

## Perception and production of [voice] contrasts in Dutch word-initial plosives

Alexandra M. Pfiffner\*

**Abstract.** This study examines relative cue weights to the [voice] contrast in Dutch word-initial plosives. Perception and production data were collected from 25 native speakers, divided into two gender-balanced age groups (22–29; 61–71). Perception stimuli were artificially-manipulated continua /pad/-/bad/ (‘trail’, ‘bath’) and /tal/-/dal/ (‘quantity’, ‘valley’), varying in amount of prevoicing and f<sub>0</sub> of the following vowel. Results show that both prevoicing and f<sub>0</sub> are used as cues to phonological voicing. Production results show an increasing rate of devoicing in younger participants, and significant f<sub>0</sub> differences in all speakers. This suggests that cue weights are in the process of changing.

**Keywords.** voicing; Dutch; cue weights; perception; production

**1. Introduction.** Phonological contrasts are realized by multiple acoustic components that can be isolated from the speech signal, called ‘cues’ (Repp 1982). These cues can be individually manipulated, but they are perceived together and integrated in a phonologically-relevant way (Fitch et al. 1980; Repp 1982). Studies show that listeners give more or less attention to certain cues when identifying speech categories, which translates to cue weights (Francis et al. 2000; Holt & Lotto 2006): the perceptually-dominant cue is considered the primary cue, and the less perceptually-dominant cues are considered secondary cues.

Phonological voicing in plosives is signaled by a large number of cues that vary by language and the overall speech context. Dutch contrasts voiced and voiceless unaspirated plosives, so the primary cue is identified as the presence or absence of glottal pulsing before the plosive burst (negative VOT or prevoicing) (Lisker & Abramson 1964; Slis & Cohen 1969). Secondary cues signalling plosive voicing in Dutch include the burst duration, burst amplitude, duration of silent interval, preceding vowel duration, and vowel formant transitions, including transition time and the range in the F1 frequency shift from plosive burst to vowel (Slis & Cohen 1969). However, more recent work (van Alphen & Smits 2004; Pinget 2015; Pinget et al. 2019), discussed below, indicates that a change in progress may be occurring, with prevoicing occasionally absent and secondary cues gaining significant differences. The present study is an examination of the current status of prevoicing and f<sub>0</sub> cues in Dutch, both in perception and production.

1.1. REALIZATION OF VOICING IN DUTCH. Lisker and Abramson (1964) examined voicing in eleven different languages, including Dutch, focusing on voice onset time in word-initial plosives. A single native speaker of standard Dutch produced 216 different plosive-initial words, and the VOT in milliseconds was measured for each. All voiced plosives were produced with prevoicing; the average length in milliseconds was 80–85, with a range of 45–145 ms. All voiceless plosives were produced with positive VOT, or voicing lag, with an average duration of 10–25 ms and range of 0–35 ms.

---

\* Many thanks to Dirk Vet and Paul Boersma for use of the Speech Lab at the University of Amsterdam. Thank you also to research assistants Andrea Alićehajić, Michelle Bak, A’idah Defilippo, Emilio Luna, and Rorick Terlouw. This research is supported by NSF Doctoral Dissertation Research award BCS-1918306, Georgetown University Dissertation Research Travel Grant, and Georgetown University GradGov Research Project Award. Author: Alexandra M. Pfiffner, Georgetown University ([amp343@georgetown.edu](mailto:amp343@georgetown.edu)).

Slis and Cohen (1969) examined multiple acoustic cues to consonant voicing in Dutch and tested their perceptual relevance. In a perception experiment with artificially-manipulated stimuli, CV syllables were created with prevoicing ranging in six steps from 40 to 200 ms, and voicing lag which varied in three steps from 10 to 40 ms. Three subjects heard each stimulus and identified which sound they heard. Results showed that /b,d/ were associated with prevoicing, and /p,t/ with voicing lag.

The f<sub>0</sub> contour following the plosive burst was also studied. Twelve naturally-produced /Ca/ syllables with voiced consonants were compared to 15 syllables with voiceless consonants, and the results showed that there were differences in the initial f<sub>0</sub> at the onset of the vowel. The vowel onset following voiceless consonants measured 6 Hz higher than that following voiced consonants.

Van Alphen and Smits (2004) expanded on the previous work. In a production study, ten speakers, five women and five men, produced 64 tokens in isolation. The presence or absence of prevoicing was examined, and considerable variation was found. Overall, 75% of tokens were produced with prevoicing. Three speakers produced 100% of tokens with prevoicing, while all other participants fell between 100% and 38%. Participant gender was a significant factor in predicting the presence or absence of prevoicing; male speakers produced more tokens with prevoicing than female speakers. However, there were no significant differences in gender when it came to the duration of prevoicing. Place of articulation was also a significant factor, with labials having significantly more prevoicing than alveolars.

Of the 64 tokens, 48 that formed minimal pairs were selected and further analyzed for six potential voicing cues, including the f<sub>0</sub> at vowel onset and the overall f<sub>0</sub> trajectory into the vowel steady-state. They found that the mean f<sub>0</sub> following voiced plosives was 160.0 Hz, and the mean f<sub>0</sub> following voiceless plosives was 176.1 Hz. A difference of 16.1 Hz is 2.5 times greater than that reported by Slis and Cohen (1969). These were non-normalized f<sub>0</sub> measurements, so it is possible that there were differences by gender groups that were not made explicit.

