# Realisation of Intervocalic /t/ in Australian English: A Snapshot

*Thomas Powell-Davies, Rosey Billington*

Australian National University

thomas.powell-davies@anu.edu.au, rosey.billington@anu.edu.au

## Abstract

In word-medial intervocalic contexts, a common realisation of contemporary Australian English /t/ is a tap, [ɾ], but research on past and present realisations of medial /t/ is limited. This study presents a preliminary exploration of the realisation of medial /t/ by 72 speakers in the Mitchell and Delbridge corpus, collected from 1959–1960. Results show a range of variants in use, including taps, fricated and affricate realisations. There are some differences in the distribution of variants according to sex, and in overall release phase durations. The study highlights the need for further phonetic research on variation and change in Australian English consonants.

**Index Terms**: phonetics, stops, tapping, flapping, VOT, aspiration, frication

## 1. Background

### 1.1. Allophones of intervocalic /t/ in English varieties

In varieties of English, the voiceless alveolar plosive /t/ has a range of allophones. Aspirated [tʰ] is a canonical variant in word-initial syllable onsets, but word-medially, particularly before unstressed vowels, various other realisations are described. In many varieties of British English, intervocalic glottalised stops and glottal stops [ʔ] are widespread in this context (e.g.[1]), with lenited fricative and affricate realisations in some regions (e.g. [2]). In North American varieties of English, intervocalic alveolar taps[1] [ɾ] are ubiquitous allophones of /t/ (and /d/) word-medially, and can also occur word-finally [4]. Compared to alveolar plosives, alveolar taps are characterised as acoustically short, with closures of around 10–40ms, and typically voiced (e.g. [5] [6]). Intervocalic tapping is also reported in other varieties of English, for example in South Africa, New Zealand, and Australia (e.g. [7]). The majority of studies on intervocalic /t/ in English varieties are based on classifying /t/-realisations as one of several main allophones such as tapped, aspirated, unaspirated, voiced, glottalised, spirantised, and fricated. However, as various authors have noted, categorising phones can be challenging, even with reference to acoustic information; productions of medial /t/ correspond to a wide range of highly gradient realisations, which could be broadly construed as various types of lenition phenomena [8].

### 1.2. Word-medial intervocalic /t/ in Australian English

Descriptions of contemporary Australian English (AusE) describe tapping of intervocalic /t/ before unstressed vowels as a common process (e.g. [9]). However, it is not yet clear to what extent tapping occurs within different communities or for different individuals, and [10] points out that intervocalic tapping is not reported in the literature on AusE until the 1980s. In sociolinguistic work based on spontaneous speech data collected with adults and teenagers in Sydney from 1977–81, [11] finds intervocalic [ɾ] in use, more often by male speakers of 'Broad' varieties. An affricated [tˢ] is also observed, primarily as a word-final pre-pausal allophone, used more often by speakers who are female, of Greek heritage, and who are middle class, and suggested to be a recent development, at the time. Intervocalic tapping is similarly observed in conversational data collected with Brisbane teenagers in the 1980s, at comparable rates for students at a middle class compared to a working class school, with comments that it is not obligatory but that it is also uncommon not to tap in a tapping environment [12]. In a study of /t/ allophones based on conversational and read speech collected in Victoria in the 1990s, [10] reports usage of allophones [t], [tˢ], [ʔ], and [ɾ] in various non-initial contexts. For word-medial intervocalic /t/ onsets in unstressed syllables, the majority of realisations are [ɾ], with some use of plosive [t] (the degree of aspiration is not described), and some tokens of [tˢ] for the middle but not lower socioeconomic group in the study (where [tˢ] here incorporates both fricatives and plosives with fricated release). Other work on /t/ production in Victoria, including intervocalic tokens, also notes occurrences of tapped and fricated realisations and strong sociolinguistic patterning (e.g. [13] [14]), and a recent study of voiceless plosives produced by children aged 5–13 in regional Victoria finds some gendered patterns, for example greater use of fricated variants by young girls, but also comparable rates of intervocalic tapping for boys and girls as the children get older [15]. Spontaneous speech data collected in the 2000s in Perth also finds a range of word-medial intervocalic realisations of /t/, including tapped and fricated variants, with fricated realisations being used more often by speakers from suburbs classified as higher socioeconomic status [16].

