Comparative Whisper Vowel Space for Singapore English and British English Accents

Hamid R. Sharifzadeh*, Iman T. Ardekani† and Ian V. McLoughlin‡
* Unitec Institute of Technology, Auckland, New Zealand
E-mail: hsharifzadeh@unitec.ac.nz
† Unitec Institute of Technology, Auckland, New Zealand
E-mail: iardekani@unitec.ac.nz
‡ The University of Kent, Kent, United Kingdom
E-mail: i.v.mcloughlin@kent.ac.uk

***Abstract*—Whispered speech, as a relatively common form of communications, has received little research effort in spite of its usefulness in everyday vocal communications. Apart from a few notable studies analysing the main whispered vowels and some quite general estimations of whispered speech characteristics, a classic vowel space determination has been lacking for whispers. Aligning with the previous published work which aimed to redress this shortfall by presenting a vowel formant space for whispers, this paper studies Singapore English (SgE) from this respect. Furthermore, by comparing the shift amounts between normal and whispered vowel formants in two different English accents, British West Midlands (WM) and SgE, the study also considers the question of generalisation of shift amount and direction for two dissimilar accent groupings. It is further suggested that the shift amounts for each vowel are almost consistent for F2 while these vary for F1, showing the role of accent in proposing a general correlation between normal and whispered vowels on first formant.**

This paper presents the results of the formant analysis, in terms of acoustic vowel space mappings, showing differences between normal and whispered speech for SgE, and compares this to results obtained from the analysis of more standard English.

I. INTRODUCTION

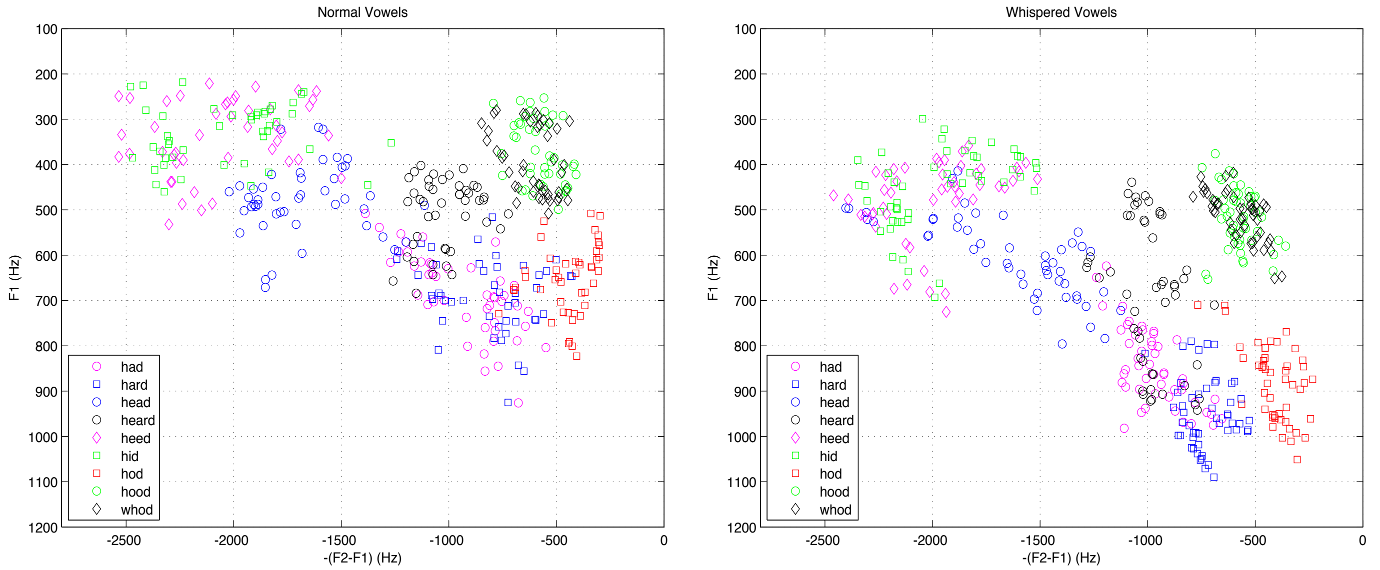
Measuring the acoustic features of phonated vowels provide foundational material for many speech related research fields. Wide research efforts [1], [2], [3], [4], [5], mainly based upon acoustic characteristics of normal vowels, show the importance of these measurements while numerous studies [6], [7], in turn, have considered formant patterns in terms of vowel diagrams and the corresponding characteristics of normal vowels.