In a perception study, sixteen participants heard the plosive and first half of the vowel from the 48 minimal pairs in the production study. They were asked to identify which phoneme they heard. The authors found a significant effect of prevoicing: voiced plosives were more often misidentified when they were produced without prevoicing. There was also a significant interaction between prevoicing and place of articulation, as more voiced labials without prevoicing were perceived as voiceless than alveolars produced without prevoicing. Van Alphen and Smits conclude that plosive voicing perception in Dutch is asymmetrical, as prevoicing presence signals phonological voicing, but its absence does not necessarily signal phonological voicelessness.

Word-initial neutralization was analyzed as a possible change in progress by Pinget and colleagues (2015; 2019). They analyzed both the devoicing of word-initial /v/ in Dutch, which is a change already in an advanced stage (see Pinget et al. 2019 and references within), and the devoicing of word-initial /b/. A production and perception experiment was conducted with 100 participants across five different regions in the Netherlands and the north part of Belgium. There were multiple production tasks, ranging from a word list to spontaneous speech, which were meant to elicit different phonetic realizations within the standard variety of Dutch. Prevoicing occurred in /b/ 75-80% of the time, though the authors note that there were some individuals “with clear devoiced realizations of (b)” (Pinget et al. 2019:670). The rate of /b/ de-

voicing was also consistent across the five regions. For the perception experiment, a continuum was made between /b/ and /p/, with varying levels of VOT, ranging from -90ms to +38ms. Participants heard each token and identified if they heard /b/ or /p/. The results show highly categorical perception, with the category boundary at approximately -10ms of VOT. The authors conclude that it is currently unclear if this plosive devoicing is a stable variation or if this is the beginning of a sound change.

1.2. CUE RE-WEIGHTING AND TONOGENESIS. Laryngeal contrasts have been frequently studied in terms of sound changes, as cues to phonological voicing can change over time and result in a shift in a language's phonological system. In particular, spectral cues from voicing contrasts can get phonologized and lead to tonogenesis, which is the introduction of phonological tone, or registrogenesis, which is the introduction of phonological register. These two processes are similar in that high tones and registers originate from voiceless stops, and low tones and registers originate from voiced stops. While tonogenesis occurs from the phonologization of  $f_0$  differences, registrogenesis can include  $f_0$ ,  $F_1$ ,  $F_2$ , VOT, vowel length, and phonation type (Brunelle & Kirby 2016; Brunelle et al. 2020).

Recent research has shown that Afrikaans is undergoing word-initial plosive voicing neutralization, which is resulting in tonogenesis (Coetzee et al. 2018). This may be relevant to Dutch, as Dutch is the mother language of Afrikaans, and the two remain very close linguistically (Heeringa et al. 2015). They have also undergone the same word-final voicing neutralization, so it is possible that the two are now undergoing the same word-initial voicing neutralization.

Coetzee and colleagues (2018) examined word-initial plosive voicing in Afrikaans in two populations: female speakers ages 21-23, and female speakers in their 40s through 60s. A production study showed younger speakers are devoicing in initial position at a rate of 83%, while older speakers are devoicing at a rate of 44% (Coetzee et al. 2018:192). There was also a clear  $f_0$  difference between vowels after underlyingly voiced and voiceless plosives in both populations, and this remains through the entire duration of the vowel.

In a perception task, twelve continua varying in VOT and the following vowel's  $f_0$  were created from natural tokens of a young female speaker of Afrikaans. One type of continuum had full prevoicing, one had no prevoicing, and one had reduced closure voicing. Results show that all participants relied on both  $f_0$  and voicing during closure to identify the word-initial plosive. However, when the two conflict, older participants relied more on the closure voicing as a cue. When the perception and production results are compared, a pattern emerges in that older speakers are more likely to produce the underlyingly voiced plosives with phonetic voicing, and they rely more on prevoicing in perception. On the other hand, younger speakers devoice more and rely on prevoicing less in perception. Thus, there appears to be a diachronic case of cue-transfer, as different generations are more heavily weighing different cues both in perception and production.

Following up on the study by Coetzee et al. (2018), Pfiffner (2020) examined word-initial plosive voicing in Afrikaans, specifically looking at age and gender differences. Speakers in this study were men and women ages 20-24 and 60-83. When devoicing rates of women and men were compared, it was found that women of all ages devoice more often than men. Additionally, there were robust differences in the following vowel's  $f_0$ , but there were larger differences in older participants.

Pfiffner (2020) further reports the results of a perception task in which four kinds of continua varied the amount of prevoicing and the  $f_0$  of the following vowel. Additionally, four different voices represented different age and gender groups. Overall, the results follow those of Coetzee et al. (2018): in the absence of prevoicing,  $f_0$  is a cue for all listener and speaker combinations, though less so for the perception of male speakers. When prevoicing is present, it is still the dominant cue.

Re-examining Dutch voicing cues in the context of cue re-weighting in Afrikaans may suggest that there is also an imminent sound change in Dutch. While prevoicing was still shown to be the primary cue to voicing in Dutch, van Alphen and Smits (2004), Pinget (2015), and Pinget et al. (2019) found a devoicing rate of approximately 25%, with notable individual variation. Additionally, Slis and Cohen (1969) found a mean difference of 6 Hz at vowel onset following voiced versus voiceless plosives, while van Alphen & Smits (2004) found a difference of 16.1 Hz. This suggests that the secondary cue of  $f_0$  may be strengthening while the primary cue is slightly weakening.