In [10] it is noted that tapped and fricated (and glottalised) /t/ allophones are not mentioned in the foundational large-scale survey conducted by Mitchell and Delbridge [M&D] based on data collected from 1959–1960 with nearly 8,000 high school students across Australia [17]. However, [13] note that fricated variants can be observed in the M&D audio-recorded materials [18], and some observations in the M&D written work also suggest other realisations. For example, in some transcribed examples of words such as *water* and *quarter*, intervocalic /t/ is represented as [d̥] and described as 'weakly articulated'.[2] With reference to phonetic transcriptions based on auditory impressionistic analyses, such realisations are used by M&D as minimal indicators of whether speakers show evidence of 'weak' consonant articulations in general, a categorisation which is not

[1] In this paper, we use the term 'tap', but note that 'flap' is also used by some authors for this kind of segment, while others reserve 'flap' for segments such as [ɽ] which require retraction followed by a forward-striking movement of the active articulator (see e.g. [3]).

[2] Specifically, M&D write that "the diacritic [ ◌̥ ] is added ... to indicate a voiceless sound somewhat more weakly articulated than the usual voiceless variety (*lenis* instead of *fortis*)."

given firm parameters but appears to broadly relate to impressions of articulatory strength. This is reportedly the case for 7% of the sample and more common for males and those classified as speakers of the 'Broad' variety of AusE. In 1960s print media, there are also orthographic representations indicative of a non-aspirated or voiced medial /t/ (e.g. 'bewdy' for beauty) [19], prompting questions about the nature of /t/ allophones and the presence of potential tap-like variants in this time period.

### 1.3. Acoustics of /t/ allophones in Australian English

As [16] note, discrete segmental categorisations of /t/ allophones are abstractions away from the complexities of speech production, and there may be sociolinguistically salient detail to uncover in more fine-grained phonetic analyses. As articulatorily complex sounds, plosives are potentially rich sources of variation, and phonetic analyses for various languages show differences on the basis of factors such as region, gender, and age, though the nature of these differences varies [20]. VOT, which is by far the most frequently analysed measure, shows high variability within and across speakers, but there is also evidence for speaker-specific VOT profiles that hold across e.g. places of articulation and speech rates ([21] [22]). For AusE, relative to the wealth of acoustic phonetic research on vowels, there has been more limited exploration of consonant acoustics. Existing research includes various phonetic examinations of coda /t/, and other coda plosives, in the context of work demonstrating the increasing prevalence of glottalisation in word-final voiceless plosives (e.g. [23]), and analyses of VOT for word-initial /t/ showing influences of factors such as ethnic identity and gender (e.g. [24]). Measures of VOT and closure duration for word-initial and word-medial plosives provide supporting evidence for the phonological contrast between two stop series in mainstream and Indigenous AusE varieties [25], and a detailed acoustic study of fricated /t/ in word-medial and word-final intervocalic contexts shows it is spectrally similar to /ʃ/. While these are not the only studies, there is clear scope for more comprehensive phonetic research on AusE plosives, particularly /t/ in word-medial intervocalic contexts, given the range of realisations possible in this environment.

## 2. Aim of this study

This study aims to provide an exploratory examination of realisations of word-medial intervocalic /t/ in AusE at the beginning of the 1960s, drawing on the M&D data. The availability of the M&D corpus offers the opportunity to revisit the impressionistic observations in [17] and develop a more detailed snapshot. While the data is limited in scope, the M&D study design has been used as a reference point in the creation of more recent corpora, and a better understanding of medial /t/ characteristics in this early corpus of AusE speech will facilitate future investigation of medial /t/ realisation over time.

## 3. Method

### 3.1. Materials

The materials for this study are taken from the corpus collected by Mitchell and Delbridge in 1959–60 [18]. For this preliminary exploration, a subset of the corpus was chosen based on available metadata, which includes school location, birthplace, and speaker sex. The age range of participants in the M&D study was 16–18. The subset included 72 speakers (44 female, 28 male) from three high schools in southern Tasmania, all born in Australia. The M&D corpus includes two sentences that participants were asked to read, designed to inform vowel analyses, but each contains a /t/ context of interest for the present study. The first – *"Let's pick a good spot near the water and spend the morning surfing and relaxing in the sun"* – provides a context with word-medial intervocalic tokens at the onset of an unstressed syllable, taken from 'water'. Three tokens were excluded from this context, as extensive devoicing of following syllables meant their environment was no longer intervocalic. The second sentence – *"The plane flew low over the runway, then increased speed and circled the aerodrome a second time"* – provides a context with word-initial tokens, taken from 'time', as a point of comparison. Altogether, there are 141 tokens – 72 word-initial and 69 word-medial.