While normal phonated vowels have been supported by a long list of literature, on the other hand, whispered speech in terms of vowel measurements has received little research effort. Apart from the few notable studies on whispered vowels [8], [9], [10] which mainly concentrate on a few main vowels /ɪ,ɛ,æ,ɒ,ʊ/ and conclude with general comments on vowel placement such as "higher formants in comparison with normal vowels", accurate acoustic measurements of the precise amount of shift for each vowel is lacking.

Whisper vowel diagrams are useful not only for common speech processing/recognition applications, but also knowing the shift amounts can particularly help those working in the biomedical engineering field of whisper-to-voice reconstruction [11], [12], [13], [14], as well as those in whisper-mode recognition and communications research [15], [16]. Our previously published work [17] tries to present an acoustic vowel space determination (a classic $F2$x$F1$ plane) for this purpose along with the shift amounts between normal and whispers for each vowel through the experiments conducted on British West Midlands (WM) accent speakers; however, the extent of generalisation of vowel shift amounts between two spaces is not yet generalised across other English accents. The current paper analyses Singapore English (SgE) to firstly present the whisper vowel diagrams for SgE, and secondly, extends the discussion to assess the generalisation of the shift amounts between the two vowel spaces for dissimilar English accents.

Many different characteristics of SgE have been considered [18], [19], [20] and the features of normal phonated vowels in SgE have also been described extensively [21], [22], but similar to other English accents, whispered vowels have not received significant research attention so far. It is also useful to make a distinction between two varieties of English spoken commonly in Singapore [23]: a) Standard Singapore English widely used by officials and educated persons, and b) Colloquial Singapore English (known as Singlish) heard in more informal situations particularly among less educated speakers. Similar to other literature in this field, Standard SgE is the focus of this paper, but in whisper mode.

The aim of this paper is to propose a classic formant plane for 9 English vowels (whispered) in SgE, through analysing the formant contours of whispered samples in a /hVd/ structure. The acoustic analysis including details of the recording, speakers, equipment and measurement methods, are described in Section 2, while Section 3 outlines the results separately for men and women. Section 3 also provides a discussion on findings including the comparison of the shift amounts obtained from the current study in SgE with the corresponding results in British WM (presented in our previous paper [17]) for each vowel in whispers and normal speech; finally, Section 4 concludes the paper.



(a) Normal data points (b) Whisper data points

Fig. 1. Values of $F1$ and $F2 - F1$ for 9 vowels from five men and five women recorded 5 times voiced and 5 times whispered. A few redundant data points have been omitted for better clarity. The words heed, hid, head, had, hard, hod, heard, hood, and who'd include vowels /ɪ,i,ɛ,æ,ɑ,ɒ,ə,ʊ,u/ respectively.

II. Acoustic Analysis

A. Subjects and Recordings

Speakers of this study consisted of ten volunteers (5 men and 5 women, aged 21 to 26 years old) born and living in Singapore all of their lives (with Chinese origin). Audio recordings were made of subjects reading lists containing 9 vowels (/ɪ,i,ɛ,æ,ɑ,ɒ,ə,ʊ,u/) in an anechoic chamber, five times with normal phonation and five times in whispered mode (total $2 * 5 * 9 * 10 = 900$ samples).

Subjects read from different randomisations of a list containing the words 'heed', 'hid', 'head', 'had', 'hard', 'hod', 'heard', 'hood', and 'who'd'. If the subjects stumbled over the samples, re-recording of the samples was allowed. Speakers could repeat the sample until an accurate pronunciation was achieved. The recorded speech was sampled at a rate of 22050 Hz with 16 bit resolution. Through a special prompt-based recording software, speech was read, and recorded directly onto a laptop computer in a sound proof booth. The microphones used were an Emkay head mounted microphone and a Telex desk microphone (for near and far field recording, respectively). An Edirol UA-5 USB sound card interface bypassed the sound card of the laptop, removing any variation in the recordings due to different hardware. An Emkay VR3294 Battery Box provided a stable bias voltage for the microphones.

