CONVERGENCE OF CREAKY VOICE USE IN AUSTRALIAN ENGLISH

Hannah White, Joshua Penney, Andy Gibson, Anita Szakay, Felicity Cox

Centre for Language Sciences, Department of Linguistics, Macquarie University
hannah.white2@hdr.mq.edu.au, joshua.penney; andy.gibson; anita.szakay; felicity.cox@mq.edu.au

ABSTRACT

Studies of phonetic accommodation have shown that speakers imitate features of their interlocutors' speech such as vowel production, voice onset time and fundamental frequency. While some individual acoustic parameters of voice quality have been examined in accommodation research, a global measure of creaky voice quality accommodation has not yet been investigated. Creak is a voice quality that appears to be increasingly prevalent in many varieties of English including Australian English and carries several social and linguistic meanings. The present study investigated whether 30 pairs of teenagers from Sydney converged in their use of creaky voice across the course of a conversation. Creaky voice was identified using automatic methods of detection. Findings suggest that creaky voice convergence was present among these speakers. Additionally, an effect of speaker sex showed females used creak at more similar rates to each other than males.

Keywords: Phonetic accommodation, creaky voice, voice quality, Australian English.

1. INTRODUCTION

It has been established that speakers imitate phonetic features of their (assumed) interlocutors in both conversational and laboratory settings and with both natural and synthetic speech in a process referred to as phonetic accommodation [2, 3, 5, 12, 24, 25]. Growing evidence suggests that accommodation is an unconscious but not automatic process influenced by many social factors [2, 3, 13, 14, 25]. Communication Accommodation Theory (CAT), as described in [13], proposes that speakers express their attitudes towards one another through strategies of convergence (their speech becomes more similar) and divergence (their speech becomes less similar). Babel [2] found evidence supporting CAT, showing that for target vowels, New Zealand English (NZE) speakers accommodated to Australian English (AusE) speakers in a shadowing task. She also found that the extent of convergence was influenced by the sentiments the NZE-speaker held about Australians. Those with more positive sentiments converged more than those with negative sentiments [2]. Dialect has also been shown to influence degree of accommodation with speakers converging more to accents different from their own [34]. Another shadowing task found that the degree of liking (measured through attractiveness ratings) influenced how much listeners converged/diverged their vowel qualities to/from model speakers in American English (AmE) [3]. Studies using more natural speech contexts have shown similar results [23, 25, 26]. Pardo [23] found that in map tasks, same-sex pairs of AmE-speakers converged in their vowel productions across the course of the conversation and that degree of convergence was influenced by speaker sex and role in the map task: female instruction givers converged more towards female instruction receivers while male instruction receivers converged more towards male instruction givers. Overall, Pardo [23] found that males converged more than females; however, findings regarding speaker sex and convergence are mixed. Studies have found that females converge more than males in certain conditions [5, 22], that females are more sensitive to lexical factors when converging [26], and that females adapt their speech more freely according to social factors than males [5].

Several phonetic features have been examined with regard to accommodation such as vowel quality [2, 3, 25], voice onset time [33], fundamental frequency (f0) [4] and pause and gap length [11]. The few studies that have examined voice quality accommodation have focused on individual acoustic voice quality parameters such as jitter and shimmer or open quotient rather than more global occurrences of different voice qualities such as creaky voice or breathy voice [18, 32]. Levitan and Hirschberg [18] investigated convergence at the conversation and the turn level in speakers of AmE by measuring the prosodic dimensions of mean and maximum intensity, mean and maximum pitch, jitter, shimmer and noise-to-harmonics ratio (NHR). At the conversation level, speakers' rates of mean intensity, shimmer and NHR became significantly more similar across the course of the conversation. At the turn level, only maximum pitch and mean pitch showed significant levels of convergence. Schweitzer et al. [32] measured

voice quality parameters that provide some detail of the glottal cycle, such as open quotient and glottal opening, to investigate turn-level convergence in conversations of female German-speaking pairs. Unlike Levitan and Hirschberg [18], Schweitzer et al. [32] found that most voice quality measures exhibited convergence. In addition, speakers' impressions of their interlocutor's competency and likeability influenced convergence on some parameters.

Creaky voice quality has been documented to varying degrees in many varieties of English, including AmE [27, 29, 37], British English [15] and AusE [7, 19, 36]. It has been labelled as a marker of phrase-finality [15, 27, 29] and is associated with mixed social functions such as sounding educated and casual [37] and less trustworthy and less competent than modal voice [1]. To our knowledge, research is yet to investigate any voice quality accommodation in AusE.

