

# A Comprehensive Vowel Space for Whispered Speech

H. R. Sharifzadeh<sup>a,\*</sup>, I. V. McLoughlin<sup>a</sup>, M. J. Russell<sup>b</sup>

<sup>a</sup>*School of Computer Engineering, Nanyang Technological University, Singapore 639798*

<sup>b</sup>*School of Electronic, Electrical & Computer Engineering, The University of Birmingham, Edgbaston Birmingham B15 2TT, United Kingdom*

---

## Abstract

Whispered speech is a relatively common form of communications, used primarily to selectively exclude or include potential listeners from hearing a spoken message. Despite the everyday nature of whispering, and its undoubted usefulness in vocal communications, whispers have received relatively little research effort to date, apart from some studies analysing the main whispered vowels and some quite general estimations of whispered speech characteristics. In particular, a classic vowel space determination has been lacking for whispers. For voiced speech, this type of information has played an important role in the development and testing of recognition and processing theories over the past few decades, and can be expected to be equally useful for whisper-mode communications and recognition systems.

This paper aims to redress the shortfall by presenting a vowel formant space for whispered speech, and comparing the results with corresponding phonated samples. In addition, since the study was conducted using speakers from Birmingham, the analysis extends to discuss the effect of the common British West Midlands (WM) accent in comparison with Standard English (RP). Thus, the paper presents the analysis of formant data showing differences between normal and whispered speech while also considering an accentual effect on whispered speech.

### *Keywords:*

Whispered speech, vowel space, British West Midlands accent, formant analysis, acoustic characteristics

---

\*Corresponding author

*Email addresses:* hami0003@ntu.edu.sg (H. R. Sharifzadeh),  
mcloughlin@ntu.edu.sg (I. V. McLoughlin), m.j.russell@bham.ac.uk (M. J. Russell)

---

## 1. Introduction

Acoustic measurements of phonated vowels and diphthongs form foundational material for the speech processing and recognition fields. Wide research efforts<sup>1,2,3,4,5,6,7</sup>, mainly based upon acoustic characteristics of normal vowels, show the importance of these measurements while numerous studies<sup>8,9,10</sup>, in turn, have considered formant patterns in terms of vowel diagrams and the corresponding characteristics of normal vowels.

Despite the strong literature supporting normal vowels, little research effort has been spent on whispered speech relating to vowel space. Apart from the studies describing the vocal mechanism of whispers' production mostly on a glottal level<sup>11,12,13,14</sup>, as well as a recent study on whispered consonants<sup>15</sup>, the few notable studies on whispered vowels<sup>16,17,18</sup> are mainly concentrated on a few main vowels /I,ε,æ,Λ,υ/ and conclude with general comments on vowel placement such as "higher formants in comparison with normal vowels", but accurate acoustic measurements of the precise amount of shift for each vowel/diphthong is lacking. Thus, whispered speech still lacks an acoustic vowel space determination (a classic  $F2 \times F1$  plane) for researchers to refer to. Whisper vowel diagrams are useful not only for common speech processing/recognition applications, but also can help those working in the biomedical engineering field of whisper-to-voice reconstruction, particularly rehabilitation of post-laryngectomised patients through restoring their normal sounding speech<sup>19,20</sup>.

The term 'whispered speech' itself encompasses two distinct classes of speech which we shall refer to as soft whispers and stage whispers<sup>21</sup>. Soft whispers (also known as quiet whispers) are produced by normally speaking people to deliberately reduce perceptibility, such as whispering into someone's ear in a theatre, and are usually spoken in a relaxed manner with little effort<sup>11</sup>. Stage whispers, however, are used if the listener is some distance away from the speaker<sup>21</sup>, and are actually a whispery voice which includes partial phonation<sup>22</sup>. The more common soft whispers, produced without vocal folds vibration, are the focus of this study.

As mentioned, the lack of vocal folds vibration is the main physical feature and the most significant acoustic characteristic of whispered speech. It implies the absence of fundamental pitch, and the harmonic relationships that are usually derived from this<sup>23</sup>. In a source filter model<sup>24</sup>, exhalation

