Sociophonetic Variation of Filled Pauses in Victoria, Australia

Liz Blackwell, Kirsty McDougall

University of Cambridge

ecb76@cam.ac.uk; kem37@cam.ac.uk

Abstract

An acoustic study of sociophonetic variation in the fillers 'um' and 'uh' in Australian Englishes (Aboriginal, AAE; Mainstream, MAE) spoken in Warrnambool, Victoria, is presented. Duration, f0 and formant frequency measurements of fillers produced by 14 AAE and 14 MAE speakers were analysed. 'uh' and 'um' exhibited different vowel qualities which also varied with gender, variety and age. 'um' demonstrated a longer vowel duration than 'uh' for male MAE speakers only. Female MAE speakers produced 'um' with a higher f0 than female AAE speakers; no such pattern was present for male speakers, nor for 'uh' produced by any speaker group.

Index Terms: fillers, filled pauses, fluency features, Australian Englishes, Australian Aboriginal Englishes

1. Introduction

1.1. Australian Englishes

At least 200 languages were spoken in Australia before European colonisation [1, 2], but following colonial suppression of First Nations culture and language and the ensuing dominance of English, there now remain an estimated 20 Indigenous languages, half of which are spoken by fewer than 1000 speakers [3]. A typology of contemporary Australian Englishes [4] suggests a broad tripartite division into the supra-regional standard ('Mainstream' Australian English, MAE), Aboriginal Australian Englishes (AAEs) and, outside the scope of the present paper, Ethnocultural Australian Englishes such as Greek or Yiddish Australian English [5].

Within MAE and AAE, there is a significant amount of social and geographical variation. MAE has historically been described as encompassing variation between 'broad' and 'cultivated' sociolects, but this is now considered outdated, and suggested to have been replaced by a new "general" prestige variety, with social and regional variation undergoing reorganisation in accordance with Australia's developing post-colonial identity [6]. Regional variation is subtle, but increasingly noted in phonetic studies [7, 8]

AAEs are a distinct group of Englishes that exhibit variation from MAE at many levels of linguistic structure, including phonology and lexicon [3] as well as notable differences in pragmatic and discoursal features [3, 9, 10]. It has been noted that AAEs may differ from MAE with respect to filler words and phrases such as "might be" or "like" [3], but 'um' and 'uh' are not mentioned in these observations. Additionally, crucial differences in the use of silence and pausing behaviour in AAEs, namely a positive view of silence and pausing in conversation by Indigenous speakers, have been documented in courtroom contexts [9]. Here, "silence fillers" such as 'um' and 'uh' are suggested to be a strategy employed more in western cultures, though a lower rate of filler use by Indigenous speakers is not directly demonstrated.

1.2. Social variation in 'um' and 'uh' distribution

Existing studies of social variation in filler use have focused on non-acoustic variables such as relative proportion and frequency. Frequency has been shown to correlate with both gender and age: in American English, men have repeatedly demonstrated a higher frequency of filler usage [11, 12], and increasing age correlates with more frequent filler use [12]. The relative proportion of 'um' and 'uh' has been shown to correlate with age and gender in American English [13], with male speakers and older speakers having a higher relative proportion of 'uh' than their female or younger counterparts. The same has been shown to be true in British English [14], together with a higher overall filler frequency in male speech, and greater relative use of 'uh' by both younger male speakers and older speakers of any gender.

Clearly, fillers may vary along social lines. However, from a forensic perspective, filler frequency has also been shown to be highly speaker-specific within a socially homogeneous group [15], with rate of 'um' production allowing for discrimination between male Southern Standard British English (SSBE) speakers at a level three times greater than chance.

1.3. Acoustic parameters of 'um' and 'uh'

The vowel quality of fillers differs across languages [16, 17], even within bilingual speakers when switching language [16]. Filler duration has also been shown to be cross-linguistically variable, with French fillers being demonstrably longer than their German equivalents [16]. However, there is little research available on these factors at the sociolinguistic level. In forensic research, acoustic properties of 'um' and 'uh' have been shown to be comparable to lexical vowels in their capacity for speaker differentiation [18], using F1-F3 formant frequencies, vowel duration, and nasal duration in 'um'. Whilst both 'um' and 'uh' were investigated, 'um' was found to be a greater locus of individual variation, whereas 'uh' was more consistent across speakers, all of whom were young, male SSBE speakers [19]. Differences in acoustic features of 'um' and 'uh' have also been reported. In French and German [16], and again in SSBE [18], the vocalic portion of 'um' is consistently shorter than the duration of 'uh', though 'um' is the longer of the two when the nasal portion is included. There have been varying conclusions on whether 'uh' and 'um' have different vowel qualities. F2 frequency has been reported significantly higher for 'uh' in German, and F1 frequency significantly higher for 'um' in both French and German [16].