The present study is an examination of the current status of Dutch word-initial plosive voicing. The goals are two-fold: (1) to see if devoicing rates and  $f_0$  differences are continuing to change, and (2) to assess the current cue weighting of prevoicing and  $f_0$  in perception. Additionally, this study builds on the prior research by specifically analyzing effects of age and gender, both in production and perception.

1.3. RESEARCH QUESTIONS AND HYPOTHESES. The following research questions guided this experiment:

1. What is the current devoicing rate in word-initial position in Dutch?
2. Has the difference in the following vowel's  $f_0$  increased from prior research?
3. Are there differences in devoicing and  $f_0$  trajectories by age and gender?
4. Which cues signal a voice contrast in perception, and what is the relative cue weighting?
5. Are there differences in perception by age and gender?

If the cue weights that signal plosive voicing contrasts are changing, then it is hypothesized that there will be differences in perception and production by different demographic groups. Sociophonetic research has shown that women tend to lead sound change, as they use innovative forms of variables more often than men (Labov 2001). In this case, it means that women would devoice more than men and have larger differences in the following vowel's  $f_0$ . Additionally, the apparent time hypothesis predicts that as language changes, younger speakers will use more innovative forms in comparison to older speakers (Labov 1975; Wagner 2012). If there is an interaction of age and gender, then it is hypothesized that younger female speakers will devoice the most and have the largest  $f_0$  differences of all age and gender groups.

Studies have shown that speech perception is also influenced by social characteristics (Niedzielski 1999; Hay et al. 2006; Hay & Drager 2010). In a study of a merger-in-progress in New Zealand English, listeners were influenced by their own social characteristics as well as perceived characteristics of the speaker (Hay et al. 2006). If perception is affected by age and gender, then it is hypothesized that younger female listeners will use  $f_0$  as a cue to voicing more than other groups. Additionally, it is hypothesized that all listeners will rely more on  $f_0$  as a cue to voicing when listening to a younger female speaker versus other speakers.

## 2. Methods.

2.1. PARTICIPANTS. Participants were 25 native speakers of standard Dutch living in Amsterdam, the Netherlands. Fourteen speakers were ages 22-29 (f=7, m=7) and eleven speakers were ages 61-71 (f=5, m=6). The experiment took place in the Speech Lab at the University of Amsterdam, and participants were compensated for their time. The experiment was conducted entirely in Dutch, and any oral instructions were given by a native Dutch speaker to limit experimenter accommodation (Hay et al. 2009, 2010).

2.2. PROCEDURE. A production task and perception task were run with PsychoPy (Peirce et al. 2019). Participants were seated in front of a computer and microphone in a sound-attenuated booth. To avoid effects of accommodation, the production task was first (Giles et al. 1973; Goldinger 1998). Participants read aloud a randomized word list twice and were recorded at a sampling rate of 44.1 kHz. During the task, one word at a time appeared on the computer 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.

The perception task consisted of four blocks, with each block presenting a different speaker voice. 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 dit blok, zal u luisteren naar een zestig-jaar-oude vrouw uit Nederland.* ‘In this block, you will listen to a twenty-year-old woman from the Netherlands.’). This was a two-alternative forced-choice task: within each block, a token automatically played and two words were shown on the screen. Participants had to choose which word they heard. 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 30 total tokens, repeated five times, as well as fillers. All stimuli were randomized for each participant. Breaks were given every 40 tokens and in between each speaker block.

2.3. STIMULI. In the production task, stimuli were 20 monosyllabic (near-)minimal pairs beginning with /p,b/ and /t,d/, balanced by the following vowel, and 60 fillers. The minimal pairs are given in Table 1.

voiceless			voiced		
piek	/pik/	‘peak’	biecht	/bixt/	‘confess’
perk	/pɛrk/	‘limit’	berk	/bɛrk/	‘birch tree’
pad	/pad/	‘trail’	bad	/bad/	‘bath’
pol	/pɔl/	‘patch of grass’	bol	/bɔl/	‘ball’
poel	/pul/	‘puddle’	boel	/bul/	‘crowd’
tien	/tin/	‘ten’	dien	/din/	‘be of service’
telg	/tɛlx/	‘descendent’	delg	/dɛlx/	‘pay off debts’
tal	/tal/	‘quantity’	dal	/dal/	‘valley’
tol	/tɔl/	‘top’	dol	/dɔl/	‘crazy’
toe	/tu/	‘shut’	doel	/dul/	‘goal’

Table 1. Minimal pairs in the production task.

continuum	stable cue	changing cue
1	no prevoicing	f0 changing from low (step 1) to high (step 5)
2	full prevoicing	f0 changing from low (step 1) to high (step 5)
3	f0 at mid-range	prevoicing changing from 0% (step 1) to 100% (step 5)

Table 2. Artificially-manipulated continua used in the perception experiment. In each continuum, one cue was held constant and one was manipulated in five equal steps.