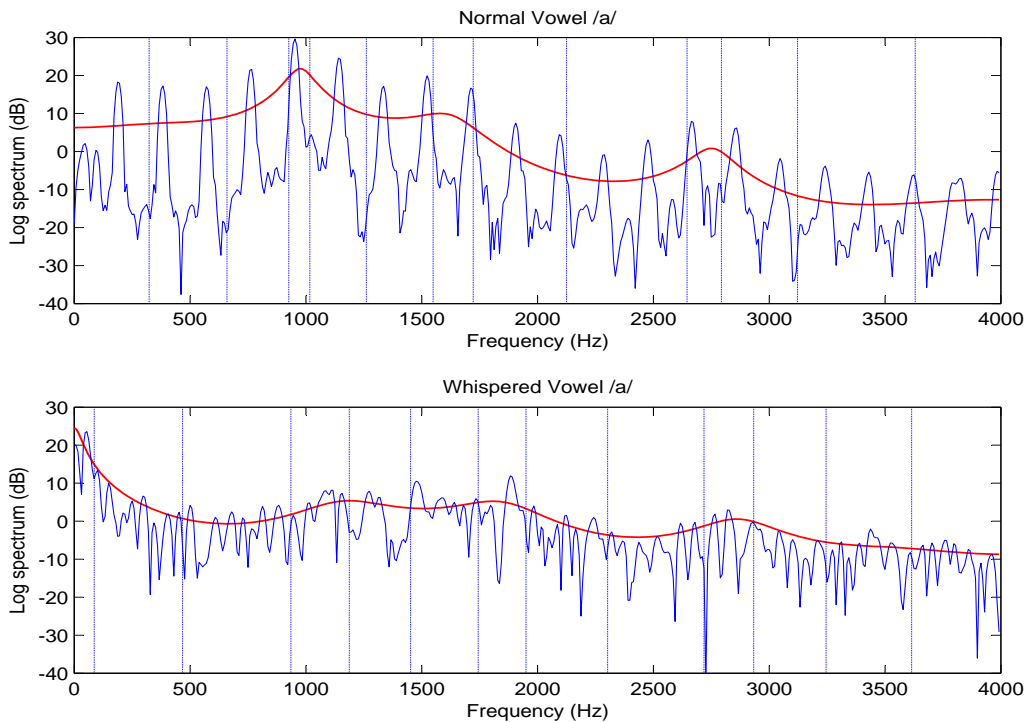


Figure 1: Comparison of the spectra for vowel /a/ in normally phonated speech (top) with whispered speech (bottom) for a single listener during a single sitting. The smoothed spectrum overlay shows formant peaks existing in similar locations, but less pronounced for whispered speech. Furthermore, LSPs overlaid as vertical lines, typically exhibit wider spacing for the whispered speech.

forms the source of excitation in whispered speech, with the shape of the pharynx adjusted so that vocal cords do not vibrate<sup>25</sup>. Exhaled air passes directly through the restricted but open larynx, causing turbulent aperiodic airflow which forms the sound source for whispers: a rich ‘hushing’ sound<sup>13</sup>.

Regarding spectral features, the spectrum of whispered speech certainly exhibits small spectral peaks at “approximately” the same frequencies as those for normally phonated speech sounds<sup>26</sup>. Such ‘formant-like’ features have a much flatter power-frequency distribution<sup>23</sup> than normal, and generally tend to be higher in frequency than the corresponding voiced speech<sup>27</sup>, particularly the first formant which shows the greatest difference between two kinds of speech. Furthermore, unlike phonated vowels where the amplitude of the higher frequency formants is usually lower than the lower frequency formants, the second formants of whispered vowels are typically as intense as

the first formants. Figure 1 shows this feature by contrasting the spectra of the vowel /a/ spoken in a whisper and in a normal voice. These differences, mainly in first formant frequency and amplitude, are thought to be due to the alteration in the shape of the posterior areas of the vocal tract including the vocal cords, which are held rigid so as to not vibrate<sup>28</sup>.

The aim of this paper is to propose a classic formant plane for all 11 English vowels (whispered), 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 as well as a brief discussion on findings including the consideration of possible accentual effects in British West Midlands accent for whispers and normal speech; finally, Section 4 concludes the paper.

## 2. Acoustic Analysis

### 2.1. Subjects and Recordings

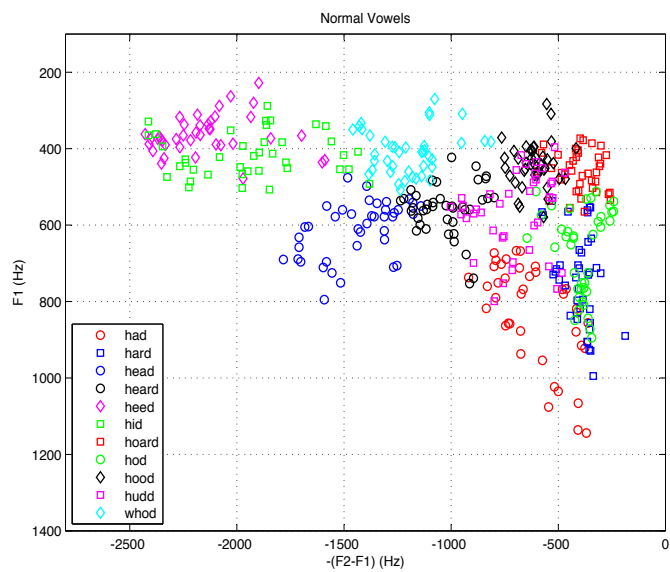
Speakers of this study consisted of ten middle-aged volunteers (5 men and 5 women, 35 to 45 years old) born and living in Birmingham all of their lives. An additional criterion of one's parents having lived in the area most of their lives was also used for the selection of volunteers.