However in British English, static formant frequency measurements of 'um' and 'uh' have been suggested to differ very little [18].

1.4. Aims

Social variation has been demonstrated in filler choice and their frequency of occurrence, and trends appear consistent across British and American English. However to date, acoustic variation in filler realisation has only been shown across languages, or between individuals, rather than socially delimited. The present study therefore examines a group of speakers from Warrnambool, Victoria, investigating whether the factors of gender, age, or variety of Australian English (MAE or AAE), correlate with the variables of duration, f0, and F1-F2 formant frequencies.

2. Method

2.1. Speakers

The speech analysed in the present study forms part of a set of sociolinguistic interviews conducted in 2015 and 2016 by Debbie Loakes at the University of Melbourne. The subset of data used in this study represents informal, spontaneous speech from 29 residents of Warrnambool, a regional centre in south-western Victoria. Warrnambool, around 3 hours' travel from Melbourne, is a small city of 35,000 residents, of which 2% - double the state average - identify as Indigenous [20]. One male MAE speaker, who produced no clear fillers in the course of his interview, was excluded from further analysis. Of the 28 remaining speakers, 14 are speakers of AAE (9 female, 5 male), and 14 speakers of MAE (7 female, 7 male). Ages range from 18 to 72 years, with a mean of 36; however, the mean age of the MAE speakers is 42, in contrast to 30 for the AAE speakers. Length of speech material from each specific participant varied greatly, ranging from 47 seconds to 30 minutes, with a median of 3 minutes and 46 seconds.

2.2. Data collection

Transcription was carried out in *Praat* [21], to tiered textgrids. The first tier was used to indicate when the target speaker began and ceased speech, and filler tokens were marked for more detailed transcription on the second tier. Following the process of [18], tokens of 'um' were split into vocalic and nasal portions, where visual inspection of the spectrogram mostly showed a clear nasal onset with sharp decreases in the amplitudes of formant frequencies. Where not spectrographically visible, onsets were marked using auditory analysis. After discarding any tokens in overlapping speech, 1204 filler tokens were available for analysis, with 310 tokens of 'uh' and 894 tokens of 'um'.

2.4. Analysis

Duration, fundamental frequency (f0), and midpoint F1-F2 frequency values were extracted using *Praat* scripting. For 'um', measurements were taken at the midpoint of the vocalic portion. For the purposes of this paper, the vowel in both 'um' and 'uh' was taken to be a short vowel, /ɐ/. Whilst the presence of nasalisation in 'um' may also affect vowel quality, in SSBE the coarticulatory effect of /m/ in 'um' has been shown to only begin around 70%-90% of the way through the /ɐ/ vowel [18], and so static midpoint measurements are less likely to demonstrate any significant effect from nasal coarticulation. Tokens returning 'undefined' or outlier f0 values were checked and discarded if found to have non-modal voice quality, removing 146 tokens from further analysis, leaving a total sample of 1058 tokens. Over half of these discarded tokens were from a single speaker with a particularly 'creaky' speech style, but over 100 clear tokens from this speaker remained for analysis. Of the 1058 tokens available for further analysis, 543 were produced by female speakers (53 'uh', 489 'um'), and 515 by male speakers (188 'uh', 344 'um'). However, due to differences in interview lengths, there was an uneven distribution of fillers between dialect groups, with 708 tokens by MAE speakers (193 uh, 531 um) and 350 tokens by AAE speakers (48 uh, 302 um) available for analysis.

Data were fitted to a series of linear mixed-effects models in *RStudio* [22], using the *lme4* package [23]. Dialect, gender, age and filler type were each used as categorical predictors, with speakers divided into two age groups of <40 years and >40 years. For models including f0 and F1-F2 formant frequencies, analysis of male and female speakers was carried out separately. A random intercept for speaker ID was also included in each model. The *RStudio* package *gtsummary* [24] was then used to estimate *p*-values resulting from the *t*-values in the output produced by the models.