### 3.2. Data processing and annotation

Textgrids were created for each .wav file from the corpus, and /t/ allophones were coded in Praat [26] with reference to waveforms and corresponding spectrograms. The data presented a range of /t/ realisations, which were categorised into variant types with with reference to their acoustic characteristics. These variants and their characteristics had similarities to what has been reported in other works (e.g. [13] [27]). The categorisation was undertaken based on the following criteria:

- **Aspirated [tʰ]** – a voiceless token where the period of closure is followed by a period of h-like frication.
- **Affricate [tˢ]** – a voiceless token where the period of closure is followed by a period of s-like frication.
- **Fricated [t̪]** – a voiceless token where there is no period of closure observed in the waveform, and the frication is observed for the entire duration of the phone.
- **Tapped [ɾ]** – a token with no break in voicing observed on the spectrogram, with the closure duration relatively short.
- **Partially voiced [d̥]** – a token that is voiced for most of the closure. Auditorily, this variant sounds similar to what is often found in English /d/. It is usually followed by a very short release phase. The voicing is what was used to distinguish these from the aspirated tokens. This category has similarities to what has been described as voiced in other work (e.g. [13]).

It is worth noting that the aspirated and affricate variants proved particularly difficult to definitively distinguish. As reported by others who have worked with similar data (e.g. [27]), in some cases there are tokens which show characteristics of both h-like and s-like frication. Additionally, a release burst, as would be typical for [tʰ], was not always clearly identifiable. In these cases, tokens were categorised based on whether they showed more h-like or s-like frication. It is worth emphasising that aspirated [tʰ] and affricate [tˢ] cannot be categorically distinct, as the degree of frication at the alveolar and glottal places is highly gradient. As abstractions from the realities of speech production, these categories are only an initial reference point in delimiting the many gradient possibilities of phonetic detail.

For /t/ realisations where there was an observable occlusion and clear release phase (however short), namely [tʰ], [tˢ], [d̥], the release phase was also annotated, with reference to the onset of a release burst or frication, and the onset of periodicity in the following vowel.

### 3.3. Analytical procedures

The .Textgrid files and corresponding .wav files were used to create a database in the EMU Speech Database Management System [28], and /t/ variant types and release phase measures for initial and medial tokens were queried using the emuR package in R [29].

## 4. Results

### 4.1. Word-initial and intervocalic realisations of /t/

Table 1: *Token numbers for /t/ variant types word-initially (in 'time') and word-medially (in 'water') by sex.*

| Phone | Word-initial | | Word-medial | | **Total** |
|---|---|---|---|---|---|
| | Female | Male | Female | Male | |
| [t$^{h}$] | 40 | 20 | 15 | 10 | **85** |
| [t$^{s}$] | 4 | 2 | 15 | 1 | **22** |
| [t̪] | 0 | 0 | 11 | 7 | **18** |
| [ɾ] | 0 | 0 | 0 | 5 | **5** |
| [d̥] | 0 | 6 | 0 | 5 | **11** |
| All | 44 | 28 | 41 | 28 | **141** |

The five categories of /t/ realisation showed different distributions, depending on word position and sex. The token counts are presented in Table 1. In word-initial position, the aspirated realisations were the most common by far (60/72, 80%), with the affricate realisations appearing in only a handful of tokens across both male (2/28, 7%) and female (4/44, 9%) speakers. However, there were six tokens of partially voiced realisations for the male speakers (21%), and none for the female speakers.

In word-medial intervocalic position, there was a similar proportion of aspirated realisations for the female (15/41, 37%) and male (11/28, 39%) speakers. The rest of the tokens for the female speakers comprised affricate (15/41, 37%) and fully fricated (11/41, 27%) realisations. The male speakers had far fewer of these realisations, with seven fricated tokens (25%) and just one affricate token (4%). Instead, they had five each (18%) of both partially voiced and tapped tokens, which were absent for the female speakers.

