Fillers and Creaky Voice Presence in Australian English

Hannah White, Joshua Penney, Felicity Cox

Department of Linguistics, Macquarie University, Australia
hannah.white@mq.edu.au; joshua.penney@mq.edu.au; felicity.cox@mq.edu.au

Abstract

The fillers *um* and *uh* have been shown to vary in terms of both their linguistic functions and their acoustics, including voice quality. The present study explores how, in Australian English, the likelihood of creaky voice presence differs depending on the phonetic realisation of the filler (*um* vs *uh*) and whether a token is a filler or a lexical item. Results suggest that different fillers do vary in their acoustics, and that social factors do not affect creaky voice presence on fillers in the same way that they do lexical items.

Index Terms: fillers, creaky voice, Australian English

1. Introduction

Fillers such as *um* and *uh* are non-lexical vocalisations, which are used widely cross-linguistically [1]. Fillers have been proposed to hold several functions in English, including planning upcoming speech, signalling upcoming delays and holding/ceding the floor [1, 2, 3], as well as the social meaning of uncertainty [3, 4].

Rates of filler use as well as preference for either *um* or *uh* vary according to macro-social categories such as gender, age and socioeconomic status [5, 6, 7, 8, 9, 10]. Across British English (BrE) [6, 7] and American English (AmE) [9] more fillers have been reported in men's speech compared to women's speech (although see [5]). Comparing *um* to *uh* across different Germanic languages, women and young speakers have been shown to prefer *um* to *uh* more than men and older speakers [6, 7, 8, 10], perhaps indicating a cross-linguistic change in progress towards *um* use as suggested by [8] and [10].

From a phonetic perspective, fillers have been investigated as potentially useful for forensic speaker comparisons as they are generally produced unconsciously and therefore less likely to be affected by a speaker's attempt to disguise their voice [11]. [11] examined formant frequencies and duration of *um* and *uh* for usefulness in forensic speaker discrimination in male speakers of BrE. Using the likelihood ratio framework, they determined that fillers were promising candidates for forensic speaker comparison when the first three formants and duration measurements were incorporated [11]. In a study of bilingual English-Māori speakers in New Zealand, [12] found that *um* and *uh* were speaker-specific in their formant and fundamental frequency (f0) characteristics with *um* discriminating between speakers more accurately compared to *uh*.

Language context can affect filler production. [13] examined the acoustics of *um* and *uh* in female L1 Dutch speakers learning English as L2. Both F1 and F2 differed between L1 Dutch and L2 English for both fillers; however, neither *um* nor *uh* differed in duration or f0 across the two language contexts [13]. Position in utterance influenced f0 of fillers with both *um* and *uh* having significantly higher f0 at the beginnings of utterances compared to in utterance-medial or final position or when surrounded by silence [13]. [14] examined the F1, F2, duration and f0 of fillers from eight different languages including Arabic, Mandarin Chinese, French and AmE. They found that while duration and f0 were relatively stable across languages, F1 and F2 frequencies were language-specific.

Many more studies have investigated the pitch characteristics of fillers across languages such as Spanish, German, French, Japanese and English [15, 16, 14, 17, 18, 19, 20]. For English, [18] found that fillers tended to have falling or level f0 contours and that their end points occurred within the lowest 15% of a speaker's f0 range regardless of whether they occurred at a phrase boundary or medially. [19, 20] provided evidence that the f0 of clause-internal fillers can be predicted by the f0 of the surrounding prosodic context, with fillers almost always having lower f0 compared to their surroundings. Position can also impact f0 with [21] finding that for both *um* and *uh* in Dutch, tokens in utterance-initial position had significantly higher f0 and longer duration than medial tokens (similar to [13]'s cross-linguistic findings). Studies have also shown differences between *um* and *uh* in Dutch, with *um* having higher mean f0 and longer mean durations [21].

