

ARTICLE

Women of All Ages Lead Tonogenesis in Afrikaans

Alexandra M. Pfiffner 

Department of Linguistics, University of California, Berkeley, USA
Email: apfiffner@berkeley.edu

Abstract

The development of a sound change can be influenced by linguistic and social factors, both within the language community and from cases of language contact. The present study is an examination of the internally generated ongoing tonogenesis process in Afrikaans, specifically analyzing production and perception of word-initial plosives among different age and gender groups. Results show that female speakers are devoicing significantly more often than male speakers, and the perception of female voices is influenced more by f_0 levels than the perception of male voices. This study finds that gender is a larger predictor overall of tonogenetic patterns than age.*

Keywords: voicing; phonological tone; Afrikaans; social factors; phonological change; speech perception; speech production

1. Introduction

Word-initial plosives in Afrikaans are undergoing voicing neutralization; what was previously prevoiced /b, d/ is frequently produced as voiceless unaspirated [p, t] (Coetzee et al. 2018). However, this neutralization is not resulting in a complete loss of phonological contrast; instead, the phonetic cues to obstruent voicing are allowing tone to emerge in a process of tonogenesis, where phonological tone arises in a previously non-tonal language (Hombert 1975).

Phonological voicing in plosives is phonetically implemented through numerous cues, both temporal and spectral. Word-initially, this includes voice onset time (VOT) as well as f_0 and F1 transitions into the vowel (Kohler 1984, Lisker 1986). There are

*Thank you, first and foremost, to all of the Afrikaans speakers who participated in this study. Thank you also to Daan Wissing and colleagues at North-West University for help with logistics, Petri Swanepoel, Roné Wierenga, Christo Alberts, and Mart Bezuidenhout-Swanepoel for help with data collection, and Harm Boer and Wesley Biggs for help with stimuli. I would also like to express my immense gratitude to Frans Hinskens and two anonymous reviewers whose comments greatly improved this work. This research was supported by NSF Doctoral Dissertation Research award BCS-1918306. All errors are my own.

numerous theories as to why this is, including articulatory differences in the degree of vocal fold tension (see, e.g., Hombert et al. 1979, Löfqvist et al. 1989), controlled perceptual enhancement of phonological contrasts (Kingston & Diehl 1994), and/or aerodynamic reasons, where the higher volume in transglottal flow in voiceless plosives results in faster vocal fold vibration (Ohala 1973, Hombert et al. 1979). Regardless of their origin, the existence of these cues allows for language users to perceive and actively control them, phonologizing the phonetic effect (Kingston & Diehl 1994). This relationship between laryngeal contrasts and pitch is widely attested in the world's languages, and many established tonal systems can be traced to historical obstruent voicing contrasts, such as the tones in Vietnamese (Haudricourt 1954, Yip 2002). Synchronically, the beginning stages of this process have been most recently reported in Seoul Korean (Kang & Han 2013, Bang et al. 2018), Afrikaans (Coetzee et al. 2018), and Dutch (van Alphen & Smits 2004, Pinget et al. 2019, Pfiffner 2021a,b).

This type of tonogenetic development in Afrikaans and Dutch stands in contrast to other types of tonal phenomena seen in Germanic languages, namely, tonal accent systems. Numerous varieties of North Germanic and West Germanic display lexical contrasts through different pitch trajectories associated with the stressed syllable, known as “pitch accent” or “tonal accent” languages (Iosad 2024). Accentual languages differ from tonal languages in that typically there is one lexically specified tone or tonal complex per word, whereas a tonal language has tone melodies that can map to numerous tone-bearing units (Yip 2002). Additionally, in the case of Germanic, the tonal accent developed out of manipulation of metrical structures versus historical obstruent voicing contrasts (Iosad 2024). Given that Afrikaans and Dutch are West Germanic languages, their development of a tonal system from laryngeal contrasts, instead of an accentual system from metrical structures, is notable.

In the case of Afrikaans, Coetzee and colleagues (2018) point out that the language has had extensive contact with local tonal languages, including Khoisan and Bantu languages, which could be a motivating factor in developing tone. However, they also note that this is unlikely given the social, economic, and political dynamics of South African society. Furthermore, the fact that Dutch is also showing variable word-initial neutralization, along with perceptual reweighting of prevoicing and f_0 cues, suggests that this cannot be a contact-induced change alone (Pfiffner 2021a). Instead, the change must also be motivated by language-internal factors.

Since tonogenesis in Afrikaans has only been reported for the first time recently, much work remains to be done to gain a fuller picture of the process and to explore what is motivating this change-in-progress. The present article aims to contribute to the literature by analyzing social differences in the production and perception of word-initial plosives in Afrikaans, providing further evidence that this is a case of tonogenesis and not simply age-graded variation.

1.1. Tonogenesis in Afrikaans

Word-initial plosive voicing in Afrikaans was traditionally characterized as a difference between prevoicing (glottal pulsing during the closure) and short-lag VOT (Wissing 2018). However, voiced plosives are frequently realized as their voiceless unaspirated counterparts, especially in younger generations (Coetzee et al. 2018,

Wissing 2018). Coetzee and colleagues (2018) examined devoicing and the emergence of tonogenesis in two populations: female speakers ages 21–23, and female speakers in their 40s through 60s. In a production experiment, they measured VOT and the f_0 of the following vowel at the boundaries of 20 equidistant intervals. Their results showed that younger speakers had a much higher devoicing rate, as they devoiced underlyingly voiced plosives approximately 83 percent of the time, while older speakers devoiced 44 percent of the time (Coetzee et al. 2018:192). In both populations, there were also clear f_0 differences in the following vowels, with lower f_0 s after underlyingly voiced plosives regardless of whether they were realized as prevoiced (with negative VOT) or devoiced (with positive VOT), and higher f_0 s after underlyingly voiceless plosives. Notably, the f_0 effects remained through the entire duration of the vowel. This suggests that a cue reweighting has occurred, where the secondary cue of f_0 has strengthened, and the once primary cue of glottal pulsing is now variably used by speakers.