### 4.2. Release phase duration by variant type

The duration of the release phase, here referring to the Voice Onset Time of plosive phones as well as the fricated release of affricate phones, is shown for partially voiced [d̥], aspirated [t$^{h}$] and affricate [t$^{s}$] tokens in Figure 1. The partially voiced tokens have very short release phases word-initially (mean 25ms, $\sigma$ 6ms, n = 6) and word-medially (mean 26ms, $\sigma$ 9ms, n = 5), typical of short-lag release (e.g. [30]). The release phases for the aspirated tokens are much longer both word-initially (mean 53ms, $\sigma$ 17ms, n = 60) and word-medially (mean 45ms, $\sigma$ 14ms, n = 25). The release phases of the affricate tokens are longer still for both word-initial tokens (mean 57ms, $\sigma$ 32ms, n = 6) and word-medial tokens (mean 64ms, $\sigma$ 14ms, n = 16). However, for the affricate tokens, the fricated release phase takes up a greater proportion of the phone relative to the closure phase (initial: mean 64%, $\sigma$ 16%, n = 6; medial: mean 73%, $\sigma$ 9%, n = 16) than the Voice Onset Time does in the aspirated tokens (initial: mean 59%, $\sigma$ 13%, n = 60; medial: mean 58%, $\sigma$ 13%, n = 25), though recalling that as noted in Section 3.2, these variant types are difficult to strictly delineate.

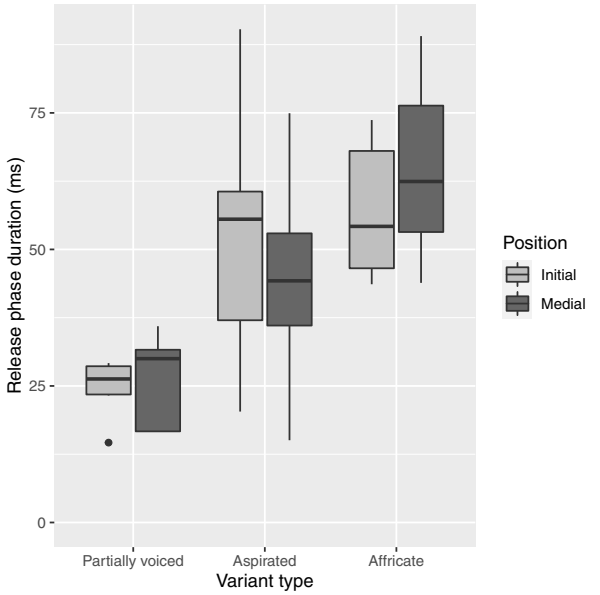


Figure 1: *Release phase duration (ms) for /t/ variants, by variant type ([d̥], [t$^{h}$], [t$^{s}$]) and word position.*

### 4.3. Initial vs. medial release phase duration by speaker

The tendency towards longer release phases for initial compared to medial aspirated [t$^{h}$] aligns with typical patterns for English voiceless plosives of longer VOT lag for stressed compared to unstressed onsets, and for (word-/utterance-) initial onsets compared to medial onsets (e.g. [31] [32]). Given that there are also (mixed) reports of gender differences in English VOT, and evidence for speaker-specific VOT profiles (Section 1.3), in Figure 2 we examine release phase duration by speaker for initial vs. medial [t$^{h}$], [t$^{s}$], and [d̥], for all speakers who produced one of these variants in both contexts. There is no connection observed between the release phase in the initial and medial positions ($R^2$ = 0.05). This is still the case when partially voiced tokens are removed ($R^2$ = 0.018). However, there are clear differences in the duration of the release phase for female compared to male speakers, with the female speakers tending to have longer release phases.

## 5. Discussion

This study builds on the auditory analyses of [17] and finds that a range of realisations of /t/ are evident in the M&D sentence data from 1959–1960. This is as expected, given the range of variants observed in more recent studies (e.g. [11] [10]), but not apparent from the higher-level categorisations (e.g. of 'weak' consonants) in the original work, and therefore adds to our understanding of /t/ realisation at an earlier stage of AusE. In particular, the data shows clear evidence that taps, as well as fricated and affricate realisations, were in word-medial use at least two decades before the earliest reports in existing research.