Audio recordings were made of subjects reading lists containing 11 vowels (/ɪ, i, ε, æ, ɑ, ʌ, ɒ, ə, ɔ, ʊ, u/) in a sound proof audio booth, five times with normal phonation and five times in whispered mode (total 10 times).

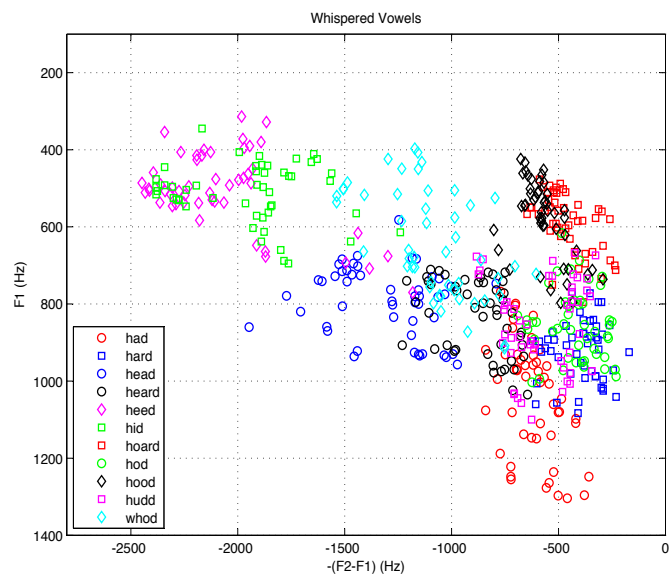
Subjects read from five different randomisations of a list containing the words 'heed', 'hid', 'head', 'had', 'hard', 'hudd', 'hod', 'heard', 'hoard', 'hood', and 'who'd'. Since the objective is to find out how *ordinary* people from Birmingham speak the vowels in the specific words, /hVd/ carrier gives actual/meaningful words in most cases, except of 'hudd' (which this also occurs as a family name) while this keeps the current study aligning with previous acoustic studies on vowels<sup>8,9,10</sup> through following the same pattern.

Furthermore, having a plosive phoneme such as 'd' at the final syllable, makes it simple to detect vowels in between from carriers within both automatic or manual methods of extraction process; particularly, due to showing a peak of energy in both whisper and normal modes after a very short silence, 'd' can be a good choice for the final syllable.

Recordings were made of 5 readings of the list in each whisper and normal modes (total  $5 * 11 * 2 * 10 = 1100$  samples). The details of the interface



(a) Normal data points



(b) Whisper data points

Figure 2: Values of  $F1$  and  $F2 - F1$  for 11 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, hudd, hod, heard, hoard, hood, and who'd include vowels /ɪ, i, ε, æ, ʌ, ɒ, ə, ɔ, u, u/ respectively.

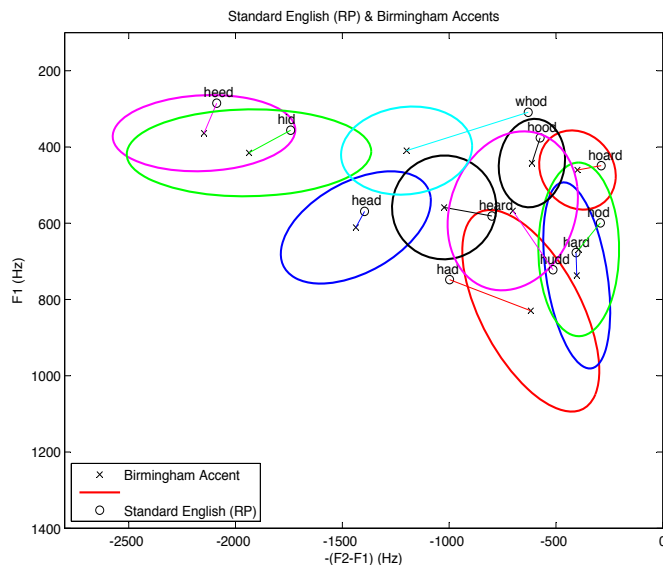


Figure 3: Average values of  $F1$  and  $F2 - F1$  for standard English and Birmingham accents. Ellipses fit to each vowel category in Birmingham accent. The average shift amounts also have been joined by a line.