3. Results

3.1. Duration

For the group at large, the vowel of 'um' was shorter than that of 'uh' ('um' mean = 0.29s, 'uh' mean = 0.33s, $\beta = -0.02s,\ p < 0.038$), and the total length of 'um' was significantly longer than that of 'uh' ('um' mean = 0.48s, 'uh' mean = 0.33s, $\beta = 0.19s,\ p < 0.001$). However, male MAE speakers had noticeably different results: the vowel in 'um' produced by these speakers was significantly *longer* than that of 'uh' ('um' mean = 0.36s, 'uh' mean = 0.30s, $\beta = 0.04s,\ p < 0.001$), leading to a greater durational difference between the two fillers for male MAE speakers.

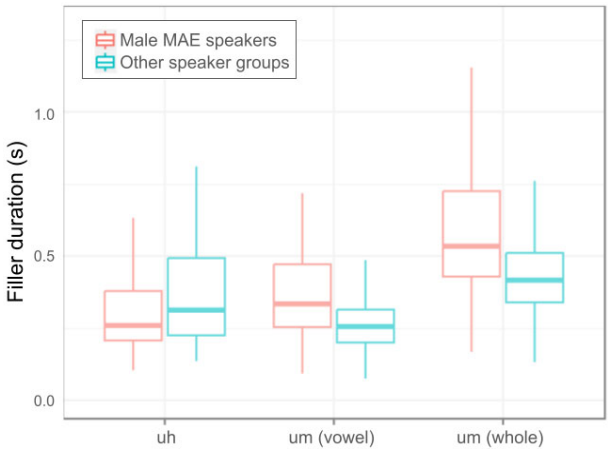

Figure 1: *Relative durations of 'uh' and 'um'.*

3.2. F0

'Um' produced by female speakers showed a significant dialect effect, with female speakers of MAE producing 'um' with a higher average f0 than female speakers of AAE (MAE mean = 181Hz, AAE mean = 156Hz, $\beta = 47Hz, p = 0.009$).

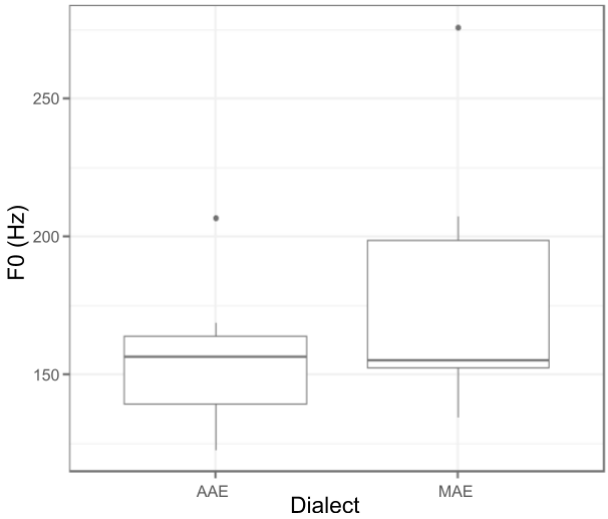

Figure 2: *F0 of 'um' by female speakers.*

There was no such interaction for 'uh' by female speakers, nor any such interaction for either filler in the male speaker group.

3.3. F1 and F2 formant frequencies

F2 of 'uh' in female MAE speakers was significantly higher in frequency than 'um', corresponding to a 'fronted' vowel quality ('um' mean = 821Hz, 'uh' mean = 784Hz, $\beta = 113Hz, p < 0.001$).

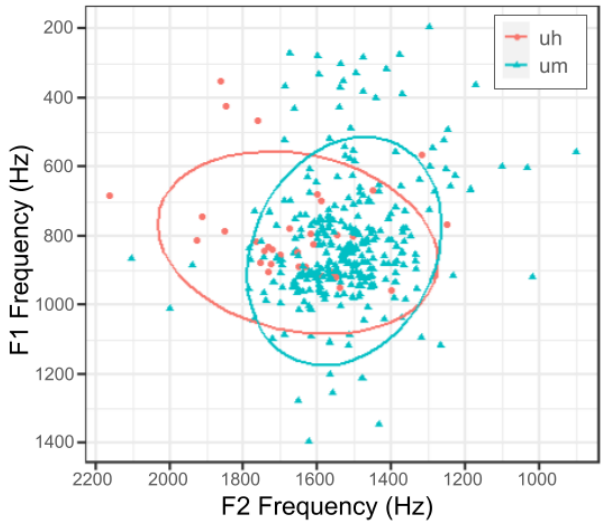

Figure 3: *95% confidence ellipses for 'uh' and 'um' produced by female MAE speakers.*

F1 of 'uh' produced by female MAE speakers under 40 was much higher than 'uh' in those over 40 (<40 mean = 851Hz, >40 mean = 698Hz, $\beta = 153Hz, p < 0.001$), corresponding to a lowered 'uh' vowel in younger female speakers.