B. Formant Contours

The automatic approach to formant analysis based on forced alignment using single emitting state phone-level HMMs to detect the vowel centres and ESPS for formant frequency measurement (such as the one described in [24]) was implemented but due to many outliers resulting from noisy nature of whispered speech [25], the more time consuming manual methods were preferred. For this purpose, different methods were combined for accurate extraction of the first two formant frequencies for each sample in the normal and whisper modes. After clipping the steady state of vowel duration by removing the /h/ and /d/ carriers, the analysis methods outline as follows: a) peak-findings through direct observation of 12-pole, 128-point linear predictive coding (LPC) spectra on every 6 ms over 12 ms Hamming windowed segments, b) looking at the results of the robust formant tracker implemented in [26], and c) observation of the gray scale spectrograms (both wide and narrow band). The decisions about formant frequencies were determined by the outcome of these three methods, as

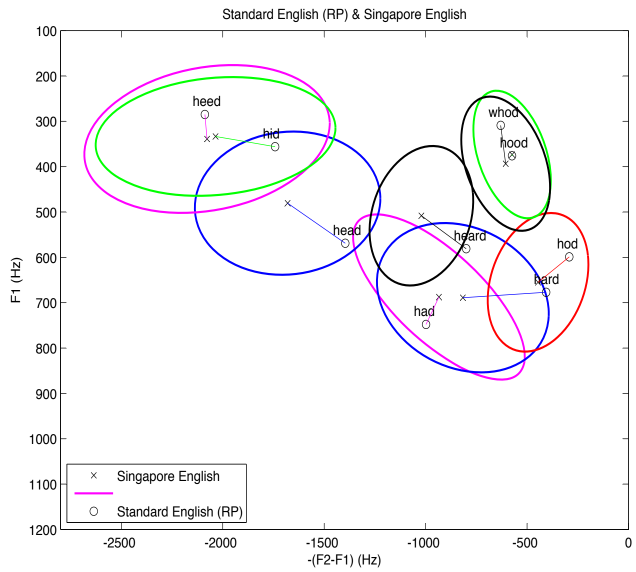
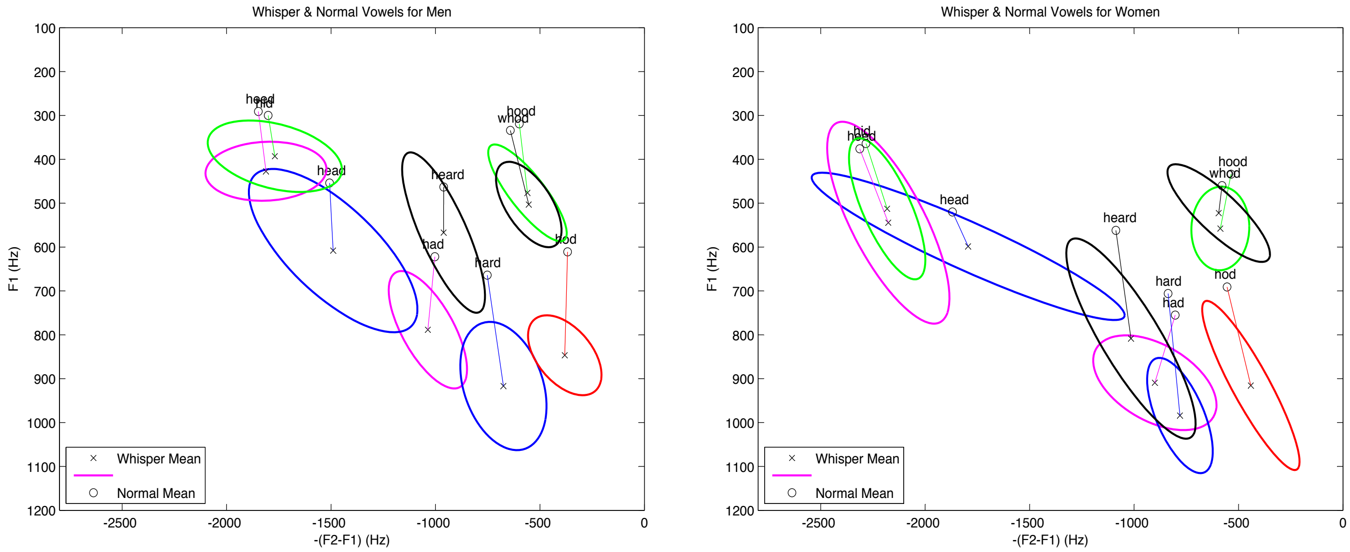