More recently, studies have examined voice quality characteristics of fillers [22, 23]. [22] compared measures of voice quality between vowels in Japanese fillers and corresponding vowels in lexical items. For males, fillers had lower mean f0 compared to lexical items; however, there was no difference for females. For both males and females, fillers had higher H1-H2 ratios (correlated with a breathier voice quality) and higher jitter and shimmer (indicating more irregularity/aperiodicity in the signal) [22]. Similar studies comparing Chinese and Japanese showed cross-linguistic differences with lower jitter and shimmer in fillers compared to lexical items for Chinese speakers [24, 23]. Both [22] and [23] found that voice quality measures were important cues to classifying fillers versus lexical vowels. Studies of French and English have noted that fillers tend to be produced with creaky voice quality [9, 14]. Creaky voice (or creak) is a voice quality generally characterised by low f0 and irregular/aperiodic glottal pulses [25]. In French, [14] found that Praat with default settings failed to detect f0 more often in fillers compared to lexical items, particularly for men, which they attribute to "an unstable voice quality which could be either vocal fry, creaky or breathy". In English, fillers have been anecdotally noted to frequently co-occur with creak [9]. [9] points out that disfluencies such as fillers can cause problems for automatic speech recognition (ASR) technology. Increased knowledge of the acoustic correlates of fillers could lead to improvements of ASR algorithms [9].

To our knowledge, the phonetic characteristics of fillers in Australian English (AusE) is yet to be examined. The aim of the present study is to explore the relationship between creak and fillers in AusE through two analyses. The first explores whether there are differences between different fillers by comparing creak occurrence between *um* (/ɐm/) and *uh* (/ɐː/). The second compares creak presence between *um* and lexical items in the same phonetic context (i.e., /ɐ/+nasal). Creak can signal phrase/turn-finality across a number of languages, including AusE [26, 27, 28, 29]. In AusE, [30] found that male speakers of Aboriginal AusE produced creakier voice quality than non-indigenous males, while [31, 32] have shown that creak use in AusE adolescents varies according to speaker gender and language background, with speakers from a monolingual English background using more creak than those with at least one parent with a first language other than English (LOTE). Among English-only background speakers, female speakers were found to use more creak than male speakers with no gender difference among LOTE background speakers [31, 32]. As previous studies have shown the acoustic characteristics of fillers may differ according to gender and language context, as well as position in utterance, these variables are included in the present analyses [14, 13, 22, 23].

2. Methods

2.1. Data

The data used in this study consist of conversations between 131 AusE-speaking teenagers from different areas of Sydney extracted from the Multicultural Australian English - Voices of Sydney (MAE-VoiS) corpus [33]. One speaker identified as neither male or female and, due to a lack of statistical power, was excluded from analysis resulting in the inclusion of 130 speakers (61f; 69m). Speakers were aged between 15 and 19 years (mean age = 15.7). 27 speakers were from monolingual English households (16f; 11m) with the remaining from LOTE backgrounds (45f; 58m). Recordings ranged in length from 5 to 35 minutes (mean length = 16 minutes; total length = 20 hours). Conversations were guided by a trained research assistant (RA). All but four of the included speakers conversed with a partner matched for gender. Speakers were generally familiar with their conversational partner (i.e., they were in the same class at school) apart from in 13 cases where the speaker conversed only with the RA.

Creaky voice was identified using the automatic optimised Union method [34]. This method combines an approach based on identifying creak through low f0 (the AntiMode method) [35, 36] with one that uses other acoustic cues to creak such as a measure of spectral tilt and residual peak prominence (the Creak Detector algorithm) [37] and has been shown to significantly improve creak identification compared to when each tool is used on its own [34]. The Union method identifies anything coded by either AntiMode or Creak Detector as creaky voice and returns a binary creak decision for every 10 ms of speech data.

All conversations were automatically transcribed using the IBM Watson Speech to Text API (https://www.ibm.com/cloud/watson-speech-to-text). Orthographic transcripts were then manually corrected by RAs and checked by trained phoneticians. Final transcripts were processed through the MAUS automatic forced-aligner [38].