The same participants then took part in a perception experiment where they heard twelve continua varying in VOT and the following vowel's f_0 , with each end of the continuum being one member of a minimal pair (e.g. [bas] '(tree) bark' and [pas] 'just now'). All tokens were created from natural utterances from a young female speaker of Afrikaans. Results show that all participants relied on both f_0 and prevoicing to identify the word-initial plosive. However, when the two conflicted, older participants relied more on VOT than younger participants. When comparing perception and production results, there is evidence for a pattern in that older speakers are more likely to produce the underlyingly voiced plosives with glottal pulsing, and glottal pulsing is a more highly weighted cue in perception for older speakers versus younger speakers. Thus, this could be evidence for a diachronic case of cue-transfer, as different generations are more heavily weighing different cues in both perception and production.

1.2. Social factors in sound change

Generational differences are not unexpected; sociolinguistic research has long shown that younger speakers tend to use more innovative variants than older speakers (Labov 1980, Milroy & Milroy 1985). In the case of Afrikaans, however, the question is whether this is simply age-graded variation or an ongoing sound change, altering the language's phonology. Comparing different age and gender groups would provide further evidence one way or another.

If this is age-graded variation, then we wouldn't expect differences in production and perception by gender. If this is a change in progress, then we would expect gender differences in production, as women tend to lead sound change by using innovative variants more often than men (Labov 2001). This has also been seen in the context of Korean tonogenesis; in production, women are more advanced in terms of their use of VOT and f_0 (Oh 2011, Kang 2014).

Speech perception data from diverse social groups would also provide further evidence for tonogenesis or age-graded variation. It is well known that speech perception is influenced by social characteristics (Niedzielski 1999, Hay et al. 2006, Hay & Drager 2010), both the social characteristics of a listener and perceived characteristics of a speaker (Hay et al. 2006). These social effects have also been seen

in the case of Seoul Korean tonogenesis, as listeners rely less on VOT and more on f_0 when listening to a female voice (Kong et al. 2011).

The previous research on Afrikaans is limited to female speakers only and the perception of a young female voice. To more completely understand the situation in Afrikaans and whether or not this is a change in progress or age-graded variation, production and perception data are needed from both male and female speakers, as well as a comparison of how listeners perceive female versus male voices.

1.3. Research questions and predictions

The current study fills the gap in the literature with production and perception data from four different age and gender groups. The research questions and hypotheses are as follows:

RQ1: Do male speakers of Afrikaans produce word-initial plosives differently from female speakers?

- H1: If this is a case of tonogenesis, then female speakers will devoice more than male speakers and have larger differences in the following vowel's f_0 to compensate for the loss of glottal pulsing.
- H2: If this is a case of age-graded variation, then there will not be systematic differences by gender.

RQ2: Are there differences in the perception of word-initial plosives based on the gender of the speaker and/or the gender of the listener?

- H3: Since women are leaders of sound change, then if this is a case of tonogenesis, listeners will be more sensitive to f_0 when hearing female voices in comparison to when they hear male voices. In terms of the listener's demographics, female listeners will be more sensitive to f_0 than male listeners.
- H4: For age-graded variation, we would not expect differences in perception based on speaker and/or listener gender, but rather age. It is hypothesized that listeners will instead be more sensitive to f_0 when listening to younger voices in comparison to older voices, and further that older listeners will be less sensitive to f_0 than younger listeners.

RQ3: Are there any interactions between age and gender?

- H5: In a case of ongoing sound change, we would expect differences both by gender and by age. It is therefore hypothesized that younger female participants will show the most devoicing and sensitivity to f_0 , and older male participants will show the least devoicing and sensitivity to f_0 . In the middle are younger male participants and older female participants. If gender is a bigger driving factor of this change, then we'd expect to see older female speakers patterning closer to younger female speakers. If age is the larger factor, then we'd expect older female participants to pattern closer to older male participants.

2. Methods

2.1. Participants

Thirty-four white native speakers of Afrikaans living in Potchefstroom, South Africa, took part in the study. Twenty speakers were ages 20–24 ($f=10$, $m=10$), and fourteen speakers were ages 60–83 ($f=7$, $m=7$). None reported any hearing or speaking difficulties. No information was collected on other languages spoken by the participants. For most participants, the study took place in a sound-attenuated booth at North-West University, though a few participants took part at a local senior citizen's activity center in a quiet conference room. Participants were paid for their time.

2.2. Procedure

The experiment was run with PsychoPy (Peirce et al. 2019). All participants completed the production task first to avoid any effects of accommodation from the perception task (Giles et al. 1973, Goldinger 1998). The experiment was conducted entirely in Afrikaans, and any verbal instructions were given in Afrikaans by a native-speaker research assistant to limit experimenter accommodation (Hay et al. 2009, 2010).

During the production task, participants were recorded with a cardioid condenser lavalier microphone and a Zoom H4n handheld recorder at a sampling rate of 44.1 kHz. Participants read aloud a randomized word list twice with 20 target stimuli and 55 fillers. One word at a time appeared on a laptop screen, and participants had two seconds to read the word out loud before the next word appeared. This was done to ensure an isolated phonological environment, control speech rate, and limit list intonation. After the word list, participants additionally read passages, which are not discussed here. In total, the production portion took 10–15 minutes to complete.

The perception task had four blocks, with each block presenting a different speaker voice. Details of the speakers are given in the following section, and stimuli measurements based on their recordings are given in table 2. Participants were told that the speakers were either 20 years old or 60 years old. In the analysis, each speaker voice is codified by their gender and apparent age (e.g. F20 for the 20-year-old female speaker).