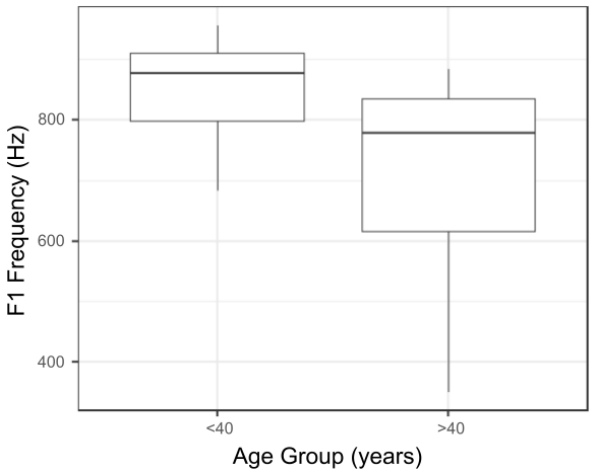

Figure 4: *F1 frequency of 'uh' by female MAE speakers by age group.*

Midpoint F1 and F2 frequencies of individual tokens of 'uh' and 'um' produced by male speakers of MAE are shown in Figure 5, with 95% confidence ellipses. Male speakers tended to have a higher F1 frequency in 'uh' than 'um', a trend reaching statistical significance for MAE speakers ('uh' mean = 598Hz, 'um' mean = 616Hz, $\beta = 27Hz, p = 0.005$), but not AAE speakers ($\beta = 39Hz, p = 0.065$).

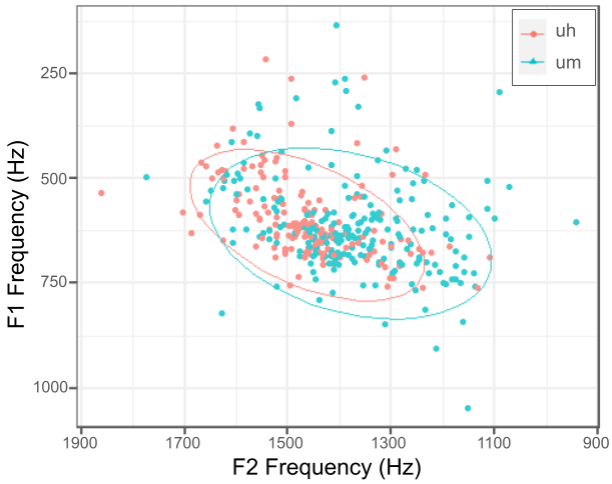

Figure 5: *95% confidence ellipses for F1 and F2 frequencies of 'uh' and 'um' produced by male speakers of MAE.*

4. Discussion

4.1. Relative duration findings

As discussed in section 1.3, previous studies of relative duration in 'um' and 'uh' [16, 18] have consistently shown the vowel in 'um' to be shorter than 'uh'. However, the findings

in section 3.1 show that although AAE speakers and female MAE speakers maintain this relationship, male MAE speakers produce 'um' with a longer vowel than 'uh', presenting the possibility that relative duration may vary according to dialectal factors. Additionally, both [18] and [16] worked with single-sex samples, and so the present study demonstrates the potential importance of gender as a factor in filler duration. A major factor affecting segmental duration is placement within the intonational phrase (IP) [25], a factor not considered in the present study. Therefore, further explanation of this finding would require examination of the positions of 'uh' and 'um' within the IP. Furthermore, participants' individual speech rates were not calculated, but no significant gender, dialect or age differences in filler frequency per minute were found [26].

4.2. F0 findings

Previous research on f0 and Warrnambool dialect groups has suggested a lower average f0 for male AAE speakers [27]. However, the present study did not find any such trend for male speakers, and f0 only displayed a significant drop for 'um' in female AAE speakers. As 'uh' displayed no such trend (p = 0.2), differences in prosodic placement once again present a potential factor, requiring further investigation. The lack of any dialect trend in f0 concurrent with wider literature is possibly due to the restricted style and content of speech in the present study.