While automatic detection methods are useful for annotating large quantities of data, they are not infallible. For this reason, phoneme boundaries of *um*, *uh* and all lexical /ɐ/+nasal tokens were manually corrected. The Union method was then checked within these tokens, with tokens coded as either creaky (if they contained creak) or not creaky. In tokens where the vowel was word-initial, creak intervals of 50 ms or less at the onset of the segment were considered to be segmental glottalisation (for tokens with an overall duration less than 100 ms, segmental glottalisation was considered present if less than half the vowel contained creak) [40, 41]. Segmental glottalisation in initial position has been proposed as a strategy of prosodic strengthening in English and differs from phrasal creak, which can be linked to conveying social meaning [26, 31, 41]. For this reason, and the fact that it is often too brief to be identified by the Union method, tokens that only contained segmental glottalisation and no other creak were coded as not creaky.

For each token of *um*, *uh* and lexical /ɐ/+nasal, position in the utterance was manually coded. This process was guided by the method used in [13]. Tokens were coded as 'isolated' if they were the speaker's entire conversational turn. If they occurred at the end of a phonological phrase, they were coded as 'phrase-final'. Similarly, if they occurred at the beginning of a phonological phrase, they were coded as 'phrase-initial'. Finally, tokens were coded as 'phrase-medial' if they interrupted a phonological phrase. Following [14], tokens were excluded if they were shorter than 40 ms (0.3% of all tokens). The duration of each token (i.e., vowel for *uh*, vowel + nasal for *um* and lexical items) were extracted in milliseconds.

2.2. Analysis

Two analyses were carried out. These were conducted with generalised linear mixed effects regression (GLMER) modelling using the lme4 [42] and lmerTest [43] packages in R [44]. The first analysis compared creak presence in *um* versus *uh*. The dependent variable was whether creak was present or not. A model was built with fixed effects of word (*um* or *uh*), speaker gender, whether the speaker had a LOTE background or not (henceforth referred to as LOTE-BG), position in utterance and duration. Three-way interactions were included for word, gender and LOTE-BG and word, position and duration. Duration was *z*-scored. A random intercept was included for speaker with random slopes for word, position and duration. The full model returned a singular fit warning so a stepwise reduction approach was taken, systematically reducing non-significant interactions followed by non-significant fixed effects and random slopes, each time comparing models using ANOVAs to ensure the most parsimonious model. The final model syntax is shown in 1. 2491 tokens (*um*=1787; *uh*=704) were included.

$$\text{Creak} \sim \text{word} + \text{pos} + \text{dur} + (1 + \text{word} \mid \text{speaker}) \tag{1}$$

The second analysis compared lexical /ɐ/+nasal to *um* (*uh* and lexical /ɐː/ were not examined here due to the larger number of *um* tokens and, as shown below, in the first analysis *um* was found to be significantly more likely to be creaky than *uh*). The initial model was the same as that in the previous analysis except word type (*um* versus lexical /ɐ/+nasal) was included as a fixed effect instead of word and an additional random intercept for word was included. In this analysis, tokens in the 'isolated' condition were removed due to low numbers of lexical items in this position (n=11). The final model syntax is shown in 2. 2732 tokens (*um*=1438; lexical=1294) were included.

$$\begin{aligned}\text{Creak} \sim \text{word type} * \text{LOTE-BG} + \text{word type} * \text{pos} +\\ \text{LOTE-BG} * \text{gender} + \text{pos} * \text{dur} +\\ (1 \mid \text{word}) + (1 + \text{word type} \mid \text{speaker})\end{aligned} \tag{2}$$

3. Results

3.1. *Um* versus *Uh*

The output of model 1 is shown in Table 1. The model shows a significant effect of word, indicating that *um* tokens are more likely to be creaky than *uh* tokens. There was also a significant simple effect of duration showing that as tokens increased in duration, so did the likelihood of creak presence. Finally, there was a significant effect of position on the likelihood of creak being present, which is illustrated in Figure 1. Comparing the different position levels, tokens in phrase-initial position were significantly less likely to contain creak compared to all other positions. There were no significant differences in creak probability between tokens in isolated, phrase-medial and phrase-final positions; however, there was a non-significant trend towards tokens in medial position being more likely to contain creak compared to isolated (p=0.091) and final (p=0.063) positions.