The order of the speaker blocks was randomized for each participant. Before each block, participants were told the age and gender of that speaker (e.g. *In hierdie gedeelte gaan jy na 'n sestigjarige vroulike Moedertaalspreker van Afrikaans luister*. 'In this block, you will hear a 60-year-old female speaker of Afrikaans'). A two-alternative forced-choice task was run with AKG K701 Harman Premium class performance headphones. Within each block, a token automatically played, and two words were shown on the screen. Participants had to choose which word they heard (e.g. *pad* or *bad*). Participants were encouraged to move quickly, but not so fast as to make mistakes. They were also told to make their best guess if they were unsure. Each speaker block contained 400 total tokens, 200 of which were stimuli (4 types of continua \times 5 steps \times 2 place of articulation base tokens \times 5 repetitions), so there were 800 stimuli in total per listener. The remaining 200 tokens consisted of words with final plosives, which were used as stimuli for a different experiment that is not discussed here. All tokens were randomized within each speaker block. Participants were given breaks every 40

Table 1. Stimuli used in the production task

Vowel	Voiceless			Voiced		
/i/	piek	/pik/	'peak'	bieg	/bix/	'confess'
/ɛ/	perk	/perk/	'limit'	berk	/bɛrk/	'birch tree'
/a/	pad	/pad/	'road'	bad	/bad/	'bath'
/ɔ/	pol	/pɔl/	'hassock'	bol	/bɔl/	'ball'
/u/	poel	/pul/	'pool'	boel	/bul/	'bulldog'
/i/	tien	/tin/	'ten'	dien	/din/	'be of service'
/ɛ/	telg	/tɛlx/	'descendent'	delg	/dɛlx/	'paying off debt'
/a/	tal	/tal/	'quantity'	dal	/dal/	'valley'
/ɔ/	tol	/tɔl/	'top'	dol	/dɔl/	'crazy'
/u/	toe	/tu/	'closed'	doel	/dul/	'purpose'

Table 2. Speaker measurements used to create the perception continua

Speaker	Base token	Lowest onset f0	Highest onset f0	Mean prevoicing
F20	pad	155 Hz	243 Hz	106 ms
	tal	174 Hz	245 Hz	
M20	pad	95 Hz	127 Hz	105 ms
	tal	88 Hz	134 Hz	
F60	pad	180 Hz	231 Hz	117 ms
	tal	184 Hz	226 Hz	
M60	pad	76 Hz	110 Hz	83 ms
	tal	77 Hz	111 Hz	

tokens and in between each speaker block. The perception task took approximately 45–55 minutes to complete, including breaks.

2.3. Stimuli

The production stimuli were 20 monosyllabic (near-)minimal pairs (table 1). The words were balanced for place of articulation (bilabial and alveolar) and following vowel. The perception task had two sets of artificially manipulated continua that varied the amount of prevoicing and the following vowel's f0. The bilabial set was made with /pad/ 'road' and /bad/ 'bath', with a natural utterance of /pad/ from a native speaker as the base. The alveolar set was made with /tal/ 'quantity' and /dal/ 'valley', with /tal/ as the base. Each continuum had a total of five steps. In continuum types 1–3, one cue was held steady and one cue changed. In continuum type 4, both

cues changed and were conflicting (i.e. highest f0 was paired with 100% prevoicing). Descriptions of the continua are given below:

1. No prevoicing; f0 changes from low (step 1) to high (step 5)
2. Full prevoicing; f0 changes from low (step 1) to high (step 5)
3. Prevoicing changes from 0% (step 1) to 100% (step 5); f0 is constant and at mid-range
4. Prevoicing and f0 conflict:
 - 0% prevoicing and low pitch, since [-voice] typically patterns with high f0
 - 25% prevoicing and medium-low pitch
 - 50% prevoicing and medium pitch
 - 75% prevoicing and medium-high pitch
 - 100% prevoicing and high pitch, since [+voice] typically patterns with low f0

The tokens were created based on recordings of four native speakers, two female speakers (ages 20 and 61) and two male speakers (ages 37 and 56). All speakers were living in Washington, DC, at the time of recording and self-identified as speakers of “Standard Afrikaans.” Speakers were recorded in a sound-attenuated booth reading a word list multiple times. The duration of any prevoicing was measured, and then means and ranges were calculated for each speaker. For each token, the f0 of the following vowel was measured at 10 percent increments throughout the vowel. A clear, modal token of /pad/ and /tal/ with the highest f0 at onset was selected for each speaker. Similarly, a clear, modal token of /bad/ and /dal/ with the lowest f0 at onset was selected for each speaker. The two tokens (per speaker) were normalized for duration, and the entire vowel contour from the highest f0 at onset was used as the fifth step of any continua with f0 changing (continua 1, 2, and 4). This became the base token. Using Praat’s PSOLA function (Boersma & Weenink 2019), the pitch tier was extracted from the base token and replaced with the pitch tier of the token with the lowest f0 at onset, creating the bottom step of continua 1, 2, and 4. Three equidistant steps were calculated and created between the two end contours, totaling five steps in the continua. The lowest (step 1) and highest (step 5) f0 for each base token and each speaker are given in table 2.

To create the prevoicing continuum, each speaker’s mean prevoicing duration was calculated. A token of /bad/ and /dal/ with clear prevoicing was selected from each speaker. The duration of the prevoicing was scaled to be the speaker’s mean, and this was spliced onto the third step of the f0 continua of /pad/ or /tal/, respectively. This token became step 5 of the prevoicing continuum, and the duration of prevoicing was scaled down in 25 percent increments, creating a continuum from no prevoicing to full prevoicing (0–25%–50%–75%–100%). To create the fourth continuum, where prevoicing and f0 conflict, scaled prevoicing durations from continuum 3 were spliced onto individual tokens from the first continuum (steps 2 through 5). Finally, all tokens were normalized for intensity at 70 dB.

2.4. Analysis

Beginning with production data, a total of 1,360 tokens (40 x 34 participants) were annotated and then aligned with the Montreal Forced Aligner (McAuliffe et al. 2017).