The perception task had two sets of artificially-manipulated continua that varied by the amount of prevoicing and the following vowel’s f0. The bilabial set was made with /pad/ and /bad/, with a natural utterance of /pad/ from a native speaker as the base. The alveolar set was made with /tal/ and /dal/, with /tal/ as the base. In each continuum, one cue was held constant and one was manipulated in five equal steps. Descriptions of the continua are given in Table 2.

The tokens were created based on recordings of four native speakers of Dutch, two female and two male. Each individual was recorded reading a word list multiple times. The duration of any prevoicing was measured, and means and ranges were calculated for each speaker. For each token, the f0 of the following vowel was measured at 10% increments throughout the vowel. The highest f0 at onset associated with a clear, modal token of /pad/ and /tal/ was selected for each speaker. Similarly, the lowest f0 at onset associated with a clear, modal token of /bad/ and /dal/ 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 continua 1 and 2. This became the base token. Using Praat’s Pitch Synchronous Overlap and Add function, 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 and 2. Three equidistant steps were calculated and created between the two ends, totalling five steps in the continua.

To create the prevoicing continuum, each speaker’s mean prevoicing duration in milliseconds 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% increments, creating a continuum from no prevoicing to full prevoicing (0-25%-50%-75%-100%).

2.4. ANALYSIS. The production task yielded 1,000 total tokens (40 per participant x 25 participants), which were annotated and forced-aligned using the Montreal Forced Aligner (McAuliffe et al. 2017). Segment boundaries were hand-corrected in Praat (Boersma & Weenink 2019). Underlyingly voiced plosives were coded as to whether or not there was any prevoicing. Following the methods of Slis and Cohen (1969) and van Alphen and Smits (2004), 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.

All following vowels were measured for f0 at 11 time points from onset to offset, totalling 11,000 measurements. A total of 199 measurement points were excluded because the Praat pitch tracking algorithm was unable to determine an f0. All f0 measurements were then analyzed by gender, with group means and standard deviations calculated. Outliers beyond 2.5

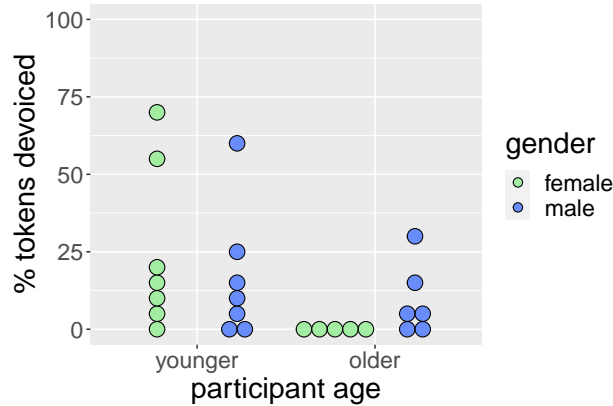


Figure 1. Percent of underlyingly voiced tokens produced without prevoicing. Each dot represents one participant.

standard deviations for each gender group were excluded, which resulted in Hertz ranges from approximately 124-286 Hz for female speakers, and 66-180 Hz for male speakers. All  $f_0$  measurements were then z-normalized based on an R script from Brunelle et al. (2020). Following their methodology, z-normalized  $f_0$  values were then converted back into Hertz-like measurements for readability. This was done using the mean and standard deviation for all speakers (group mean + z-score \* group SD, see Brunelle et al. 2020:8).

The perception task was analyzed as follows. Each individual participant response in the two alternative forced choice task was 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 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). Data from a fourth continuum, where  $f_0$  and prevoicing were varied together, was also collected, but only results from the first three are reported here.

### 3. Results.

3.1. PRODUCTION. Of the 500 underlyingly voiced tokens, 69 were produced without prevoicing (13.8%) and 431 were produced with some amount of prevoicing (86.2%). There was variation between individuals and between age and gender groups. Overall, younger female speakers had the highest devoicing rate (25%), then younger male speakers (16.43%), then older male speakers (9.17%). Older female speakers did not devoice at all. Individuals ranged from 0% to 70% (Figure 1). There was also a difference between place of articulation: 11.6% of all /b/ tokens were devoiced, while 16% of all /d/ tokens were devoiced, though this difference was not significant. A two-way ANOVA was run to examine the interaction between participant age group (younger/older) and gender (male/female) in predicting the presence or absence of prevoicing. The interaction was significant ( $F(1,496)=8.626, p<0.005$ ). An analysis of main effects showed that age category was a significant predictor of prevoicing ( $F(1,496)=27.207, p=2.69e-07$ ), though gender alone was not. A post-hoc Tukey HSD test showed significant differences between younger and older female speakers ( $p<0.001$ ), younger female and older male speakers ( $p<0.001$ ), and older female and younger male speakers ( $p<0.002$ ).