2. AIMS

The present study aims to investigate creaky voice accommodation in male and female same-sex conversations between teenage speakers of AusE. Creaky voice can be realised through a wide range of acoustic parameters and different combinations thereof [16]. While previous studies have investigated acoustic voice quality measures independently of each other, this study will provide insight into creaky voice accommodation as a global voice quality. This is particularly relevant given the suggestion that creak prevalence is continuing to increase in varieties of English [8]. We also investigate whether there are sex differences in creaky voice accommodation.

3. METHODS

3.1. Data

A total of 30 recordings ranging in length from 9.4 – 36.6 minutes were included in the analysis. These recordings were collected as part of a study of Multicultural AusE from various schools throughout Sydney. Speakers were drawn from a variety of areas differing in linguistic diversity and socioeconomic status; however, these factors are not examined in this paper. Recordings consist of conversations between same-sex pairs (16f; 14m) of students aged 15 – 16 years facilitated by a research assistant (RA). Pairs of students attended the same school and were familiar with each other prior to the conversation. All speakers involved in the conversation including the RAs were native speakers of AusE.

Conversations were orthographically transcribed using the Speech to Text API from IBM Watson, which provides automatic time-aligned orthographic transcriptions using an Australian model diarised for each speaker. The orthographic transcriptions of the teenagers were corrected, ensuring that the content of the speech was faithful to the original recordings and also that the speech was assigned to the correct speaker. Speech from the RA was not corrected. Recordings were then processed through WebMAUS to generate textgrids containing segmented and aligned phonemes [31].

Creaky voice was identified in the speech of the teenage speakers using a combination of two automated approaches described in White et al. [36] as the optimised Union method. This method consists of combining the outputs of the AntiMode (AM) method, an f0-based approach to detect creak [7, 9], and a Creak Detector (CD) algorithm, which makes decisions on the presence of creak based on acoustic cues including H2-H1 and residual peak prominence [10]. As different phonetic realisations of creak can be characterised by various combinations of acoustic cues and each tool employs different acoustic parameters, it is likely that the AM method would identify instances of creak that the CD algorithm would miss and vice versa. Steps can be taken to further improve creak detection such as optimising the automatic Creak Detection algorithm, which involves manually annotating a subset of data (also shown in [21] and [35]) and by limiting the analysis to sonorant segments only (including vowels, liquids, nasals and glides) [35, 36]. White et al. [36] showed that the optimised Union method outperformed the use of AM or CD alone. This (mostly) automatic process was carried out on the present data, the output of which is a binary creak decision for every 10 ms of input audio.

3.2. Analysis

To analyse whether speakers converged in their use of creaky voice across a conversation, each conversation was divided evenly into thirds. Dividing the conversations into thirds meant that the amount of speech from each speaker in each bin was more even compared to when the conversations were divided into quarters or fifths (mean ratios of speech from each speaker were close to 0 in all bins). This reduces the likelihood that one speaker dominates in any bin.

Creaky voice prevalence was then calculated per speaker per bin. Creak prevalence was determined as the number of 10 ms intervals spent in creak during sonorant segments divided by the total number of 10 ms sonorant segment intervals used by that speaker.

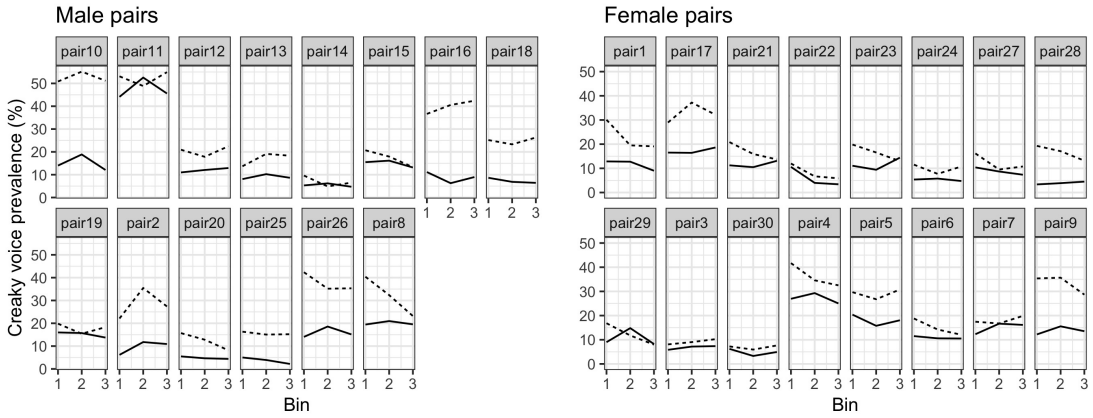

Figure 1: Raw data showing prevalence of creaky voice across thirds of the conversations (i.e., bins). Dotted line = creakier speaker. Male pairs are shown on the left and female pairs are on the right.