Segment boundaries were hand-corrected, and underlyingly voiced plosives were coded as to whether or not prevoicing was present. Prevoicing was determined to begin at the start of visible periodic voicing, no matter the amplitude, and the end of prevoicing was indicated by the plosive burst. This included tokens that had prevoicing followed by a period of voicelessness before the burst. All following vowels were measured for f_0 at eleven time points from onset to offset, totaling 14,960 measurements. A total of 860 measurement points were excluded because the Praat pitch-tracking algorithm was unable to determine an f_0 . An additional 8 tokens were excluded for mispronunciation or noise in the recording. All f_0 measurements were then analyzed by gender, with group means and standard deviations calculated. Outliers beyond 2.5 standard deviations for each gender group were excluded. All f_0 measurements were then z-normalized and converted back to Hertz-like measurements for readability using the mean and standard deviation for all speakers ($\text{group mean} + \text{z-score} * \text{group SD}$; see Brunelle et al. 2020:8). The production data was then fitted with mixed-effects regression models, as detailed below.

For the perception data, individual responses in the perception task were coded as 0 if the participant chose the voiceless option (*pad* or *tal*), and 1 if they chose the voiced option (*bad* or *dal*). Each continuum was separately fitted with a mixed-effects logistic regression model in a step-up-step-down procedure using the *glmer* function of the *lme4* package (Bates et al. 2015) in R (R Core Team 2013), with the dependent variable being token choice (0 or 1). The fixed effects that were tested include (as relevant for each continuum) f_0 level (1–5), prevoicing level (1–5), continuum (*pad* or *tal*), listener age (younger or older), listener gender (female or male), speaker age (younger or older), speaker gender (female or male), and interactions between age and gender, for both speakers and listeners. Participant was included as a random effect. Full model outputs, for both production and perception, can be found in Pfiffner (2021a).

3. Results

3.1. Production

The percentage of underlyingly voiced tokens produced with glottal pulsing was highly individual; 12 participants devoiced approximately 70–100 percent of tokens, and all others devoiced less than 50 percent of tokens. Three younger speakers, two female and one male, devoiced all underlyingly voiced tokens, while one younger male speaker did not devoice at all. The groups of “voicers” and “devoicers” had at least one participant from each of the four demographic groups. These groupings and individual participants’ devoicing means are shown in figure 1. With this bimodal distribution, the average proportion of devoicing comes out at 46 percent of tokens.

The data was fitted with a mixed-effects logistic regression model to predict (1) devoicing or (0) maintenance of prevoicing. Fixed effects were speaker age, speaker gender, the interaction of age and gender, and place of articulation. Random effects were the token word and the participant number. Only gender was significant ($p=0.04$).

Time-normalized and z-score transformed trajectories of the following vowel are shown in figure 2. Overall, vowels following underlyingly voiceless plosives were produced with higher f_0 s than vowels following underlyingly voiced plosives, regardless of their voicing realization (i.e. prevoiced or devoiced). At onset, the f_0 s of vowels following devoiced plosives were also consistently higher than vowels

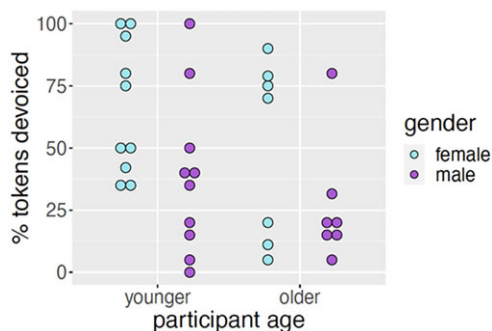


Figure 1. Percentage of underlyingly voiced tokens that were devoiced. Each dot represents one participant.

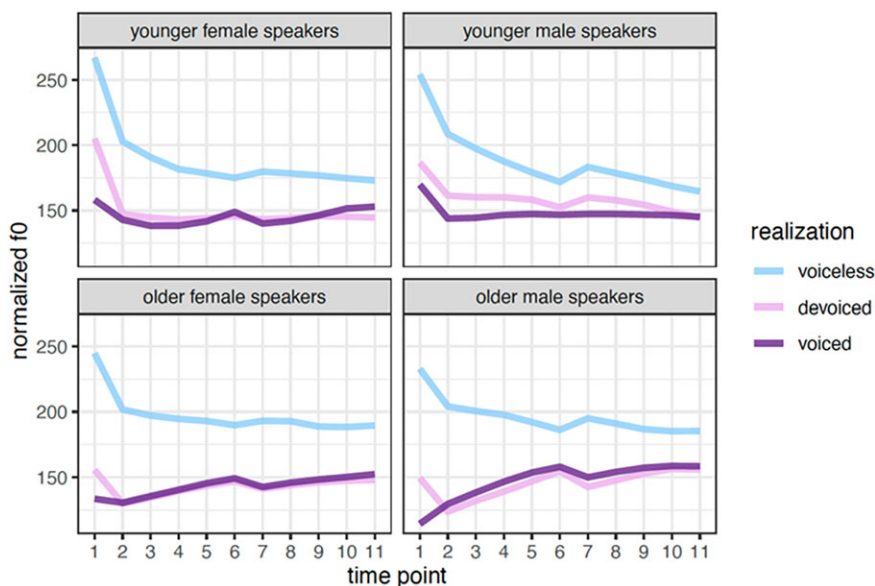


Figure 2. Time-normalized and z-score transformed f_0 trajectories of vowels following initial plosives. The different colors represent vowels following voiceless /p, t/, devoiced /b, d/, and prevoiced /b, d/ plosives. Time point 1 is the onset of the vowel, and time point 11 is the offset. Each grid shows a different speaker group.

following prevoiced plosives, but their average f_0 immediately lowered and was approximately even with the prevoiced realizations by the second time point, which is 10 percent of the way through the vowel. The lessened perturbation at onset may be partially attributed to articulatory reasons, but the decrease in f_0 and the maintenance of the f_0 difference across the vowel suggest that this is a controlled cue.