Fig. 2. Average values of $F1$ and $F2 - F1$ for standard English and Singapore English. Ellipses fit to each vowel category in Singapore English. The average shift amounts also have been joined by a line.



(a) Whisper vowels versus normal vowels for men

(b) Whisper vowels versus normal vowels for women

Fig. 3. Average values of $F1$ and $F2-F1$ for normal and whispered vowels in: a) men, b) women. Ellipses fit to each vowel category in Singapore English for a) men, b) women. The average shift amounts also have been joined by a line.

well as by comparing the results to select the most accurate representation. Thus, all reported results have been verified manually one-by-one.

Figure 1 shows the individual data points of the measured first and second formants through the combined approach for a) normal samples and b) whispered data while a few redundant points have been omitted for clarity.

III. Results and Discussion

Acoustic measurements on formant values of the /hVd/ samples for both normal and whisper modes are presented separately in this section for men and women. Since the data were collected in Singapore, the amount of vowel variation in SgE, compared with the average formant frequencies in Standard English (Received Pronunciation, RP) is also provided for referencing purposes, in addition to normal and whisper variations which are the primary aim of the paper. RP formant values were obtained from Wells' work [27]. Figure 2 shows the average frequencies of normal phonation for $F1$ and $F2-F1$ along with ellipses showing the standard deviation within each vowel category in SgE. The variations between SgE and average formant frequencies in RP accent also have been shown. In figure 3, the variations between normal and whispered vowels are separately illustrated for male and female speakers. The corresponding acoustic vowel diagrams on a $F1$x$F2$ space are presented in figure 4 based on average formant frequency. Again, this shows normal and whisper samples for a)men, and b)women. The discussion and the analysis of these data is presented in III-A.

Tables I and II provide the precise percentage of shift amounts of first and second formants for each vowel in SgE averaged within normal and whisper phonation, again separately for men and women.

TABLE I

Average formant values in normal and whispered vowels for men (N: Normal, W: Whisper, S.A: Shift amount in %)

		/ɪ/	/i/	/ɛ/	/æ/	/ɑ/	/ɒ/	/ə/	/ʊ/	/u/
F1	N	296	300	447	621	665	616	468	316	334
	W	427	392	608	788	916	846	567	476	503
	S.A	0.44	0.31	0.36	0.27	0.37	0.38	0.21	0.50	0.50
F2	N	2145	2120	1959	1655	1445	990	1436	925	961
	W	2238	2162	2098	1825	1591	1228	1528	1036	1056
	S.A	0.04	0.02	0.07	0.10	0.10	0.24	0.06	0.12	0.10