For each bin, the difference between each speaker's creak prevalence was calculated. The closer a value was to zero, the more similar the speakers' rates of creaky voice in that third of the conversation.

A linear mixed effects regression (LMER) model was run using the lme4 [6] and lmerTest packages [17] in R [28]. First, a maximal model was built with the difference in creak prevalence as the dependent variable. Independent variables were the third of the conversation (i.e., bin number, treated as a continuous variable henceforth called "bin"), speaker sex (pairs were same-sex), the proportion of the conversation that the RA spoke, the interaction of these three factors, the conversation length and the ratio of speech from each speaker in each bin. Pair was included as a random intercept. All continuous variables (bin, RA proportion, conversation length and speaker ratio) were transformed to z-scores. The data in this study was not collected specifically for the purpose of analysing accommodation. Hence, in each conversation the RA had different levels of engagement with the teenage speakers. Raw means and standard deviations show that, overall, RAs spoke more in the first bin than in the second and third bins and there is considerable variation between conversations (bin 1: mean=21.8% of conversation was RA, sd=9.8; bin 2: mean=16.8%, sd=10.2; bin 3: mean=16.2%, sd=9.4). We proposed that the more an RA talked in a conversation, the greater likelihood that they may have influenced accommodation of the teenagers. We included the proportion of the conversation that the RA spent talking in the model as a way of mitigating their possible influence on creaky voice accommodation. In order to balance model power and Type I error rate, we followed a stepwise elimination process, reducing non-significant interactions followed by non-significant fixed effects, removing terms with highest p-values, and comparing models with ANOVAs to ensure that the removal of each term did not significantly worsen the model [20].

4. RESULTS

Figure 1 shows the raw creaky voice prevalence for each speaker by pair and speaker sex. There appears to be a general trend for speakers to become less creaky over the course of a conversation, as well as evidence of convergence for many of the pairs. This initial impression was tested using LMER modelling.

Our final model included the significant simple fixed effects of bin and gender (with no significant interaction) and the random intercept for pair. The proportion of RA speech, conversation length and the ratio of speech from each speaker in each bin were not significant predictors. The model output is presented in Table 1.

	Est.	Std. error	df	t value	p-value
(Int.)	0.81	0.18	41.9	4.46	<0.001***
bin	-0.10	0.04	60	-2.83	0.006**
sex	0.53	0.25	30	2.15	0.040*

Table 1: Model output showing simple effects of bin and speaker sex on creaky voice convergence. Pair is included as a random intercept. Male is the reference level for sex.

The effect of bin (Figure 2) shows that over the course of a conversation, speakers become more similar to each other in their creaky voice usage. Although the interaction between bin and speaker

sex was not significant, there was a significant simple effect of speaker sex shown in Figure 3. Overall, females were more similar to each other than males.

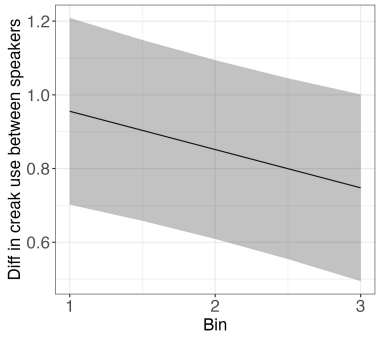

Figure 2: LMER model estimates of the difference in creaky voice prevalence between speakers in a pair across conversations. Values closer to zero indicate convergence.

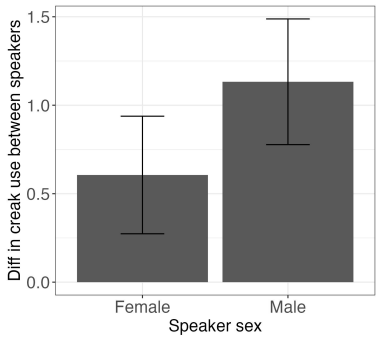

Figure 3: LMER model estimates of the difference in creaky voice prevalence between speakers in a pair for females and males. Values closer to zero indicate more similar creak levels.