To test for significance, a linear mixed-effects model was fitted to the normalized f_0 data. The independent variable was normalized f_0 at onset (time point 1 in figure 2). Fixed effects were realization (voiceless, devoiced, voiced), place of articulation, speaker age, speaker gender, and the interaction between speaker age and gender. Participant number was a random effect. There were significant differences between

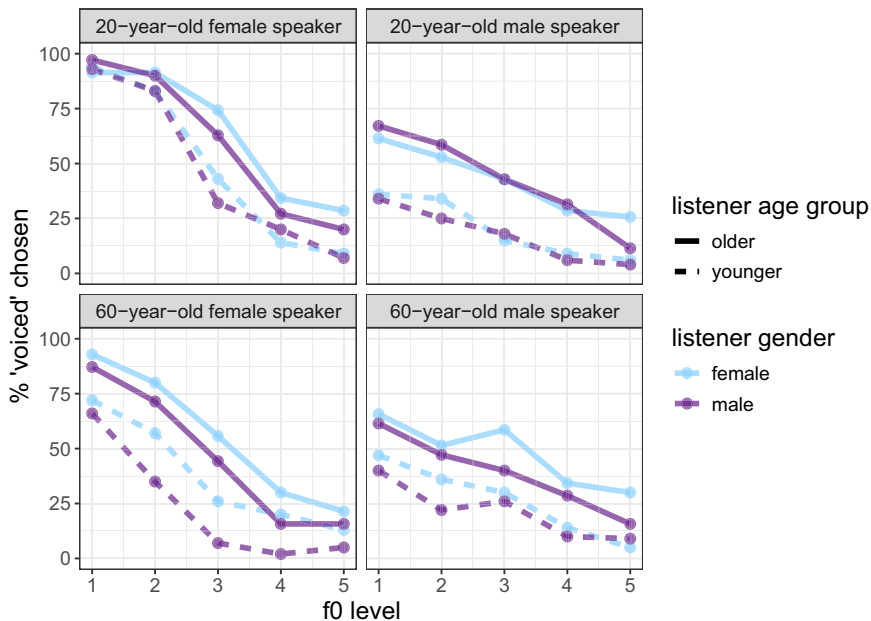


Figure 3. Perception of continua with no prevoicing and f0 changing from low (1) to high (5) in equal steps. Each grid shows younger versus older listeners' responses to each speaker voice.

f0 at onset following devoiced and voiceless plosives ($p < 2e-16$), with f0s following devoiced plosives being significantly lower, and between f0 at onset following devoiced and voiced plosives ($p < 0.001$), with f0 following devoiced stops being significantly higher. Age was also significant ($p = 0.009$); younger participants had overall higher f0s following prevoiced and devoiced plosives. None of the other factors were found to be significant.

3.2. Perception

Continuum 1: No prevoicing

The first type of continuum had no prevoicing, and f0 changed in five equal steps from low (1) to high (5). There was a significant effect of f0 ($p < 2e-16$) for all listener groups and speaker voices: Lower f0s were more often perceived as underlyingly voiced, even in the absence of prevoicing, and higher f0s were often perceived as underlyingly voiceless (figure 3). There were significant differences between listener age groups ($p = 1.62e-0$), as younger participants in general were more likely to perceive a token as underlyingly voiceless. This is particularly noticeable in the perception of the younger male speaker, older female speaker, and older male speaker, suggesting that for these voices, younger listeners are expecting prevoicing to signal underlying voicing. Compare that to the perception of the younger female speaker, where a low f0 cues voicing at ceiling.

There were other differences between the perception of the four different speakers; speaker age ($p < 2e-16$), speaker gender ($p = 2.84e-0$), and their interaction

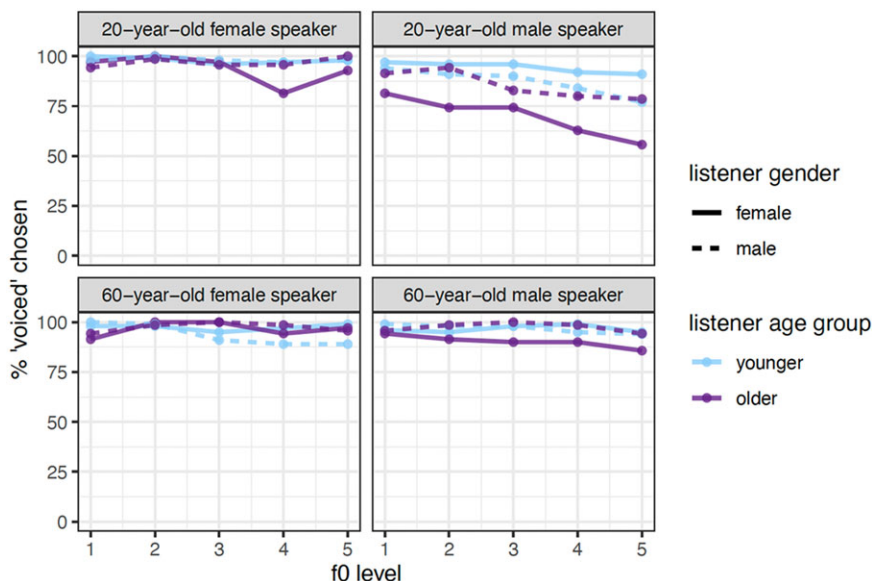


Figure 4. Perception of continua with full prevoicing and f_0 changing from low (1) to high (5) in equal steps. Each grid shows younger versus older listeners' responses to each speaker voice.

($p < 2e-16$) were all significant predictors in the perception of underlying voicing. Noticeably, the perception of tokens from both female speakers was more strongly influenced by f_0 levels than tokens from their male counterparts. The perception of the young female speaker was the most influenced by f_0 levels.

Continuum 2: Full prevoicing

The second type of continuum had full prevoicing (100%) while f_0 changed in five equal steps from low (1) to high (5). The results show that prevoicing is still a dominant cue for perceiving word-initial voicing contrasts. Regardless of f_0 level, nearly all tokens were perceived as voiced (figure 4). One exception can be seen for listeners hearing the 20-year-old male speaker. Older female listeners perceived the lowest f_0 level as voiced 81 percent of the time, and the highest f_0 level as voiced 56 percent of the time. This 25 percent decrease shows that higher f_0 levels lead to ambiguity, with listener responses nearly at chance. The downward trend is seen in all listener groups, though to a much smaller magnitude among the younger listeners and the older male listeners.