Table 1: *GLMER model output showing effect of word, position in utterance and duration on creak presence. Significant effects are shaded. Ref. levels: word = uh; position = isolated.*

	Est.	Std. Err.	z-value	p-value
(Int.)	-0.45	0.16	-2.80	0.005
wordUm	0.43	0.14	3.07	0.002
posMed	0.25	0.15	1.69	0.091
posFin	-0.03	0.16	-0.16	0.876
posInt	-0.36	0.13	-2.77	0.006
dur	0.17	0.05	3.21	0.001

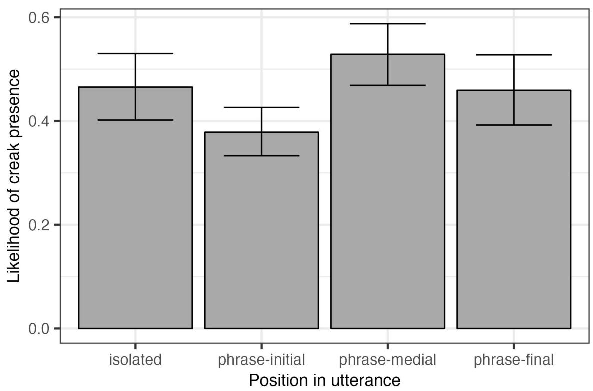

Figure 1: *GLMER model estimates of creak presence by position in utterance.*

3.2. *Um* versus lexical items

The output of model 2 is shown in Table 2. It shows a significant interaction between duration and position on creak likelihood. Figure 2 shows that tokens across all positions are generally more likely to be creaky the longer their duration; however, there is less of an effect of duration in initial position.

There was also a significant interaction of LOTE-BG and gender. *Post hoc* pairwise comparisons were carried out using the emmeans package [45] in R. These showed that regardless of word type, females with non-LOTE backgrounds were significantly more likely to creak than those with LOTE backgrounds

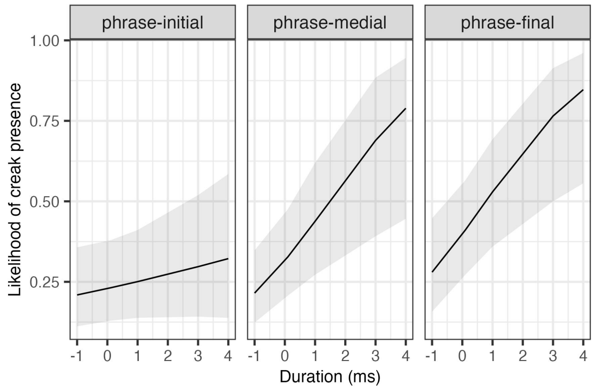

Figure 2: *GLMER model estimates of creak presence by position in utterance and token duration.*

Table 2: *GLMER model output showing effects of word type, LOTE-BG and position in utterance on creak presence. Significant effects are shaded. Ref. levels: word type = lexical; LOTE-BG = no; gender = female; position = phrase-medial.*

	Est.	Std. Err.	z-value	p-value
(Int.)	-0.38	0.39	-0.97	0.334
wordUm	0.54	0.68	0.79	0.428
LOTEYes	-1.68	0.41	-4.16	<0.001
genderM	-0.64	0.39	-1.64	0.101
posFin	1.16	0.27	4.27	<0.001
posInt	-0.48	0.46	-1.04	0.297
dur	0.52	0.17	3.07	0.002
wordUm:LOTEYes	1.30	0.37	3.52	<0.001
LOTEYes:genderM	1.10	0.44	2.48	0.013
wordUm:posFin	-1.53	0.42	-3.65	<0.001
wordUm:posInt	0.07	0.53	0.14	0.889
posFin:dur	0.01	0.24	0.04	0.973
posInt:dur	-0.41	0.20	-2.04	0.041

(p<0.001). Among speakers with a LOTE background, males were more likely to have creak than females (p=0.033).