5. DISCUSSION

This study explored accommodation of creaky voice in same-sex conversations of teenager speakers of AusE. Results support other studies showing that speakers do tend to phonetically converge towards each other [2, 3, 4, 11, 18, 25, 32, 33]. Unlike other studies of accommodation [5, 22, 23, 26], the present study did not find an interaction of convergence with speaker sex. However, a main effect of sex shows that female pairs were more similar to each other in their levels of creaky voice throughout the conversation than males. In our data, there is greater variance in creaky voice prevalence in males than females (males: mean=19.9%, sd=13.9; females: mean=14.9%, sd=8.45). It is possible that the sex effect is a reflection of this variance; males have more variable creak prevalence, increasing the likelihood that levels of creaky voice use will be more variable between male pairs compared to female pairs. An alternative explanation highlighted by Babel [5] draws on psychology literature indicating that females show more in-group bias than males [30], suggesting that females might be more motivated to accommodate to other females for social reasons.

The present study is limited by the presence of the RA in the conversations. Despite the amount of speech contributed by the RA not significantly influencing creaky voice accommodation, future research should be carried out on data without an RA as it is likely that aspects of the RA's presence other than amount of speech contribute to their influence on accommodation (e.g., RA creak prevalence or attitudes towards the RA). Additionally, it is unknown how participants in this study felt towards their interlocutor. Research has shown that social preferences influence accommodation [2, 3, 5, 25, 32] so future work could investigate whether the same effects exist for creaky voice accommodation and whether this interacts with speaker sex.

6. ACKNOWLEDGEMENTS

This research was supported by a Macquarie University Research Excellence Scholarship to the first author and by Australian Research Council Grant DP190102164 and Australian Research Council Future Fellowship Grant FT180100462 to the fifth author.

7. REFERENCES

[1] R. C. Anderson, C. A. Klofstad, W. J. Mayew, and M. Venkatachalam, "Vocal fry may undermine the success of young women in the labor market," *PLoS ONE*, vol. 9, no. 5, p. e97506, 2014.

[2] M. Babel, "Dialect divergence and convergence in New Zealand English," *Language in Soc.*, vol. 39, no. 4, pp. 437–456, 2010.

[3] ——, "Evidence for phonetic and social selectivity in spontaneous phonetic imitation," *J. Phonetics*, vol. 40, no. 1, pp. 177–189, 2012.

[4] M. Babel and D. Bulatov, "The role of fundamental frequency in phonetic accommodation," *Language and Speech*, vol. 55, no. 2, pp. 231–248, 2012.

[5] M. Babel, G. McGuire, S. Walters, and A. Nicholls, "Novelty and social preference in phonetic accommodation," *Laboratory Phonology*, vol. 5, no. 1, pp. 123–150, 2014.