The f_0 level was a significant predictor of listener responses ($p = 2.97e-1$), though this is likely to be due to the pattern seen for the 20-year-old male speaker voice, as the speaker age and gender interaction was significant ($p = 3.22e-1$), but neither factor was individually significant. There were also significant differences between listener age ($p = 6.54e-0$), listener gender ($p = 0.03$), and their interaction ($p = 0.0005$). This is again likely to be weighted by the older female listeners' perception of the 20-year-old male speaker.

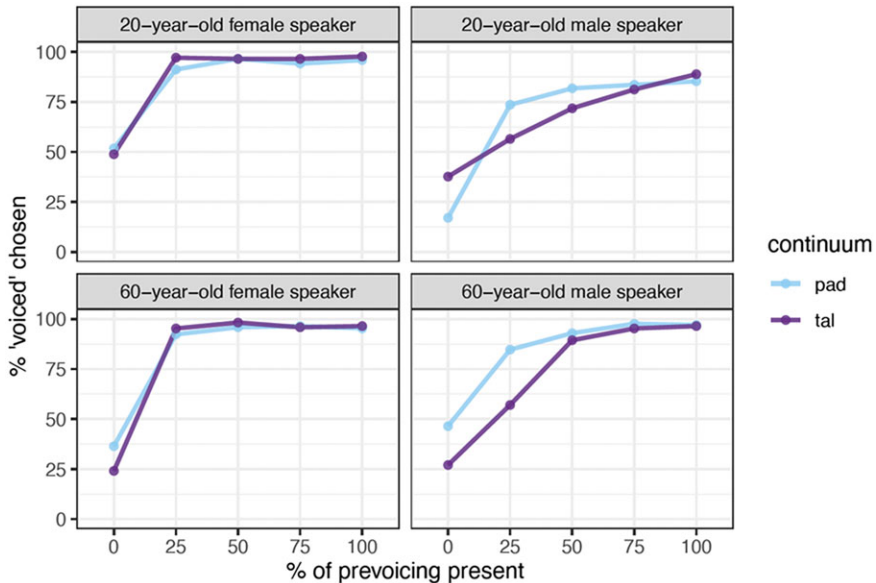


Figure 5. Perception of the bilabial and alveolar continua with ambiguous f_0 and prevoicing changing from 0 to 100 percent in equal steps. Each grid shows a different speaker voice.

Continuum 3: Ambiguous f_0 , prevoicing changing

The third type of continuum had f_0 held at mid-range (equal to step 3 of the f_0 -changing continua), which was meant to be ambiguous in signaling underlying voicing. The percentage of prevoicing varied from none (0%) to full (100%) in 25 percent increments. Overall, no prevoicing led to “voiced” judgements from 17–69 percent of tokens (figure 5). Older listeners had average “voiced” judgments nearly at chance (43–69%). Additionally, the difference between no prevoicing and each prevoicing level was significant ($p < 2e-16$), as the presence of prevoicing led to the perception of voiced stops. There was also a slight difference by continuum, as the bilabial tokens were more likely to be perceived as voiced at prevoicing levels 25–50 percent in comparison to the alveolar tokens. This difference was significant ($p = 6.03e-1$).

No listener characteristics were significant for this type of continuum. However, there were significant differences by speaker age ($p < 0.0001$), speaker gender ($p < 0.0001$), and their interaction ($p < 2e-16$). Essentially, any amount of prevoicing led to the perception of voiced plosives when listening to both female speakers, but this pattern was not seen with the male speakers. The perception of the 60-year-old male speaker hit ceiling at approximately 50 percent prevoicing, while perception of the 20-year-old male speaker reached ceiling at full prevoicing.

Continuum 4: F_0 and prevoicing in conflict

The fourth continuum type had both f_0 and prevoicing changing, and the two were in conflict. The lowest f_0 was paired with no prevoicing, while the highest f_0 was paired

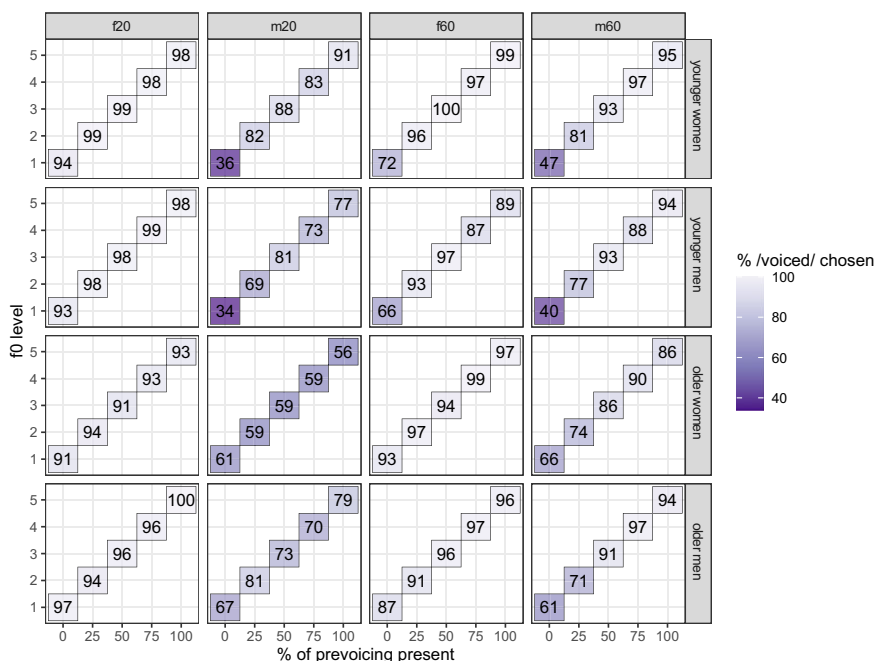


Figure 6. Listener groups' perception of continua with prevoicing and f_0 in conflict. Each column shows a different speaker voice, codified by their gender (f or m) and apparent age (20 or 60). Each row shows a different listener group.