Additionally, there was a significant interaction between word type and LOTE-BG, which is illustrated in Figure 3. Pairwise comparisons show creak was significantly more likely in *um* compared to lexical items among speakers with a LOTE background (p=0.026), but no difference for non-LOTE background speakers. Within lexical items, creak was significantly more likely in speakers with a non-LOTE background (p=0.001); however, LOTE-BG did not significantly affect creak likelihood in *um*.

Finally, there was a significant interaction between word type and position in utterance on creak likelihood, shown in Figure 4. Pairwise comparisons indicate that lexical items in phrase-final position were more likely to be creaky than those in phrase-initial (p=0.001) or phrase-medial (p<0.001) position. Among *um* tokens, there were no significant differences.

4. Discussion

This study has explored whether creak presence differs between the fillers *um* and *uh*, and between *um* and lexical /ɐ/+nasal. Focusing first on the filler only analysis, no social variables affected

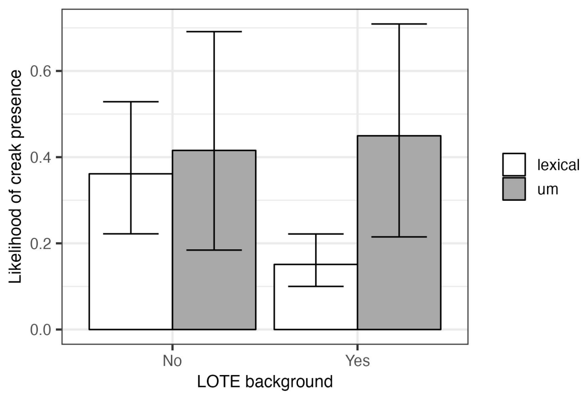

Figure 3: *GLMER model estimates of creak presence by word type and LOTE-BG.*

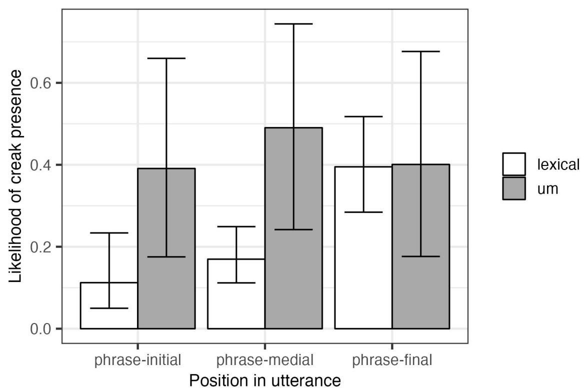

Figure 4: *GLMER model estimates of creak presence by word type and position in utterance.*

the use of creak suggesting that both male and female speakers from monolingual English and LOTE backgrounds use creak to similar degrees on *um* and *uh*. Tokens of *um* were significantly more likely to contain creak compared to *uh*. This supports findings from previous work showing that different fillers can vary in their acoustic characteristics, in this case voice quality [21, 11, 12]. The finding that both *um* and *uh* are less likely to contain creak in initial position is consistent with previous research showing that utterance-initial fillers have higher f0 than those later in the utterance [13, 21]. As low f0 is an acoustic characteristic of creak, it is possible that speakers are using creak to access their lower f0 range in tokens closer to the ends of utterances. Creak presence has been found to increase with phrase-finality (and turn-finality) across varieties of English, including AusE, and in other languages [27, 28, 29]. This could be a contributing factor to the duration effect where longer tokens were found to have higher likelihood of containing creak, due to the process of declination in f0: as airflow decreases across the course of an utterance (or in this case a long token), conditions that favour creak increase [26, 46]. It is important to bear in mind that this analysis does not take into account the location of the creak within a token and therefore further investigation of this point is needed. Although only approaching significance for isolated and final positions, the trend for medial tokens to have a higher likelihood of creak presence is consistent with research showing that clause-internal fillers are lower in f0 compared to their surrounding prosodic context [19, 20].