4.3. F1 and F2 formant frequency findings

4.3.1. Differences in placement of 'uh' and 'um'

As discussed in section 1.3, differences in vowel quality between 'um' and 'uh' have been reported in French and German [16], but were not found in SSBE [18]. However, the present study provides preliminary evidence that 'uh' and 'um' may have significant, if slight, differences in vowel placement. Based on findings for F1 (section 3.3), male MAE speakers produced 'uh' with a slightly higher vowel quality than 'um', a trend potentially present, but below the level of significance, for male AAE speakers. In contrast, findings for F2 showed that for female MAE speakers, 'uh' had a more fronted quality than 'um' (Fig. 3). Therefore, 'uh' and 'um' may differ in vowel placement in some dialects of English, and this placement may also be sensitive to social factors such as gender. However, if onset of nasalisation in AusE is comparatively earlier than that of SSBE, where /ʊ/ in 'um' typically begins to nasalise in the final third of the vowel [18] then it is also possible that nasal coarticulation had an effect on the midpoint formant frequencies.

4.3.2. Age effect and potential connection to short front vowel lowering

An age effect was also present for female MAE speakers, with younger participants producing 'uh' with much lower vowel quality (Fig. 4). No such trend was demonstrated for speakers of AAE, or for 'um'. This finding potentially relates to the ongoing change of short front vowel lowering found in Sydney speakers of MAE [28], a change that is not particularly evident in the speech of AAE speakers in Warrnambool [29]. The exclusivity of the trend to 'uh', and not 'um', may relate to the fronting of 'uh' by these speakers as discussed in section 4.3.1, or potentially may be due to the apparent resistance of nasalised vowels to the change, as previously demonstrated for nasalised [æ̃] [30]. However, the small sample size of 'uh' from female MAE speakers (n = 33 tokens) may have also been a contributing factor, so further data are necessary for any comprehensive investigation.

5. Conclusions

This paper has presented a preliminary investigation of sociophonetic variation in filled pauses produced by speakers of two varieties of Australian English, MAE and AAE. Gender, dialect and age were shown to be potential factors of variation in vowel duration, f0, and vowel placement. Findings also support the treatment of 'uh' and 'um' as separate entities when investigating vowel quality of fillers, as vowel quality differed between fillers, and relative differences in vowel quality were also shown to be socially variable. The vowel duration of 'um' was longer than 'uh' in male MAE speakers, a novel finding that contrasts with previous studies of British English, French and German. Whilst a tentative connection was made between short front vowel lowering and the lowered quality of 'uh' for some speakers, further research and a larger dataset are necessary for any strong conclusion. For any further research regarding social variation in duration and f0, investigation of the relationship between social factors and prosodic placement of fillers is also needed.

6. Acknowledgements

The authors are grateful to Debbie Loakes, The University of Melbourne, for making the data set available for this study.

7. References

[1] Malcolm, I. G., "Aboriginal English research: an overview", Asian Englishes, 3(2):9–31, 2000.

[2] Burridge, K., "History of Australian English", in L. Willoughby and H. Manns [Eds], Australian English Reimagined, 175–192, Routledge, 2019.

[3] Butcher, A., "Linguistic aspects of Australian Aboriginal English", Clinical Linguistics & Phonetics, 22(8):625–642, 2008.

[4] Cox, F. and Fletcher, J., Australian English Pronunciation and Transcription, Cambridge University Press, 2017.

[5] Clyne, M., Eisikovits, E., and Tollfree, L. "Ethnic varieties of Australian English", in D. Blair and P. Collins [Eds], English in Australia, 223–238, John Benjamins, 2001.

[6] Cox, F. and Palethorpe, S., "Standard Australian English: the sociolinguistic broadness continuum", in R. Hickey [Ed], Standards of English, 294–317, 2012.

[7] Cox, F. and Palethorpe, S., "Vowel variation across four major Australian cities", in S. Calhoun, P. Escudero, M. Tabain and P. Warren [Eds], Proc. 19th International Congress of Phonetic Sciences 2019, 577–581, ASSTA, 2019.

[8] Loakes, D., Clothier, J., Hajek, J. and Fletcher, J., "An investigation of the /el/-/æl/ merger in Australian English: a pilot study on production and perception in South-West Victoria", Australian Journal of Linguistics, 34(4):436–452, 2014.