TABLE II

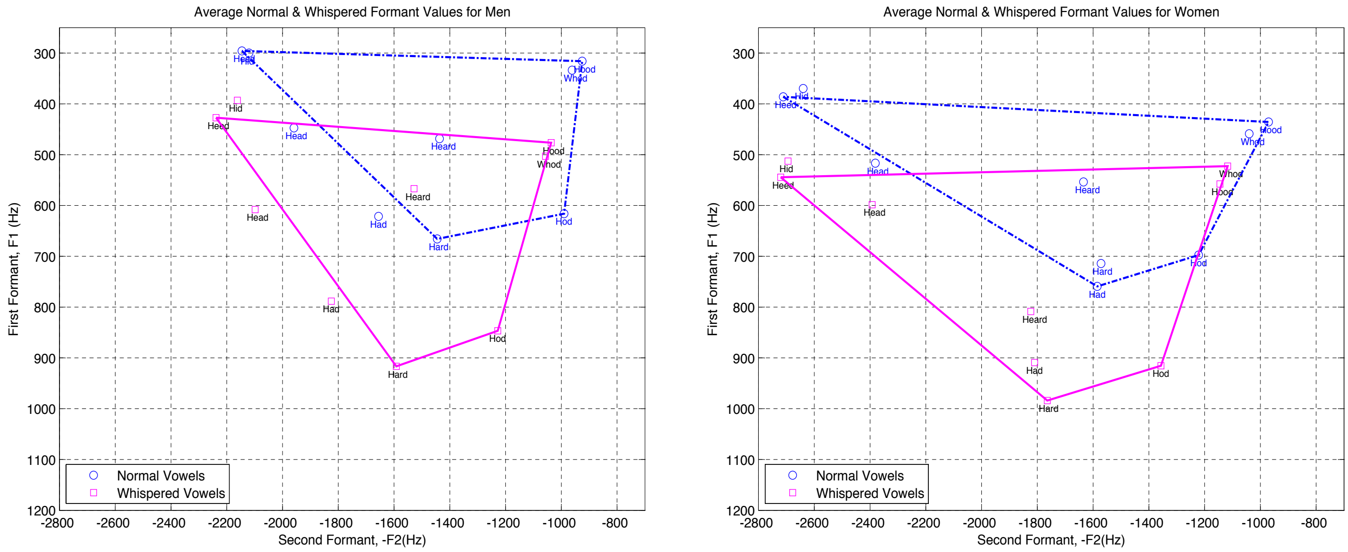
Average formant values in normal and whispered vowels for women (N: Normal, W: Whisper, S.A: Shift amount in %)

		/ɪ/	/i/	/ɛ/	/æ/	/ɑ/	/ɒ/	/ə/	/ʊ/	/u/
F1	N	386	370	516	759	714	697	554	435	458
	W	544	512	598	909	984	915	808	558	522
	S.A	0.41	0.39	0.16	0.20	0.38	0.31	0.46	0.28	0.14
F2	N	2711	2639	2381	1585	1571	1221	1634	971	1040
	W	2721	2694	2393	1810	1764	1356	1823	1144	1117
	S.A	0.003	0.02	0.004	0.14	0.12	0.11	0.11	0.18	0.07

A. Discussion

As the main vowel characteristics of SgE, Deterding [22] points out the lack of distinction between the long and short vowel pairs; so this neutralisation of length distinctions often causes contrasts between {/ɪ/ and /i/}, {/ʊ/ and /u/}, and {/ɑ/ and /ɒ/} may not be found. These features can also be observed in figure 1(a) and figure 2 as these vowel pairs respectively in {'heed' and 'hid'}, {'hood' and 'whod'}, and {'hod' and 'hard'} are highly overlapping. Furthermore, the absence of distinction between /ə/ and /æ/ mentioned by Bao [21] as one of the SgE features can also be seen in partly overlapping 'had' and 'heard' in figure 2.

Comparing with our previous study [17], this merging effect



(a) Vowel diagrams for men

(b) Vowel diagrams for women

Fig. 4. Acoustic vowel diagrams showing average formant frequencies for normal and whispered vowels from male and female speakers.

on long and short vowel pairs causes the vowel such as /ʌ/ in ‘hudd’ which highly overlaps with ‘hard’ and ‘hod’, to be affected; thus, to keep the clarity of the graphs and avoid repetitive values, the data corresponding to /ʌ/ and /ɔ/ were excluded while the remaining 9 vowels were considered. Furthermore, third formant has not been considered due to not showing any significant shift as discussed in the previous work[17].

Shifting and more convergence of adjacent vowels is evident in the whispered samples both for men and women. As shown in figure 3, vowel groups such as {/u/ and /ʊ/} or {/ɪ/ and /i/} retain the similarity in terms of formant characteristics (as the clues of SgE) while /ɛ/ (in ‘head’) is about to merge with /ɪ/ and /i/. Furthermore, {/ɑ/ and /ə/ and /æ/} show a high degree of overlapping in figure 3(b) meaning the decrease of distinction between ‘hard’, ‘heard’, and ‘had’ in samples of Singaporean women.

It can be seen from the diverse amount of shifts in figure 3 that each vowel has its own variation when converted to whispered speech and this amount also varies in terms of formant number. Tables I and II summarise these variations for the first two formants in whisper and normal speech for men and women, respectively.