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

[7] K. Dallaston and G. Docherty, "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.
[8] ——, "The quantitative prevalence of creaky voice (vocal fry) in varieties of English: A systematic review of the literature," *PloS ONE*, vol. 15, no. 3, p. e0229960, 2020.
[9] K. Dorreen, "Fundamental frequency distributions of bilingual speakers in forensic speaker comparison," Master's thesis, University of Canterbury, Christchurch, NZ, 2017.
[10] T. Drugman, J. Kane, and C. Gobl, "Data-driven detection and analysis of the patterns of creaky voice," *Comput. Speech and Language*, vol. 28, no. 5, pp. 1233–1253, 2014.
[11] J. Edlund, M. Heldner, and J. Hirschberg, "Pause and gap length in face-to-face interaction," in *Proc. INTERSPEECH*, 2009, pp. 2779–2782.
[12] I. Gessinger, E. Raveh, I. Steiner, and B. Möbius, "Phonetic accommodation to natural and synthetic voices: Behavior of groups and individuals in speech shadowing," *Speech Communication*, vol. 127, pp. 43–63, 2021.
[13] H. Giles and T. Ogay, "Communication accommodation theory," in *Explaining communication: Contemporary theories and exemplars*, B. Whaley and W. Samter, Eds. Lawrence Erlbaum, 2006, pp. 293–310.
[14] H. Giles, D. M. Taylor, and R. Bourhis, "Towards a theory of interpersonal accommodation through language: Some Canadian data," *Language in Soc.*, vol. 2, no. 2, pp. 177–192, 1973.
[15] C. Henton and A. Bladon, "Creak as a sociophonetic marker," in *Language, speech, and mind: Studies in honour of Victoria A. Fromkin*, L. M. Hyman and C. N. Li, Eds. Routledge, 1988, pp. 3–29.
[16] P. Keating, M. Garellek, and J. Kreiman, "Acoustic properties of different kinds of creaky voice," in *Proc. 18th ICPhS*. the University of Glasgow, 2015, pp. 821.1–5.
[17] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *J. Statistical Softw.*, vol. 82, no. 13, pp. 1–26, 2017.
[18] R. Levitan and J. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," in *Proc. INTERSPEECH*, 2011, pp. 3081–3084.
[19] D. Loakes and A. Gregory, "Voice quality in Australian English," *JASA Express Letters*, vol. 2, no. 8, p. 085201, 2022.
[20] H. Matuschek, R. Kliegl, S. Vasishth, H. Baayen, and D. Bates, "Balancing Type I error and power in linear mixed models," *J. Memory and Language*, vol. 94, pp. 305–315, 2017.
[21] O. Murton, S. Shattuck-Hufnagel, J.-Y. Choi, and D. D. Mehta, "Identifying a creak probability threshold for an irregular pitch period detection algorithm," *J. Acoust. Soc. Am.*, vol. 145, no. 5, pp. EL379–EL385, 2019.
[22] L. L. Namy, L. C. Nygaard, and D. Sauerteig, "Gender differences in vocal accommodation: The role of perception," *J. Language and Social Psychology*, vol. 21, no. 4, pp. 422–432, 2002.
[23] J. S. Pardo, "On phonetic convergence during conversational interaction," *J. Acoust. Soc. Am.*, vol. 119, no. 4, pp. 2382–2393, 2006.
[24] ——, "Measuring phonetic convergence in speech production," *Frontiers in Pyschology*, vol. 4, p. 559, 2013.
[25] J. S. Pardo, R. Gibbons, A. Suppes, and R. M. Krauss, "Phonetic convergence in college roommates," *J. Phonetics*, vol. 40, no. 1, pp. 190–197, 2012.
[26] J. S. Pardo, A. Urmanche, S. Wilman, J. Wiener, N. Mason, K. Francis, and M. Ward, "A comparison of phonetic convergence in conversational interaction and speech shadowing," *J. Phonetics*, vol. 69, pp. 1–11, 2018.
[27] R. J. Podesva, "Gender and the social meaning of non-modal phonation types," *Ann. Meet. Berkeley Linguistics Soc.*, vol. 37, no. 1, pp. 427–448, 2013.
[28] R Core Team, "R: A language and environment for statistical computing," 2021. [Online]. Available: https://www.R-project.org/
[29] L. Redi and S. Shattuck-Hufnagel, "Variation in the realization of glottalization in normal speakers," *J. Phonetics*, vol. 29, no. 4, pp. 407–429, 2001.
[30] L. A. Rudman and S. A. Goodwin, "Gender differences in automatic in-group bias: Why do women like women more than men like men?" *J. Personality and Social Psychology*, vol. 87, no. 4, pp. 494–509, 2004.
[31] F. Schiel, C. Draxler, and J. Harrington, "Phonemic segmentation and labelling using MAUS technique," in *New tools and methods for very-large-scale phonetics research workshop*, Philadelphia, PA, USA, 2011.
[32] A. Schweitzer, W. Wokurek, and M. Pützer, "Convergence of harmonic voice quality parameters in spontaneous dialogues," in *Proc. 19th ICPhS*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds. ASSTA Inc., 2019, pp. 363–367.
[33] K. Shockley, L. Sabadini, and C. A. Fowler, "Imitation in shadowing words," *Perception and Psychophysics*, vol. 66, pp. 422–429, 2004.
[34] A. Walker and K. Campbell-Kibler, "Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task," *Frontiers in Psychology*, vol. 6, 2015.
[35] H. White, J. Penney, A. Gibson, A. Szakay, and F. Cox, "Optimizing an automatic creaky voice detection method for Australian English-speaking females," in *Proc. INTERSPEECH*, 2021, pp. 1384–1388.
[36] ——, "Evaluating automatic creaky voice detection methods," *J. Acoust. Soc. Am.*, vol. 152, no. 3, pp. 1476–1486, 2022.
[37] I. P. Yuasa, "Creaky Voice: A New Feminine Voice Quality for Young Urban-Oriented Upwardly Mobile American Women?" *American Speech*, vol. 85, no. 3, pp. 315–337, 2010.