In the word-medial intervocalic position of particular interest here, aspirated realisations are most common, followed by affricate and fricative realisations. Tapped and partially voiced realisations are present in smaller numbers. This is a much more diverse set of variants than found in the word-initial context, in-

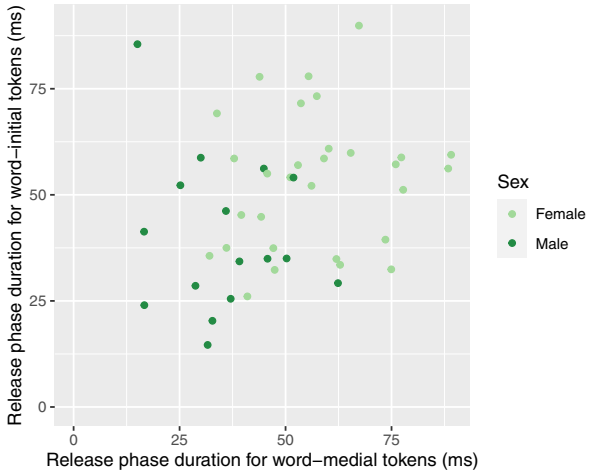


Figure 2: *Release phase duration of [$t^h$], [$t^s$], and [$\underset{\circ}{d}$] in initial versus medial positions, by speaker and sex.*

dicating higher variability in the word-medial intervocalic context here, when /t/ is the onset of an unstressed syllable. Tapped realisations are much less common than more recent reports suggest (e.g. [12]), noting however that the speech style of the two elicited sentences is likely to differ from the more conversational style of speech analysed in some later studies.

The use of different variants in this data appears to differ according to speaker sex. Both female and male speakers produced aspirated, affricate and fricated tokens, but none of the female speakers produced any of the partially voiced or tapped variants, whereas these were well represented in the data for the male speakers. This has similarities to findings reported by [11] on the greater use of taps by 'Broad' male speakers.

The differences in variant distribution may be related to the differences in release phase duration. While there appears to be no correlation between a speaker's release phase duration in word-initial position and word-medial intervocalic position (as might have been expected given other kinds of evidence for speaker-specific VOT profiles [22]), it is apparent that the male speakers have shorter release phases in general. This has also been observed, albeit inconsistently, in other varieties of English [20]. The shorter release phases and the presence of partially voiced tokens for male speakers could potentially show the earlier phases of a process that has progressed into the widespread distribution of taps observed in present-day Australian English [9]. However, much more detailed and comprehensive analyses, together with direct comparison with data from later timepoints, will be needed to explore this.

As part of more detailed and comprehensive analyses, further examination of the phonetic characteristics of word-medial intervocalic /t/ will be of value, and complement the increasing number of phonetic studies on word-initial and word-final /t/. The precise processes at play in this environment, at the onset of unstressed syllables, can be difficult to characterise, given that the phone realisations are highly gradient (and arguably correspond to a range of different types of lenition, e.g. [10] [8]). However, additional fine-grained phonetic detail is likely to significantly add to our understanding of socially and linguistically structured variation for AusE /t/ and other plosives.

## 6. Conclusion

The data presented sheds light on the realisation of /t/ in Australian English, with a focus on word-medial intervocalic segments at the onset of unstressed syllables. It shows that multiple variants of intervocalic /t/ were already in use at the time Mitchell & Delbridge collected data for their foundational study of Australian English speech, well before this range of variants was described in the scholarly literature. This raises questions about the nature of /t/ realisation in Australian English over time, and to what extent there is evidence for stable variation compared to changing patterns in /t/ realisation. More broadly, this study also highlights the need to continue adding to sociophonetic research on consonants in Australian English, to complement the extensive research on variation and change in Australian English vowels.

## 7. Acknowledgements

We wish to thank two anonymous reviewers for their generous comments.

## 8. References

[1] Docherty, G. J. and Foulkes, P. "Glottal variants of /t/ in the Tyne-side variety of English", in W. J. Hardcastle, and J. Beck, [Eds], *A figure of speech: A festschrift for John Laver*, 213–240. Routledge, 2014.

[2] Marotta, G., Barth, M., et al. "Acoustic and sociolinguistic aspects of lenition in Liverpool English", Studi Linguistici e Filologici On Line, 3(2):377–413, 2005.

[3] Ladefoged, P. and Maddieson, I. The sounds of the world's languages. Blackwell, 1996.

[4] De Jong, K. J. "Flapping in American English", in Oostendorp, M., Ewen, C. J., Hume, E., and Rice, K., editors, *The Blackwell Companion to Phonology*, 1–19. Wiley, 1 edition, 2011.

[5] Zue, V. W. and Laferriere, M. "Acoustic study of medial /t, d/ in American English", Journal of the Acoustical Society of America, 66(4):1039–1050, 1979.

[6] Turk, A. "The American English flapping rule and the effect of stress on stop consonant durations", Cornell Working Papers in Phonetics, 7:103–133, 1992.