Normalized  $f_0$  trajectories of the vowel following the plosive are shown in Figure 2. The

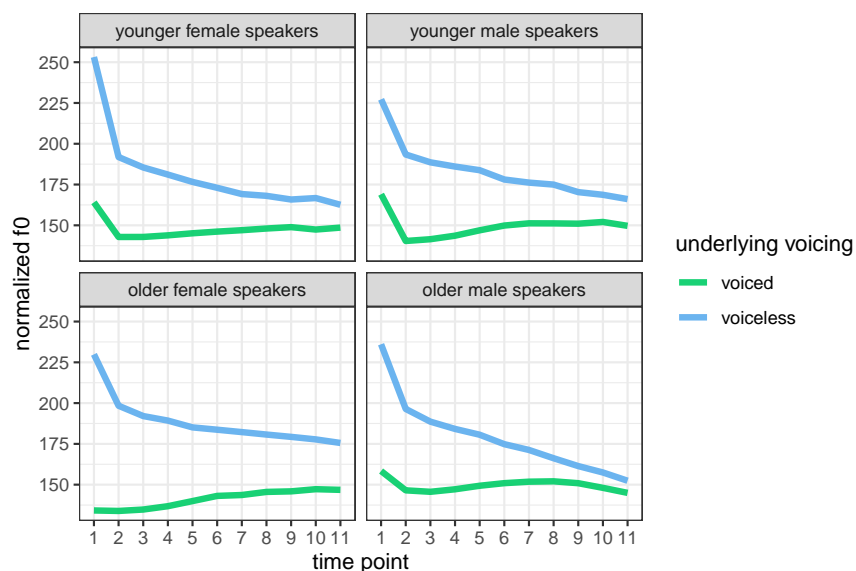


Figure 2. Normalized f0 trajectories following underlyingly voiced and voiceless plosives. Time point 1 is the onset of the vowel, and time point 11 is the offset. Each grid shows a different speaker group.

speaker group	mean Hz following /p,t/	mean Hz following /b,d/	difference
younger female	238	204	34 Hz
younger male	141	122	19 Hz
older female	237	181	56 Hz
older male	140	119	21 Hz

Table 3. Non-normalized f0 values following voiced and voiceless plosives, divided by speaker group.

f0 following a voiceless plosive is consistently higher than the f0 following a voiced plosive, from vowel onset (time point 1) to vowel offset (time point 11). In the previous studies of Dutch plosive voicing, the transition from the onset to the steady-state resulted in a negative slope for vowels following voiceless plosives (i.e. the f0 falls from onset to steady-state) and a positive slope for vowels following voiced plosives (f0 rises from onset to steady-state). This is not the case for all demographic groups in this study; all younger speakers and older male speakers have a negative slope from vowel onset to steady-state for all plosives. Only older female speakers show a positive slope for voiced plosives. Regardless of slope, the differences between f0 at onset are larger in all female speakers than male speakers. For the purpose of comparison with previous studies, non-normalized values in Hz are given in Table 3.

For all speakers, a linear mixed-effects model was fitted to the data. Normalized f0 at onset was the dependent variable, and fixed effects were underlying voicing and participant age \* gender interaction. Participant identity was a random effect. Only underlying voicing was a significant factor in predicting f0 at onset ( $p < 2e-16$ ).

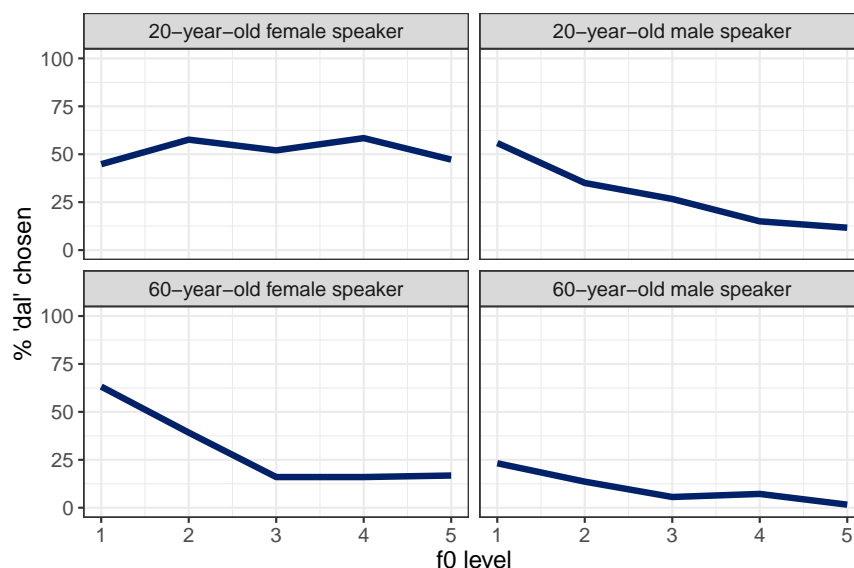


Figure 3. Perception of artificially-manipulated stimuli with no prevoicing and f0 changing from low (1) to high (5) in equal steps. Each grid shows all participants' responses to each speaker voice.

### 3.2. PERCEPTION.

3.2.1. CONTINUUM 1: NO PREVOICING. In the absence of any prevoicing, f0 can serve as a cue to underlying voicing, but the strength of the effect is dependent upon the speaker. Figure 3 shows participants' reactions to the /tal/-/dal/ continuum with no prevoicing, and f0 changing in five equal steps. Each grid shows a separate speaker voice, with all listener ages and genders collapsed. The percent that 'dal' was chosen is shown on the y-axis, and the f0 level from low (1) to high (5) is shown on the x-axis. The difference in speaker identity is clear, as there is no effect of f0 for the 20-year-old female speaker, as nearly all responses are at chance. On the other hand, the three other speakers have high f0s associated with voicelessness, and low f0s associated with voicing, even in the absence of any prevoicing. The 20-year-old male speaker and 60-year-old female speaker exhibit the largest f0 effect.