As shown in tables I and II, all first and second formants are shifted upwards. The shift amounts range from 21% in /ə/ to 50% in /u/ for men and from 14% in /u/ to 46% in /ə/ for women within the first formants and from 2% in /i/ to 24% in /ɒ/ for men and from 0.3% in /ɪ/ to 18% in /ʊ/ for women within the second formants. Furthermore, significant shifts occur in the first formants with average of 37% (σ : 10%) and 30% (σ : 11%) while these numbers are 9% (σ : 6%) and 8% (σ : 6%) for the second formants for men and women, respectively. From the tables, it can be observed that the extreme closed vowels either front or back (as in /ɪ/ and

TABLE III

Average shift amounts of each vowel between normal and whispers in SgE and British WM

		/ɪ/	/i/	/ɛ/	/æ/	/ɑ/	/ɒ/	/ə/	/ʊ/	/u/
F1 Shift	WM	0.38	0.23	0.30	0.25	0.25	0.28	0.49	0.31	0.57
	SgE	0.43	0.35	0.25	0.23	0.38	0.34	0.35	0.38	0.30
F2 Shift	WM	0.02	0.04	0.02	0.14	0.15	0.20	0.08	0.08	0.07
	SgE	0.02	0.02	0.03	0.12	0.11	0.17	0.09	0.15	0.08

/ʊ) show greater amount of shift than central open-mid and close-mid vowels.

Moving from spoken vowels to whispers, the size of quadrilaterals in figure 4 also show different changes in terms of area. For both men and women, the area expands in height while the width remains almost the same. In fact, the significant changes appear in diagrams particularly on the height of the quadrilaterals corresponding to whispers.

The average change in size of quadrilateral for men is 31% in height and less than 2% in width at the extremes while the height shows increase but width decreases by this amount when moving from voiced to whispered mode. These amounts are 8% decrease in width and 17% increase in height for quadrilaterals of women’s vowels. As mentioned, the significant changes occur in height of both diagrams by increasing 31% and 17% in whispered speech.

To consider whether the amount of shifting between whispers and normal speech for each vowel are consistent, table III compares the results of this study in terms of average shift amounts (combined men and women) with the one recently conducted in UK [17]. This helps explore how much generalisation might be taken into account when shifting between whispers and normal phonated vowels across accents, but would of course need to extend to other accents before it can be considered definitive. However some interesting trends can be seen: the F1 of each vowel has quite different degrees of shift ranging from 23% to 43% in SgE and 23% to 57% in British WM. F2 shows more consistent figures in which seven out of nine vowels have almost the same amount of shifting (with 2% margin). We can summarise with the comment that early results indicate some consistency in F2 shifts irrespective of accent, but that correlation of F1 shifts between whisper and normal speech for different accents is likely to be low.

IV. Conclusion

A vowel formant space for whispered SgE speech has been established through experimentation with Singaporean subjects. By comparing whispered vowels with the corresponding phonated samples separately for men and women, amounts of shift for each vowel and formant have been presented, and the distribution of formant values for normal and whispered SgE samples also illustrated. Acoustic vowel diagrams were also presented showing increase in height of quadrilaterals for both men and women while the width remained almost the same.

In terms of generalising these shift amounts between whisper and normal speech across different accent groups, F1 and F2 results from the current study were compared to previously published British WM results. Although further studies on different accent groups might be required before definitive conclusions can be made on generalisation, the comparative analysis in this paper suggests that the shift amounts for F1 depend largely upon accent whereas F2 shifts show more consistency.

References

[1] J. D. Miller, "Auditory-perceptual interpretation of the vowel," *Journal of the Acoustical Society of America*, vol. 85, pp. 2114–2134, 1989.

[2] T. Nawka, L. C. Anders, M. Cebulla, and D. Zurakowski, "The speaker's formant in male voices," *Journal of Voice*, vol. 11, pp. 422 – 428, 1997.

[3] I. V. Bele, "The speaker's formant," *Journal of Voice*, vol. 20, pp. 555 – 578, 2006.

[4] T. M. Nearey, "Static, dynamic, and relational properties in vowel perception," *Journal of the Acoustical Society of America*, vol. 85, pp. 2088–2113, 1989.