with full prevoicing. The results of these continua (figure 6) show that prevoicing is still the dominant cue; when prevoicing is present, listeners overwhelmingly report hearing a voiced token; this is true for all listener groups and speaker voices. However, there is still an effect of f_0 , as tokens with no prevoicing but low f_0 s were still perceived as voiced in high amounts, especially when listening to the 20-year-old female speaker ("f20" in figure 6). Additionally, some of the highest percentages of perceived voiced responses are at steps 3 or 4, and the percentage of voiced responses decreases at a higher step. For example, younger male listeners perceived the 60-year-old female speaker ("f60") as voiced 97 percent of the time at step 3, but 87 percent at step 4 and 89 percent at step 5. While this is a small difference, it might suggest that higher f_0 s can occasionally outrank high amounts of prevoicing. One final point concerns gender: Non-prevoiced tokens from the two male speakers were perceived as voiced less often in comparison to non-prevoiced tokens from the two female speakers.

Numerous factors were significant; crucially, the differences between 0 percent prevoicing and all other levels of prevoicing were highly significant ($p < 2e-16$). There was also a significant difference between the two continua ($p = 2.85e-05$), as all tokens in the alveolar continuum were less often perceived as "voiced." Social characteristics were also significant. First, there was a significant interaction between listener age and gender, as younger male speakers overall perceived tokens as less voiced ($p = 0.0131$). Additionally, speaker age ($p = 1.91e-08$), speaker gender ($p < 2e-16$), and their interaction ($p < 2e-16$) were all significant.

4. Discussion

Returning to the research questions, the results of this experiment provide further evidence that the neutralization of the plosive voicing contrast in Afrikaans is leading to tonogenesis. While there are clear differences in production and perception by age, reinforcing the results of Coetzee et al. (2018), there are also significant differences by gender, answering **RQ1**. On average, female speakers, young and old, devoiced underlyingly voiced plosives significantly more often than male speakers. In terms of f_0 , all speakers had significant and reliable f_0 differences, not only between underlyingly voiced and voiceless plosives, but also between underlyingly voiceless and surface devoiced plosives. Thus, even in the absence of prevoicing, the cue to phonological voicing in the f_0 differences remained steady between generations, further supporting Coetzee and colleagues' (2018) interpretation of f_0 as an already-strengthened cue.

RQ2 asks if there are differences in the perception of word-initial plosives based on gender. The results show that speaker gender is a significant factor in perception; in the absence of prevoicing, f_0 effects were stronger in the perception of female speaker voices versus male speaker voices. This echoes findings in the perception of Seoul Korean stops undergoing tonogenesis, where Kong et al. (2011) found that listeners were more sensitive to f_0 when attending to a female voice. In Afrikaans, there was also a significant gender difference when f_0 was ambiguous and prevoicing changed; listeners perceived any amount of prevoicing in the female voices as "voiced," while there was some ambiguity when the male voices had prevoicing levels of 25–50 percent. When considered in conjunction with the f_0 results, it appears that listeners are anticipating f_0 differences to signal voicing in the speech of female speakers, so the presence of prevoicing further enhances the perception of voicing. Finally, these effects of gender were only seen in relation to speaker voices; there were no consistent patterns regarding the listener's gender.

RQ3 asks if there are interactions between age and gender. In a case of sound change, we expect younger speakers to use the innovative variant more than older speakers, but at the same time, we know that female speakers tend to lead sound change. The results of this study show there is a clear interaction of age and gender. Female speakers of all ages devoiced more than male speakers, which supports the hypothesis that women lead sound change, but at the same time, younger male speakers devoiced more than older male speakers. Hypothesis 1 further predicted that gender would be a factor in the magnitude of f_0 differences at onset of the vowel following the plosive. However, this prediction was proven false, as older speakers had larger differences in f_0 than younger speakers. Perhaps this is a case of age-graded variation; generational cue flux, where different age groups have different cue weightings for the same phonological contrast, has been seen before in cases of stable sociophonetic variation (Pfiffner 2021a, forthcoming).

In terms of perception, it was hypothesized that if there was an interaction of age and gender, then we should see distinctions by group regarding sensitivity to f_0 as a cue to underlying voicing, with younger female speakers being the most sensitive, followed by younger male and/or older female speakers, and lastly older male speakers. This was not seen in the results; overall, the social characteristics of the listeners did not appear to impact their perception. On the other hand, the apparent

social characteristics of the speaker voices were frequently significant. In particular, f_0 was used as a cue to voicing more often when perceiving any female voice, young or old.

Taken together, the results of this study show that both age and gender are significant factors in the production and perception of word-initial plosives in Afrikaans. This provides further evidence for tonogenesis in progress, with women of all ages leading the sound change.

5. Conclusion

This study asked about the relationship between age and gender in the perception and production of word-initial plosives in Afrikaans. What was a historical obstruent voicing contrast is now variably produced, and with a strengthened f_0 cue; this suggests that this is a case of tonogenesis, where the secondary f_0 cues of the original obstruent voicing contrast are being rephonologized as tone. Previous research has been limited to the production and perception of this contrast in female speakers, so the present study fills the gap by contributing information on both age and gender differences. The results ultimately show that gender and age interact; while there are generational differences between older (ages 60+) and younger (ages 20–24) speakers, reinforcing the age differences in female speakers reported by Coetzee et al. (2018), gender appears to be a larger factor, as older and younger female speakers on average devoiced more than male speakers. Additionally, listeners are more sensitive to f_0 as a cue when hearing female voices, young or old, in relation to male voices. Overall, the evidence supports the interpretation of Afrikaans undergoing tonogenesis rather than simply having age-graded differences in initial obstruent production and perception.