The strength of the f0 effect was assessed with a mixed-effects logistic regression model. Token choice was the independent variable, and the dependent variables were f0 level (1-5), place of articulation (bilabial/alveolar), speaker age group (younger/older), speaker gender, and the interaction between participant age group and participant gender. F0 level was a significant predictor ( $p < 2e-16$ ), with higher f0 levels associated with voicelessness, and lower f0 levels associated with voicing. There was also a significant difference by place of articulation ( $p < 2e-16$ ). Overall, the alveolar continuum showed some effect of f0, while the bilabial continuum did not. In terms of participant (listener) identity, gender was significant ( $p = 0.027$ ) as well as its interaction with age group ( $p = 0.017$ ), as male speakers were overall more likely to perceive a token as voiced. The speaker voice was also significant both in terms of age and gender ( $p < 2e-16$ ).

3.2.2. CONTINUUM 2: FULL PREVOICING. The second continuum had full (100%) prevoicing and f0 changing in five equal steps. Overall, the effect of prevoicing was large, as nearly all

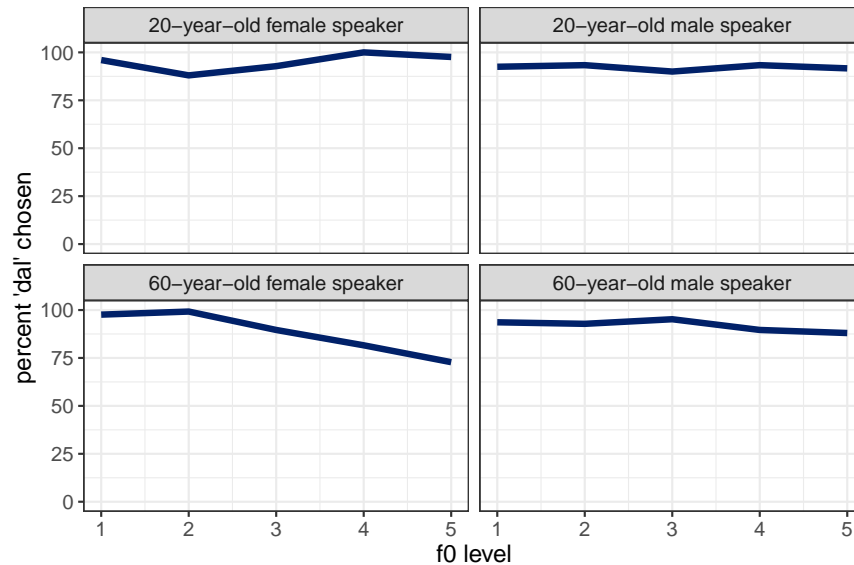


Figure 4. Perception of artificially-manipulated stimuli with full prevoicing and f0 changing from low (1) to high (5) in equal steps. Each grid shows all participants' responses to each speaker voice.

token responses were 'voiced.' Figure 4 shows token choice at near ceiling, with each speaker voice in a separate grid. One exception is with the 60-year-old female speaker. In her case, higher f0 levels elicited approximately 25% 'tal' responses. All participant groups reacted in the same way to this speaker.

A mixed effects model was fitted to the data through a step-up-step-down process. The model with the best fit had only f0 level and speaker as fixed effects, and participant number as a random effect. F0 level was significant ( $p=8.89e-05$ ) as well as speaker ( $p=3.39e-13$ ). Participant demographics were not significant in earlier models that were fitted in the analysis.

3.2.3. CONTINUUM 3: AMBIGUOUS F0, PREVOICING CHANGING. In the third continuum, the f0 was held at an ambiguous pitch (step 3 of continua 1 and 2) and the prevoicing amount changed in five equal steps. Figure 5 shows that when f0 is ambiguous, no prevoicing is associated with voiceless plosives. One exception is when listening to the 20-year-old female speaker, who elicited nearly equal amounts of voiced and voiceless responses when there was no prevoicing. There is some ambiguity with small amounts of prevoicing (25%), but 50% and higher lead to the perception of voiced plosives.

This model was fitted with prevoicing level as a fixed effect, as well as participant age, participant gender, place of articulation, speaker age, speaker gender, and their interaction. Participant number was a random effect. Prevoicing level was significant ( $p<2e-16$ ) as well as place of articulation ( $p=1.44e-11$ ). Participant gender was not significant, but age group was marginally significant ( $p=0.04$ ). Speaker age ( $p<0.001$ ), gender ( $p=2.77e-09$ ), and their interaction ( $p<0.001$ ) were all significant.