Turning to the analysis comparing *um* to lexical /ɐ/+nasal, the finding that female speakers with non-LOTE backgrounds were more likely to produce creaky tokens than female speakers with LOTE backgrounds, regardless of the type of item, is consistent with previous work on creak in AusE [31, 32]. While gender did not significantly impact creak by word type, LOTE-BG did have an effect. Interestingly, while speakers with a LOTE background had significantly less chance of creaking on lexical items compared to non-LOTE speakers, there was no difference in creak likelihood by LOTE-BG for *um*. This could suggest that on lexical items, creak can be used socially to signify aspects of a speakers' identity such as their cultural/linguistic heritage as suggested by [31]. However, creak is just as likely to occur on *um* for speakers of LOTE and non-LOTE backgrounds suggesting that these social meanings of creak are limited to when they occur on lexical items.

We now turn to significant interactions involving position in utterance. There was a significant interaction between position and duration regardless of word type. While tokens in initial position were less affected by duration when it comes to creak presence, those in medial or final positions were more likely to be creaky as duration increased. Again, this could be related to declination in f0: at the beginning of an utterance airflow is generally high whether tokens are short or long; however, as the utterance progresses, airflow decreases, increasing likelihood of creak, especially when tokens are long or drawn out [26, 46]. The interaction between word type and position suggests that while position has an effect on creak likelihood in lexical items, this is not the case for *um*. Higher probability of creak presence on lexical items in final position compared to initial or medial is consistent with previous studies suggesting creak is a marker of phrase-finality [27, 28, 29].

5. Conclusions

The present study has suggested that, in AusE, there is a relationship between creaky voice presence and the fillers *um* and *uh*. While previous work has shown that *um* and *uh* differ in their acoustics with regard to f0 and duration [11, 12, 21], we have shown that this finding extends to creaky voice quality. Additionally, our findings present evidence that *um* differs in the presence of creaky voice compared to lexical /ɐ/+nasal sequences. These results have potential implications for improving ASR technology if creaky voice quality can be used to aid in distinguishing between fillers and lexical items. It is also possible that creak could be considered alongside formant, duration and f0 information in using fillers for forensic use, especially considering recent research suggests interspeaker variation exists in the acoustics of creak in Dutch [47]. Results suggest that although macro-social factors play a role in the likelihood of creak occurring on lexical items, they do not have the same effect for *um*. This could be due to potential sociolinguistic functions of fillers such as conveying uncertainty [3, 4], which has also been linked to creaky voice [48]; however further research is required to explore this.

6. Acknowledgements

This research was supported by Australian Research Council Grant DP190102164 and Australian Research Council Future Fellowship Grant FT180100462 to the fifth author.

7. References

[1] Clark, H. H. and Fox Tree, J. E., "Using uh and um in spontaneous speaking," *Cogn.*, vol. 84, pp. 73–111, 2002.

[2] Levelt, W. J. M., "Monitoring and self-repair in speech," *Cogn.*, vol. 14, no. 1, pp. 41–104, 1983.

[3] Smith, V. L. and Clark, H. H., "On the course of answering questions," *J. Memory Lang.*, vol. 32, no. 1, pp. 25–38, 1993.

[4] Kirkland, A., Lameris, H., Székely, É., and Gustafson, J., "Where's the uh, hesitation? The interplay between filled pause location, speech rate and fundamental frequency in perception of confidence," in *Proc. Interspeech*, 2022, pp. 4990–4994.

[5] Tottie, G., "On the use of *uh* and *um* in American English," *Functions of Lang.*, vol. 21, no. 1, pp. 6–29, 2014.

[6] Foulkes, P., Carrol, G., and Hughes, S., "*Sociolinguistics and acoustic variability in filled pauses* [Conference presentation]," Ann. Conf. Int. Assoc. Forensic Phonetics and Acoustics, 2004.