[5] M. P. Gelfer and V. A. Mikos, "The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels," *Journal of Voice*, vol. 19, pp. 544 – 554, 2005.

[6] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *Journal of the Acoustical Society of America*, vol. 24, pp. 175–184, 1952.

[7] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of american english vowels," *Journal of the Acoustical Society of America*, vol. 97, pp. 3099–3111, 1995.

[8] K. J. Kallail and F. W. Emanuel, "Formant-frequency difference between isolated whispered and phonated vowel samples produced by adult female subject," *Journal of Speech and Hearing Research*, vol. 27, pp. 245–251, 1984.

[9] S. T. Jovicic, "Formant feature differences between whispered and voiced sustained vowels," *Acta Acustica united with Acustica*, vol. 84, pp. 739–743, 1998.

[10] F. W. Smith, "A formant study of whispered vowels," Ph.D. dissertation, University of Oklahoma, 1973.

[11] H. R. Sharifzadeh, I. V. McLoughlin, and F. Ahmadi, "Reconstruction of normal sounding speech for laryngectomy patients through a modified celp codec," *IEEE Transactions on Biomedical Engineering*, vol. 57, pp. 2448–2458, 2010.

[12] R. W. Morris and M. A. Clements, "Reconstruction of speech from whispers," *Medical Engineering and & Physics*, vol. 24, pp. 515 – 520, 2002.

[13] J. Li, I. V. McLoughlin, L. Dai, and Z. Ling, "Whisper-to-speech conversion using restricted boltzmann machine arrays," *Electronics Letters*, vol. 50, no. 24, pp. 1781 – 1782, 2014.

[14] I. V. McLoughlin, H. R. Sharifzadeh, S. Tan, J. Li, and Y. Song, "Reconstruction of phonated speech from whispers using formant-derived plausible pitch modulation," *ACM Transactions on Accessible Computing*, vol. 6, no. 4, pp. 12:1–12:21, 2015.

[15] B. P. Lim, "Computational differences between whispered and non-whispered speech," Ph.D. dissertation, University of Illinois, 2010.

[16] L. Xuan, D. Wee, T. Hilary, B. P. Lim, N. Yih, and M. Bin, "A whispered mandarin corpus for speech technology applications," in *INTERSPEECH*, 2014.

[17] H. R. Sharifzadeh, I. V. McLoughlin, and M. J. Russell, "A comprehensive vowel space for whispered speech," *Journal of Voice*, vol. 26, no. 2, pp. e49 – e56, 2012.

[18] M. J. Tay, "The phonology of educated singapore english," *English World-Wide*, vol. 3, pp. 135 – 145, 1982.

[19] D. Deterding, "The intonation of singapore english," *Journal of the International Phonetic Association*, vol. 24, pp. 61–72, 1994.

[20] ——, "The measurement of rhythm: a comparison of singapore and british english," *Journal of Phonetics*, vol. 29, pp. 217–230, 2001.

[21] Z. Bao, *English in New Cultural Contexts: Reflections from Singapore*. Oxford University Press, 1998, ch. The sounds of Singapore English, pp. 152–174.

[22] D. Deterding, "An instrumental study of the monophthong vowels of singapore english," *English World-Wide*, vol. 24, pp. 1–16, 2003.

[23] L. Wee, *A Handbook of Varieties of English*. Mouton de Gruyter, 2004, ch. Singapore English: phonology, pp. 1017–1033.

[24] S. D'Arcy, "The effect of age and accent on automatic speech recognition performance," Ph.D. dissertation, University of Birmingham, 2007.

[25] H. R. Sharifzadeh, "Reconstruction of natural spunding speech from whispers," Ph.D. dissertation, Nanyang Technological University, Singapore, 2012.

[26] K. Mustafa and I. C. Bruce, "Robust formant tracking for continuous speech with speaker variability," *IEEE Transactions on Speech and Audio Processing*, vol. 14, pp. 435– 444, 2006.

[27] J. C. Wells, *Accents of English, Volume 2: The British Isles*. Cambridge: Cambridge University Press, 1982, ch. England, RP revisited, pp. 279 – 300.