[7] Wells, J. C. Accents of English, volume 1–3. Cambridge University Press, 1982.

[8] Ashby, M. and Przedlacka, J. "The stops that aren't", Journal of the English Phonetic Society of Japan, 1:14–15, 2011.

[9] Cox, F. and Fletcher, J. Australian English pronunciation and transcription. Cambridge University Press, 2 edition, 2017.

[10] Tollfree, L. "Variation and change in Australian English consonants: Reduction of /t/", English in Australia, 26:45–67, 2001.

[11] Horvath, B. M. Variation in Australian English: The sociolects of Sydney. Cambridge University Press, 1985.

[12] Ingram, J. "Connected speech processes in Australian English", Australian Journal of Linguistics, 9(1):21–49, 1989.

[13] Loakes, D., McDougall, K., and Gregory, A. "Variation in /t/ in Aboriginal and Mainstream Australian Englishes", in *Pro-ceedings of the Eighteenth Australasian International Conference on Speech Science and Technology*, 61–65. Australasian Speech Science and Technology Association, 2022.

[14] Loakes, D. and McDougall, K. "Individual variation in the frication of voiceless plosives in Australian English: A study of twins' speech", Australian Journal of Linguistics, 30(2):155–181, 2010.

[15] Ford, C. Acquisition of gender-specific sociophonetic cues in the speech of primary school-aged children. PhD Thesis, La Trobe University, 2018.

[16] Docherty, G., Foulkes, P., Gonzalez, S., and Mitchell, N. "Missed connections at the junction of sociolinguistics and speech processing", Topics in Cognitive Science, 10(4):759–774, 2018.

[17] Mitchell, A. G. and Delbridge, A. The speech of Australian adolescents: A survey. Angus & Robertson, 1965.

[18] Mitchell, A. G. and Delbridge, A. The speech of Australian adolescents: Research data and recordings collected by A.G. Mitchell and Arthur Delbridge in 1959 and 1960 [available via https://hdl.handle.net/2123/31585], 1998. University of Sydney.

[19] Army: The Soldier's newspaper. Here's a 'Bewdy', Nov 1964.

[20] Chodroff, E. and Foulkes, P. "Sociophonetics and stops". In Strelluf, C., editor, *The Routledge handbook of sociophonetics*, pages 143–175. Routledge, 2024.

[21] Theodore, R. M., Miller, J. L., and DeSteno, D. "Individual talker differences in voice-onset-time: Contextual influences", The Journal of the Acoustical Society of America, 125(6):3974–3982, 2009.

[22] Chodroff, E. and Wilson, C. "Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English", Journal of Phonetics, 61:30–47, 2017.

[23] Penney, J., Cox, F., Miles, K., and Palethorpe, S. "Glottalisation as a cue to coda consonant voicing in Australian English", Journal of Phonetics, 66:161–184, 2018.

[24] Clothier, J. and Loakes, D. "Coronal stop VOT in Australian English: Lebanese Australians and mainstream Australian English speakers". In *Proceedings of the Seventeenth Australasian International Conference on Speech Science and Technology*, pages 13–16. Australasian Speech Science and Technology Association, 2018.

[25] Mailhammer, R., Sherwood, S., and Stoakes, H. "The inconspicuous substratum: Indigenous Australian languages and the phonetics of stop contrasts in English on Croker Island", English World-Wide, 41(2):162–192, 2020.

[26] Boersma, P. and Weenink, D. Praat: doing phonetics by computer, 2022.

[27] Buizza, E. and Plug, L. "Lenition, fortition and the status of plosive affrication: the case of spontaneous RP English", Phonology, 29(1):1–38, 2012.

[28] Winkelmann, R., Harrington, J., and Jänsch, K. "EMU-SDMS: Advanced speech database management and analysis in R", Computer Speech & Language, 45:392–410, 2017.

[29] R Core Team. R: A Language and Environment for Statistical Computing, 2021.

[30] Cho, T. and Ladefoged, P. "Variation and universals in VOT: Evidence from 18 languages", Journal of Phonetics, 27(2):207–229, 1999.

[31] Lisker, L. and Abramson, A. S. "Some effects of context on voice onset time in English stops", Language and Speech, 10(1):1–28, 1967.

[32] Cho, T. and Keating, P. "Effects of initial position versus prominence in English", Journal of Phonetics, 37(4):466–485, 2009.