[7] Tottie, G., "*Uh* and *Um* as sociolinguistic markers in British English," *Int. J. Corpus Linguistics*, vol. 16, no. 2, pp. 173–197, 2011.

[8] Acton, E. K., "On gender differences in the distribution of um and uh," *University of Pennsylvania Working Papers in Linguistics*, vol. 17, no. 2, p. article 2, 2011.

[9] Shriberg, E., "To 'errrr' is human: Ecology and acoustics of speech disfluencies," *J. Int. Phonetic Assoc.*, vol. 31, no. 1, pp. 153–169, 2001.

[10] Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J., and Liberman, M., "Variation and change in the use of hesitation markers in Germanic languages," *Lang. Dyn. Change*, vol. 6, no. 2, pp. 199–234, 2016.

[11] Hughes, V., Wood, S. and Foulkes, P., "Strength of forensic voice comparison evidence from the acoustics of filled pauses," *Int. J. Speech, Lang. Law*, vol. 23, no. 1, pp. 99–132, 2016.

[12] Wong, S. G.-J. and Papp, V., "*Transferability of non-lexical hesitation markers across languages: Evidence from te reo Māori-English bilinguals* [Conference presentation]," 27th Ann. Conf. Int. Assoc. for Forensic Phonetics and Acoustics, 2018.

[13] de Boer, M. M. and Heeren, W. F. L., "Cross-linguistic filled pause realization: The acoustics of uh and um in native Dutch and non-native English," *J. Acoust. Soc. Am.*, vol. 148, pp. 3612–3622, 2020.

[14] Candea, M., Vasilescu, I., and Adda-Decker, M., "Inter- and intra-language acoustic analysis of autonomous fillers," in *Proc. DiSS*, 2005, pp. 47–52.

[15] Adell, J., Bonafonte, A. and Escudero-Mancebo, D., "Modelling filled pauses prosody to synthesise disfluent speech," in *Proc. Speech Prosody*, 2010.

[16] Belz, M. and Reichel, U. D., "Pitch characteristics of filled pauses," in *Proc. DiSS, ICPhS Satellite Meeting*, 2015.

[17] Maekawa, K., "Prediction of f0 height of filled pauses in spontaneous Japanese: a preliminary report," in *Proc. DiSS*, 2013, pp. 41–44.

[18] O'Shaughnessy, D., "Recognition of hesitations in spontaneous speech," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*. IEEE, 1992, pp. 521–524.

[19] Shriberg, E. E. and Lickley, R. J., "The relationship of filled-pause f0 to prosodic context," in *Proc. IRCS Workshop on Prosody in Natural Speech*, 1992, pp. 201–209.

[20] ——, "Intonation of clause-internal filled pauses," *Phonetica*, vol. 50, pp. 172–179, 1993.

[21] Swerts, M., "Filled pauses as markers of discourse structure," *J. Pragmatics*, vol. 30, pp. 485–496, 1998.

[22] Maekawa, K. and Mori, H., "Voice-quality difference between the vowels in filled pauses and ordinary lexical items," in *Proc. Interspeech*, 2016, pp. 3171–3175.

[23] Li, X., Ishi, C. T., Fu, C., and Hayashi, R., "Prosodic voice quality analyses of filled pauses in Japanses spontaneous conversation by Chinese learners and Japanese native speakers," in *Proc. Speech Prosody*, 2022, pp. 550–554.

[24] Maekawa, K., 'ya Nishikawa, K., and Tseng, S.-C., "Phonetic characteristics of filled pauses: A preliminary comparison between Japanese and Chinese," in *Proc. DiSS*, 2017, pp. 41–44.

[25] Keating, P., Garellek, M., and Kreiman, J., "Acoustic properties of different kinds of creaky voice," in *Proc. 18th ICPhS*, The Scottish Consortium for ICPhS 2015, Ed. University of Glasgow, 2015, Conference Proceedings, pp. 821.1–5.

[26] Podesva, R. J., "Gender and the social meaning of non-modal phonation types," *Ann. Meet. Berkeley Linguistics Soc.*, vol. 37, no. 1, pp. 427–448, 2013.