**4. Discussion.** The results of the perception and production experiments suggest that Dutch word-initial plosives are undergoing a cue re-weighting; both prevoicing and f0 can be used to signal phonological voicing, though prevoicing is still the dominant cue in both perception and

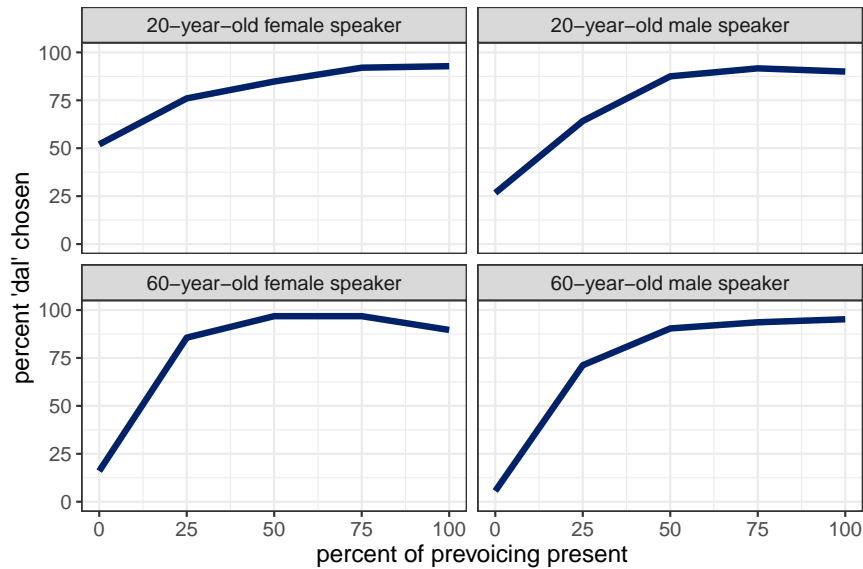


Figure 5. Perception of artificially-manipulated stimuli with an ambiguous  $f_0$  and prevoicing changing from 0% (no prevoicing) to 100% (full prevoicing) in five equal steps. Each grid shows all participants' responses to each speaker voice.

production. This change in cue weights is not equal across age and gender groups.

Beginning with the production cues, previous research has shown that speakers devoice approximately 25% of word-initial plosives, though there is a lot of individual variation. The results of this study suggest that the prevoicing cue has remained relatively unchanged since 2004. A total of 13.8% of plosives were devoiced in this study, but individual variation ranged from 0-25%. The speakers of this study also included people in their 60s through 80s, who were not expected to devoice as much as speakers in their 20s. This prediction was born out, as younger speakers devoiced significantly more often.

While prevoicing has remained steady, the cue of the following vowel's  $f_0$  appears to have strengthened. Prior studies showed small differences between the  $f_0$  value following voiced versus voiceless plosives, from 6 Hz (Slis & Cohen 1969) to 16 Hz (van Alphen & Smits 2004). When divided by age and gender, the production results in this study showed that male speakers have a 23-24 Hz difference, while female speakers have a 35-59 Hz difference. When the  $f_0$  values were z-normalized to compare across all participants, the differences by underlying voicing were significant, though speaker age and gender were not. Thus, all participants show significant differences in the following vowel's  $f_0$ , and the raw values show that these differences are larger than those found in previous studies. This suggests that  $f_0$  is becoming more prominent as a cue to word-initial plosive voicing.

The perception results also show that both prevoicing and  $f_0$  are used as cues to underlying voicing, though the strength of the effect varies by place of articulation and speaker voice. Overall,  $f_0$  is a stronger cue to voicing in alveolars than in bilabials. This fits with the production results in that alveolars were more often devoiced than bilabials, which leads to a compensatory relationship. Since alveolar plosives are more often lacking prevoicing,  $f_0$  strengthens to signal voicing.

In the absence of prevoicing, high f0s cue voicelessness, and low f0s cue voicing. If this is an ongoing sound change, we might expect to see a stronger effect of f0 when listening to female versus male speakers, or when listening to younger versus older speakers. This was not the case; listeners did not respond to changing f0 in the voice of the 20-year-old female speaker. They did respond to changing f0 when listening to the other three speakers, though to different extents. F0 cued voicing more when listening to the 20-year-old male speaker and the 60-year-old female speaker, and less when listening to the 60-year-old male speaker.

When prevoicing was present, most listeners perceived all the tokens as voiced. The one exception was when listening to the 60-year-old female speaker, for whom some tokens with high f0 were perceived as voiceless. This speaker effect is unexpected, as it was hypothesized that f0 would have the largest effect for the younger female speaker, not the older.

Finally, when f0 is ambiguous and prevoicing is presented in increments, a certain amount of prevoicing is needed to cue phonological voicing. From 0-25%, there is ambiguity, especially in the 20-year-old male speaker. These results suggest that prevoicing is slightly weakening in cue weighting; the absence of prevoicing does not necessarily signal voicelessness, and the presence of prevoicing does not necessarily signal voicing.

While there were not significant differences overall between the listener demographic groups, there were clear differences in how the four speakers' voices were perceived. It is unclear if the differences are due to the age and gender groups that the individual speakers represent, or if there is something in particular about a speaker's voice that did or did not signal voicing. The stimuli in this study were created from four native speakers' productions. While this resulted in natural sounding stimuli, a drawback is that the f0 ranges and average prevoicing values were unique to each speaker. For example, the two female speakers had similar f0 ranges at onset in their /tal/-/dal/ continua. The younger female speaker had a difference between f0 steps 1 and 5 of approximately 135 Hz, while the older female speaker had a difference of 145 Hz. However, the older female speaker's highest f0 was 274 Hz, while the younger speaker's was 246 Hz. It is possible that differences like these affected perception. This research would benefit from a follow-up study with methodology similar to Hay et al. (2006), where listeners hear identical acoustic stimuli, but different participant groups are presented with different descriptions of each speaker in terms of social characteristics.

