: f0 shift versus voice timing. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25–33). Orlando: Academic Press.

Abramson, A. S., & Whalen, D. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics, 63*, 75–86.

Amemiya, T. (1985). *Advanced econometrics*. Cambridge, MA: Harvard University Press.

Bailey, G. (2002). Real and apparent time. In J. K. Chambers, N. Schilling-Estes, & P. Trudgill (Eds.), *The handbook of language variation and change* (pp. 312–332). Oxford: Blackwell.

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2015). The effect of word frequency on the timecourse of tonogenesis in Seoul Korean. In T. S. C. f. I. 2015 (Ed.), *ICPhS XVIII: Proceedings of the 18th international congress of phonetic sciences*. London: International Phonetic Association. (https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0843.pdf).

Bates, D., Maechler, M., & Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes, R package version 0.999999-2. <http://cran.r-project.org/package=lme4> (Last downloaded June 9, 2013).

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language, 85*, 785–821.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, 17*, 97–110.

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer (Version 5.3.56)*. [Computer Program.] Retrieved May 23, 2012, from <http://www.Praat.org/>.

Brunelle, M. (2009). Contact-induced change? Register in three Cham dialects. *Journal of the Southeast Asian Linguistics Society, 2*, 1–22.

Collins, B., & Mees, I. (1982). A phonetic description of the consonant system of Standard Dutch. *Journal of the International Phonetic Association, 12*, 2–12.

Collins, B., & Mees, I. (2003). *The phonetics of English and Dutch* (5th Revised Edition). Leiden: Brill.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics, 54*, 35–50.

de Boor, C., 1978. *A practical guide to splines*. (Applied Mathematical Sciences series, Vol. 27.) New York: Springer-Verlag.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics, 49*, 77–95.

Ewan, W. G., & Krones, R. (1974). Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics, 2*, 327–335.

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: A roughness penalty approach*. London: Chapman & Hall.

Gussenhoven, C. (1999). Illustrations of the IPA: Dutch. In *Handbook of the International Phonetic Association*. Cambdrige: Cambdrige University Press, pp. 74–44.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *Quarterly Progress Report, MIT Research Lab. of Electronics, 101*, 198–213.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America, 47*, 613–617.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *Journal of the Acoustical Society of America, 125*, 425–441.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America, 123*, 2825–2835.

Hombert, J.-M. (1977). Consonant types, vowel height and tone in Yoruba. *Studies in African Linguistics, 8*, 173–190.

Hombert, J.-M. (1978). Consonant types, vowel quality, and tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 77–111). New York: Academic Press.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language, 55*, 37–58.

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in f0 control. *Language and Speech, 42*, 401–411.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25*, 105–113.

Jessen, M. (1999). *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: John Benjamins.

Jun, S.-A. (1996). Influence of microprosody on macroprosody: a case of phrase initial strengthening. *UCLA Working Papers in Phonetics, 92*, 97–116.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics, 45*, 76–90.

Kang, Y., & Han, S. (2013). Tonogenesis in early Contemporary Seoul Korean: A longitudinal case study. *Lingua, 134*, 62–74.

Kim, M.-R. (2000). *Segmental and tonal interactions in English and Korean: A phonetic and phonological study* (Unpublished Ph.D. Dissertation). University of Michigan.

Kingston, J. (2005). The phonetics of Athabaskan tonogenesis. In S. Hargus & K. Rice (Eds.), *Athabaskan prosody* (pp. 137–184). Amsterdam: John Benjamins.

Kingston, J. (2011). Tonogenesis. In M. van Oostendorp, C. J. Ewen, & K. Rice (Eds.). *Blackwell companion to phonology* (Vol. 4, pp. 2304–2334). Oxford: Blackwell.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language, 70*, 419–454.

Kirby, J. P. (2013). The role of probabilistic enhancement in phonologization. In A. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 228–246). Oxford: Oxford University Press.

Kirby, J. P., & Ladd, D. R. (2015). Stop voicing and f0 perturbations: evidence from French and Italian. In T. S. C. f. I. 2015 (Ed.), *Proceedings of the 18th international congress of phonetic sciences*. London: International Phonetic Association.

Kirby, J. P., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel F0: Evidence from "true voicing" languages. *Journal of the Acoustical Society of America, 140*, 2400–2411.

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics, 59*, 40–57.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models, R package version 2.0-33. <https://cran.r-project.org/web/packages/lmerTest/index.html> (Last downloaded on May 25, 2017).

Labov, W. (1963). The social motivation of a sound change. *Word, 19*, 273–309.

Labov, W. (1965). On the mechanism of linguistic change. *Georgetown Monographs on Language and Linguistics, 18*, 91–114.

Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Le Roux, T. H., & Pienaar, P. d. V. (1927). *Afrikaanse Fonetiek*. Cape Town: Juta & Co.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America, 33*, 419–425.

Lindblom, B., Guion, S., Hura, S. L., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica, 7*, 5–37.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384–422.