[27] Redi, L. and Shattuck-Hufnagel, S., "Variation in the realization of glottalization in normal speakers," *J. Phonetics*, vol. 29, no. 4, pp. 407–429, 2001.

[28] Ogden, R., "Turn transition, creak and glottal stop in finnish talk-in-interaction," *J. Int. Phonetic Assoc.*, vol. 31, no. 1, pp. 139–152, 2001.

[29] White, H., Penney, J., Gibson, A., Szakay, A., and Cox, F., "Creak prevalence and prosodic context in Australian English," in *Proc. Interspeech*, 2023.

[30] Loakes, D. and Gregory, A., "Voice quality in Australian English," *JASA Express Letters*, vol. 2, no. 8, p. 085201, 2022.

[31] White, H., Gibson, A., Penney, J., Szakay, A., and Cox, F., "*Differences in prevalence of creaky voice in mono- and multilingual communities in Sydney* [Conference presentation]," 14th Int. Symposium on Bilingualism, Sydney, Australia, 2023.

[32] White, H., "Creaky voice in Australian English," [PhD thesis, Macquarie University], 2023.

[33] Cox, F. and Penney, J., "Multicultural Australian English - The new voice of Sydney," *Aus. J. Linguistics*, forthcoming.

[34] White, H., Penney, J., Gibson, A., Szakay, A., and Cox, F., "Evaluating automatic creaky voice detection methods," *J. Acoust. Soc. Am.*, vol. 152, no. 3, pp. 1476–1486, 2022.

[35] Dorreen, K., "Fundamental frequency distributions of bilingual speakers in forensic speaker comparison," [Master's thesis, University of Canterbury], 2017.

[36] Dallaston, K. and Docherty, G., "Estimating the prevalence of creaky voice: A fundamental frequency-based approach," in *Proc. 19th ICPhS*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds. ASSTA Inc, 2019, pp. 581.1–5.

[37] Drugman, T., Kane, J., and Gobl, C., "Data-driven detection and analysis of the patterns of creaky voice," *Comput. Speech Lang.*, vol. 28, no. 5, pp. 1233–1253, 2014.

[38] Schiel, F., Draxler, C., and Harrington, J., "Phonemic segmentation and labelling using MAUS technique," in *New tools and methods for very-large-scale phonetics research workshop*, Philadelphia, PA, USA, 2011.

[39] Boersma, P. and Weenink, D., "Praat: Doing phonetics by computer [Computer program]," 2023, version 6.3.09. Online: http://www.praat.org/

[40] Garellek, M., "Perception of glottalization and phrase-final creak," *J. Acoust. Soc. Am.*, vol. 137, no. 2, pp. 822–831, 2015.

[41] Dilley, L., Shattuck-Hufnagel, S., and Ostendorf, M., "Glottalization of word-initial vowels as a function of prosodic structure," *J. Phonetics*, vol. 24, no. 4, pp. 423–444, 1996.

[42] Bates, D., Maechler, M., Bolker, B., and Walker, S., "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, 2015.

[43] Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B., "lmerTest package: Tests in linear mixed effects models," *J. Stat. Softw.*, vol. 82, no. 13, pp. 1–26, 2017.

[44] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2022. Online: https://www.R-project.org/

[45] Lenth, R., "emmeans: Estimated marginal means, aka least-squares means," 2018, R package version 1.8.9. Online: https://CRAN.R-project.org/package=emmeans

[46] Ladd, D. R., "Declination: A review and some hypotheses," *Phonology Yearbook*, vol. 1, pp. 53–74, 1984.

[47] van Hugte, T. B. R. and Heeren, W. F. L., "Exploring interspeaker variation in creaky voice in Dutch," *J. Voice*, 2024.

[48] Yuasa, I. P., "Creaky Voice: A New Feminine Voice Quality for Young Urban-Oriented Upwardly Mobile American Women?" *Am. Speech*, vol. 85, no. 3, pp. 315–337, 2010.