**5. Conclusion.** This study found that both prevoicing and f0 are used as cues to phonological voicing in Dutch word-initial plosives. In comparison to previous studies, the amount of devoicing has not changed. However, in this study, the individual speakers with the largest devoicing rates were all in the younger age group. Following the apparent time hypothesis, this could be a signal that a change is underway. Significant differences between voiced and voiceless plosives were found in f0 at onset. The difference in Hertz values was also the largest that has been reported so far, providing further evidence that f0 may be becoming a more important cue to voicing. Prevoicing was significantly different by age and gender. Additionally, all participants had a significant difference in f0 following voiced versus voiceless plosives, but the size of the effect did not differ across speaker groups.

The perception test shows that prevoicing is still the dominant cue to initial voicing in Dutch, though f0 can also be used by all listeners to distinguish between underlyingly voiced and voiceless plosives. The largest effect occurs in the absence of prevoicing, but mid-range f0 values and short prevoicing lengths can lead to ambiguity. While van Alphen and Smits (2004)

concluded that the absence of prevoicing does not necessarily mean voicelessness, the present study adds that the presence of some degree of prevoicing does not always signal phonological voicing.

Combining these results with previous studies, it appears as though the prevoicing cue has stayed stable in production, and it has slightly weakened in perception. At the same time, the f<sub>0</sub> cue has strengthened in both perception and production. While the overall cue weighting is changing in Dutch, it is not equal through all demographic groups. Whether or not this cue re-weighting will lead to tonogenesis as in Afrikaans remains to be seen, though given the languages' relation and overall similarities, it is possible that this will happen in the future.

## References

- van Alphen, Petra M. & Roel Smits. 2004. Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics* 32(4). 455–491. <https://doi.org/10.1016/j.wocn.2004.05.001>.
- Bates, Douglas, Martin Mächler, Benjamin M. Bolker & Steven C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Boersma, Paul & David Weenink. 2019. Praat: doing phonetics by computer [Computer program] (Version 6.1). Retrieved 13 July 2019 from <http://www.praat.org>.
- Brunelle, Marc & James Kirby. 2016. Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass* 10(4). 191–207. <https://doi.org/10.1111/lnc3.12182>.
- Brunelle, Marc, Ta Thanh Tan, James Kirby & Dinh 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. <https://doi.org/10.5334/labphon.278>.
- 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. <https://doi.org/10.1016/j.wocn.2017.09.009>.
- Fitch, Hollis L., Terry Halwes, Donna M. Erickson & Alvin M. Liberman. 1980. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics* 27. 343–350. <https://doi.org/10.3758/BF03206123>.
- Francis, Alexander L., Kate Baldwin & Howard C. Nusbaum. 2000. Effects of training on attention to acoustic cues. *Perception & Psychophysics* 62. 1668–1680. <https://doi.org/10.3758/BF03212164>.
- 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. <https://doi.org/10.1017/S0047404500000701>.
- Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2). 251–279. <https://doi.org/10.1037/0033-295x.105.2.251>.
- Hay, Jennifer & Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48(4). 865–892. <https://doi.org/10.1515/LING.2010.027>.
- 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. <https://doi.org/10.1080/07268600902823128>.
- Hay, Jennifer, Katie Drager & Paul Warren. 2010. Short-term exposure to one dialect affects processing of another. *Language and Speech* 53(4). 447–471.

- <https://doi.org/10.1177/0023830910372489>.
- 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.  
<https://doi.org/10.1016/j.wocn.2005.10.001>.
- Heeringa, Wilbert, Febe de Wet & Gerhard B. van Huyssteen. 2015. Afrikaans and Dutch as closely-related languages: A comparison to West Germanic languages and Dutch dialects. *Stellenbosch Papers in Linguistics Plus* 47. 1–18. <https://doi.org/10.5842/47-0-649>.
- Holt, Lori L. & Andrew J. Lotto. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America* 119(5). 3059–3071. <https://doi.org/10.1121/1.2188377>.
- 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. 2001. *Principle of linguistic change, volume 2: Social factors*. Malden, MA: Blackwell Publishers.
- Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *WORD* 20(3). 384–422.  
<https://doi.org/10.1080/00437956.1964.11659830>.
- McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner & Morgan Sonderegger. 2017. Montreal Forced Aligner [Computer program] (Version 0.9.0). Retrieved 1 May 2020 from <http://montrealcorpus-tools.github.io/montreal-forced-aligner/>.
- Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18. 62–85.  
<https://doi.org/10.1177/0261927X99018001005>.
- 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. 2020. Tonogenesis in Afrikaans: Age and gender differences in cue weighting. Poster presented at the 17th Conference on Laboratory Phonology.
- Pinget, Anne-France. 2015. *The actuation of sound change*. Utrecht University dissertation.
- 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. <https://doi.org/10.1177/0023830919880206>.
- R Core Team. 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. <http://www.R-project.org/>.
- Repp, Bruno H. 1982. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin* 92(1). 81–110.  
<https://doi.org/10.1037/0033-2909.92.1.81>.
- Slis, I. H. & Amos M. Cohen. 1969. On the complex regulating the voiced-voiceless distinction I. *Language and Speech* 12. 80–102. <https://doi.org/10.1177/002383096901200202>.
- Wagner, Suzanne Evans. 2012. Age grading in sociolinguistic theory. *Language and Linguistics Compass* 6. 371–382. <https://doi.org/10.1002/lnc3.343>.