[9] Eades, D., "Understanding Aboriginal silence in legal contexts", in H. Kotthoff and H. Spencer-Oatey [Eds], Handbook of Intercultural Communication, 285–301, Mouton de Gruyter, 2007.

[10] McDougall, K., Paver, A., Duckworth, M., Blackwell, L. and Loakes, D. "Patterns of silent pausing in Aboriginal and

Mainstream Australian Englishes spoken in Warrnambool", Proc. 19th Australasian International Conference on Speech Science and Technology, Melbourne, this volume, 2024.
[11] Shriberg, E., "Disfluencies in Switchboard", Proc. International Conference on Spoken Language Processing, 11–14, Philadelphia, PA: IEEE, 1996.
[12] Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F. and Brennan, S. E., "Disfluency rates in conversation: effects of age, relationship, topic, role, and gender", Language and Speech, 44(2):123–147, 2001.
[13] Acton, E. K., "On gender differences in the distribution of um and uh", University of Pennsylvania Working Papers in Linguistics, 17(2):2, 2011.
[14] Tottie, G., "Uh and um as sociolinguistic markers in British English", International Journal of Corpus Linguistics, 16(2):173–197, 2011.
[15] McDougall, K. and Duckworth, M. "Profiling fluency: an analysis of individual variation in disfluencies in adult males", Speech Communication, 95:16–27, 2017.
[16] Lo, J. J. H., "Between äh(m) and euh(m): the distribution and realization of filled pauses in the speech of German-French simultaneous bilinguals", Language and Speech, 63(4):746–768, 2020.
[17] Candea, M., Vasilescu, I. and Adda-Decker, M., "Inter- and intra-language acoustic analysis of autonomous fillers", Proc. Disfluency in Spontaneous Speech Workshop, 47–51, Aix-en-Provence, 2005.
[18] Hughes, V., Wood, S. and Foulkes, P., "Strength of forensic voice comparison evidence from the acoustics of filled pauses", International Journal of Speech, Language and the Law, 23(1):99–132, 2016.
[19] Nolan, F., McDougall, K., de Long, G. and Hudson, T., "The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research", International Journal of Speech, Language and the Law, 16(1):31–57, 2009.
[20] Australian Bureau of Statistics, "2021 Census All persons QuickStats", 2021. Online: https://www.abs.gov.au/census/find-census-data/quickstats/2021/2021, accessed on 07 Jul 2024.
[21] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer", 2024. Online: http://www.praat.org/.
[22] Posit Team, "RStudio: integrated development environment for R." Posit Software, PBC, Boston, MA, 2023. Online: http://www.posit.co/.
[23] Bates, D., Mächler, M., Bolker, B. and Walker, S., "Fitting linear mixed-effects models using lme4.", Journal of Statistical Software, 67(1):1–48, 2015.
[24] Sjoberg, D. D., Whiting, K., Curry, M., Lavery, J. A. and Larmarange, J., "Reproducible summary tables with the gtsummary package", The R Journal, 13(1):570–580, 2021.
[25] Nolan, F., "Intonation", in B. Aarts and A. McMahon [Eds], The Handbook of English Linguistics, 433–457, Blackwell, 2006.
[26] Blackwell, L., "A study of sociophonetic variability in filled pauses in Southeastern Australia", Undergraduate dissertation, University of Cambridge, 2024.
[27] Loakes, D. and Gregory, A., "Voice quality in Australian English", Journal of the Acoustical Society of America Express Letters, 2(8):085201, 2022.
[28] Cox, F. and Palethorpe, S., "Reversal of Short Front Vowel Raising in Australian English", Proc. Annual Conference of the International Speech Communication Association, 342–345, Interspeech, 2008.
[29] Loakes, D., Fletcher, J., Hajek, J., Clothier, J. and Volchok, J. "Short vowels in L1 Aboriginal English spoken in Western Victoria.", in C. Carignan and M. Tyler [Eds], Proc. 16th Australasian International Conference on Speech Science and Technology, 33–36, ASSTA, 2016.
[30] Cox F. and Palethorpe, S., "Phonologisation of vowel duration and nasalised /æ/ in Australian English", in J. Hay and E. Parnell [Eds], Proc. 16th Australasian International Conference on Speech Science and Technology, 33–36, ASSTA, 2014.