The findings allow us to understand more about how this voicing neutralization and the emergence of tonogenesis is spreading through the community. The change in progress is internally motivated, as it arises from a reweighting of acoustic cues for a particular phonological contrast, and it is also affected by social factors within its own community of speakers. Women are using the more innovative variants by devoicing, and listeners are paying more attention to f_0 when they hear a female voice.

This study is not without limitations; in particular, both this experiment and that of Coetzee et al. (2018) examined the phenomenon in Potchefstroom, South Africa. While examining the same relative population provides an opportunity to easily compare results, it is also not a full view of all the dialects of Afrikaans, but rather a representation of one small part of the Afrikaans-speaking population. In future work, more communities should be studied, and additional social factors such as ethnicity, region, and socioeconomic status should be taken into consideration. Finally, natural conversational data should be examined, as previously only word lists and read speech have been collected and analyzed.

References

- Alphen, Petra M. van & Roel Smits. 2004. Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics* 32, 455–491.

- Bang, Hye-Young, Morgan Sonderegger, Yoonjung Kang, Meghan Clayards, & Tae-Jin Yoon. 2018. The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics* 66, 120–144.
- Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48.
- Boersma, Paul & David Weenink. 2019. *Praat: Doing phonetics by computer* [Computer program] (version 6.1). www.praat.org (accessed July 13, 2019).
- Brunelle, Mark, Tạ Thành Tấn, James Kirby, & Đinh Lu Giang. 2020. Transphonologization of voicing in Chru: Studies in production and perception. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1), 1–33.
- Coetzee, Andries W., Patrice Speeter Beddor, Kerby Shedden, Will Styler, & Daan Wissing. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66, 185–216.
- Giles, Howard, Donald M. Taylor, & Richard Y. Bourhis. 1973. Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society* 2(2), 177–192.
- Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2), 251–279.
- Haudricourt, André-Georges. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 242, 69–82.
- Hay, Jennifer & Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48(4), 865–892.
- Hay, Jennifer, Katie Drager, & Paul Warren. 2009. Careful who you talk to: An effect of experimenter identity on the production of the NEAR/SQUARE merger in New Zealand English. *Australian Journal of Linguistics* 29(2), 269–285.
- Hay, Jennifer, Katie Drager, & Paul Warren. 2010. Short-term exposure to one dialect affects processing of another. *Language and Speech* 53(4), 447–471.
- Hay, Jennifer, Paul Warren, & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4), 458–484.
- Hombert, Jean-Marie. 1975. Towards a theory of tonogenesis: An empirical, physiological, and perceptually-based account of the development of tonal contrasts in language. PhD thesis, University of California, Berkeley.
- Hombert, Jean-Marie, John J. Ohala, & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55, 37–58.
- Iosad, Pavel. 2024. Suprasegmental phenomena in Germanic: Tonal accent. In Sebastian Kürschner & Antje Dammel (eds.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.965>
- Kang, Yoonjung. 2014. Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics* 45, 76–90.
- Kang, Yoonjung & Sungwoo Han. 2013. Tonogenesis in early contemporary Seoul Korean: A longitudinal case study. *Lingua* 134, 62–74.
- Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70, 419–454.
- Kong, Eun Jong, Mary E. Beckman, & Jan Edwards. 2011. Why are Korean tense stops acquired so early? The role of acoustic properties. *Journal of Phonetics* 39(2), 196–211.
- Kohler, Klaus J. 1984. Phonetic explanation in phonology: The feature fortis/lenis. *Phonetica* 41, 150–174.
- Labov, William. 1975. On the use of the present to explain the past. In Luigi Heilmann (ed.), *Proceedings of the 11th International Congress of Linguists*, 825–851. Bologna: Il Mulino.
- Labov, William. 1980. *Locating language in time and space*. New York: Academic Press.
- Labov, William. 2001. *Principle of linguistic change*, vol. 2: *Social factors*. Malden, MA: Blackwell.
- Lisker, Leigh. 1986. “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29(1), 3–11.
- Löfqvist, Anders, Thomas Baer, Nancy S. McGarr, & Robin Seider Story. 1989. The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America* 85(3), 1314–1321.
- McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, & Morgan Sonderegger. 2017. *Montreal Forced Aligner* [Computer program] (Version 0.9.0). Retrieved May 1, 2020 from <http://montrealcorpus.github.io/montreal-forced-aligner/>
- Milroy, James & Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21, 339–384.

- Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18, 62–85.
- Oh, Eun Jong. 2011. Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics* 39, 59–67.
- Ohala, John J. 1973. The physiology of tone. In Larry M. Hyman (ed.), *Consonant types and tone* (Southern California Papers in Linguistics 1), 1–14. Los Angeles: University of Southern California.
- Peirce, Jonathan W., Jeremy R. Gray, Sol Simpson, Michael R. MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, & Jonas Kristoffer Lindeløv. 2019. Psychopy2: Experiments in behavior made easy. *Behavior Research Methods* 51, 195–203.
- Pfiffner, Alexandra M. 2021a. Cue-based features: Modeling change and variation in the voicing contrasts of Minnesotan English, Afrikaans, and Dutch. PhD thesis, Georgetown University.
- Pfiffner, Alexandra M. 2021b. Perception and production of [voice] contrasts in Dutch word-initial plosives. *Proceedings of the Linguistic Society of America* 6(1), 340–353.
- Pfiffner, Alexandra M. Forthcoming. Acoustic cues and obstruent devoicing in Minnesotan English. *American Speech*. <https://doi.org/10.1215/00031283-10867196>
- Pinget, Anne-France, René Kager, & Hans Van de Velde. 2019. Linking variation in perception and production in sound change: Evidence from Dutch obstruent devoicing. *Language and Speech* 63(3), 660–685.
- R Core Team. 2013. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Wagner, Suzanne Evans. 2012. Age grading in sociolinguistic theory. *Language and Linguistics Compass* 6, 371–382.
- Wissing, Daan P. 2018. Afrikaans. *Journal of the International Phonetic Association* 50, 127–140.
- Yip, Moira. 2002. *Tone*. New York: Cambridge University Press.