Liu, M. (2015). *The Production and Identification of Mandarin Tones in Context* (Unpublished M.A.). Indiana University.

Llanos, F., Dmitrieva, O., Shultz, A. A., & Francis, A. L. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *Journal of the Acoustical Society of America, 134*, 2213–2224.

Löfqvist, A. (1975). Intinsic and extrinsic f0 variations in Swedish tonal accents. *Phonetica, 31*, 228–247.

Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America, 85*, 1314–1321.

Maddieson, I., & Pang, K.-F. (1993). Tone in Utsat. *Oceanic Linguistics Special Publications, 24*, 75–89.

Maran, L. R. (1973). On becoming a tonal language: A Tibeto-Burman model of tonogenesis. In L. M. Hyman (Ed.), *Consonant types and tone. Southern California occasional papers in linguistics* (pp. 98–114). Los Angeles: University of Southern California.

Massaro, D. W., & Cohen, M. H. (1976). The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. *Journal of the Acoustical Society of America, 60*, 704–717.

Massaro, D. W., & Cohen, M. H. (1977). Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction. *Perception & Psychophysics, 22*, 373–382.

Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. M. Hyman (Ed.), *Consonant types and tones* (pp. 72–95). Los Angeles: Linguistics Program, University of Southern California.

Myers, S. (1999). Tone association and f_0 timing in Chichewa. *Studies in African Linguistics, 28*, 215–239.

Myers, S. (2001). F_0 timing in Kinyarwanda. *Phonetica, 60*, 71–97.

Ohala, J. J. (1973). The physiology of tone. In L. Hyman (Ed.), *Consonant types and tone* (pp. 1–14). Los Angeles: Linguistics Program, University of Southern California.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Chicago Linguistic Society: Papers from the Parasession on language and behavior* (pp. 178–203). Chicago: Chicago Linguistic Society.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 235–278). London: Longman.

Pearce, M. (2005). Kera tone and voicing. In M. Pearce & N. Topintzi (Eds.). *University College London working papers in linguistics* (vol. 17, pp. 61–82). London: University College London.

Pearce, M. (2009). Kera tone and voicing interaction. *Lingua, 119*, 846–864.

Pinget, A.-F. (2015). *The actuation of sound change* (Unpublished Ph.D. Dissertation). Utrecht University.

R Core Team (2013). *R: A language and environment for statistical computing*. 3-900051-07-0. Vienna, Austria: R Foundation for Statistical Computing. URL http://www.r-project.org/.

Sankoff, G. (2006). Age: Apparent time and real time. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., pp. 110–116). Amsterdam: Elsevier.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics, 52*, 183–204.

Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *Journal of the Acoustical Society of America, 132*, EL95–EL101.

Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology, 23*, 287–308.

Simon, E. (2009). Acquiring a new second language contrast: an analysis of the English laryngeal system of native speakers of Dutch. *Second Language Research, 25*, 377–408.

Slis, I. H., & Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction. *Language and Speech, 12*, 80–102.

Sun, J., Loader, C., & McCormick, W. P. (2000). Confidence bands in generalized linear models. *The Annals of Statistics, 28*, 429–460.
Svantesson, J.-O., & House, D. (2006). Tone production, tone perception and Kammu tonogenesis. *Phonology, 23*, 309–333.
Tagliamonte, S. (2012). *Variationist sociolinguistics: Change, observation, interpretation*. Oxford: Wiley-Blackwell.
Thurgood, G. (1996). Language contact and the directionality of internal drift: The development of tones and registers in Chamic. *Language, 72*, 1–31.
Thurgood, G. (2002). Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica, 19*, 333–363.
van Alphen, P. M. (2004). *Perceptual relevance of prevoicing in Dutch*. Katholieke Universiteit Nijmegen.
van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics, 32*, 455–491.
van Wyk, E. B. (1977). *Praktiese Fonetiek vir Taalstudie: 'n Inleiding*. Durban: Butterworths.
Verhoeven, J. (2005). Belgian Standard Dutch. *Journal of the International Phonetic Association, 35*, 243–247.
Vietti, A., Alber, B., & Vogt, B. (to appear). Initial laryngeal neutralization in Tyrolean. *Phonology*.
Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America, 73*, 1322–1336.
Wagner, S. E. (2012). Age grading in sociolinguistic theory. *Language and Linguistics Compass, 6*, 371–382.
Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America, 93*, 2152–2159.
Wissing, D. (1982). *Algemene en Afrikaanse Generatiewe Fonologie*. Johannesburg: Macmillan.
Yu, A. C. L. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 201–227). Oxford: Oxford University Press.
Xu, Y. (1999). F0 peak delay: Where, when, and why it occurs. In *Proceedings of the XIVth international congress of phonetic sciences* (pp. 1881–1884).
Xu, Y. (2001). Fundamental frequency peak delay in Mandarin. *Phonetica, 58*, 26–52.