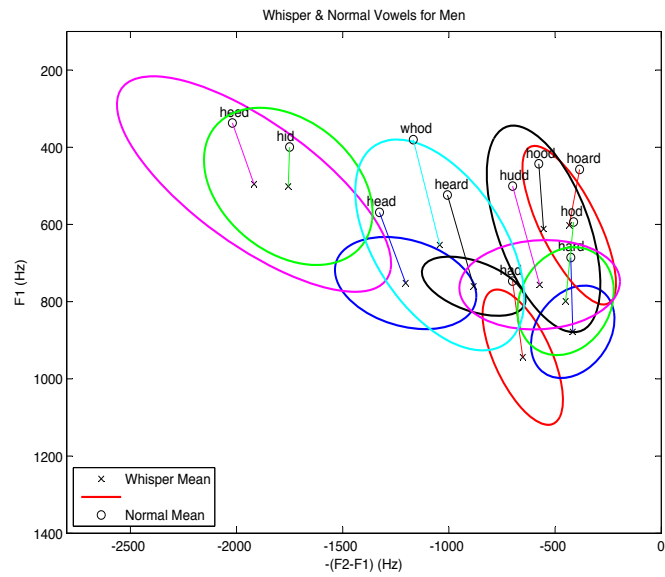
are described in subsection 2.2. If the subjects stumbled over the samples, re-recording of the samples was allowed. Speakers could repeat the sample until an accurate pronunciation was achieved.

## 2.2. Equipments and Interface

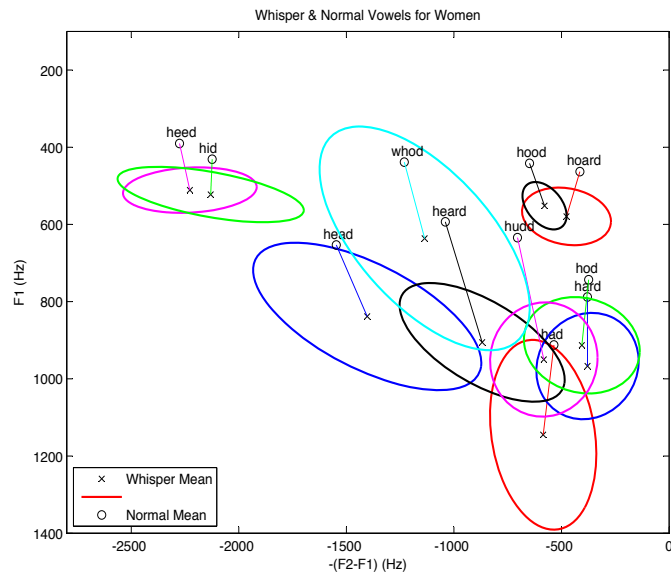
Speech was read, and recorded directly onto a laptop computer in a sound proof booth. The microphones used were an Emkay head mounted microphone and a Telex desk microphone (for near and far field recording, respectively). An Edirol UA-5 USB sound card interface bypassed the sound card of the laptop, removing any variation in the recordings due to different hardware. An Emkay VR3294 Battery Box provided a stable bias voltage for the microphones.

A special prompt-based recording software, developed by the University of Birmingham, was used as the recording application. Any set of prompts specified in a separate xml-formatted file with different login options, can be loaded into the prompt-recorder at run time; so the randomised lists of vowels in /hVd/ carriers as mentioned in subsection (2.1) were customised using this application.

The recorded speech is sampled at a rate of 22050 Hz with 16 bit resolution. The application included level meters for both microphone inputs. At



(a) Whisper vowels versus normal vowels for men



(b) Whisper vowels versus normal vowels for women

Figure 4: Average values of  $F_1$  and  $F_2 - F_1$  for normal and whispered vowels in: a)men, b)women. Ellipses fit to each vowel category in Birmingham accent for a)men, b)women. The average shift amounts also have been joined by a line.

the beginning of each recording session, the subject spoke for a few seconds





by the person recording the speech, so the subjects need only concentrate on reading the text in front of them. Subjects were seated in front of the laptop and the headset microphone placed on their head with the microphone angled about 5 cm away from the right corner of their mouth. The desk-mounted microphone was placed to the left of the laptop.

### *2.3. Formant Contours*

Segmentation/extraction of normal speech at a phonemic or sub phonemic level has been generally an attractive research field in speech recognition. The different approaches and solutions proposed by researchers can be mainly categorised into two major recognition classes: manual and automatic<sup>29,30,31,32,33</sup>. Based upon specific acoustic cues, these approaches try to identify/classify different groups of phonemes. The process of segmentation in whispered speech is more complicated<sup>15</sup>. For the purpose of this research, a combined approach based on manual observation has been taken as described in the following.

Different methods were combined for accurate extraction of the first three formant frequencies for each sample in the normal and whisper modes. After manually clipping the steady state of vowel duration by removing the /h/ and /d/ carriers, the analysis methods, which are mainly based upon manual observation of the results, outline as follows:

- 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 (256-point)
- looking at the results of the robust formant tracker implemented in<sup>34</sup>
- observation of the gray scale spectrograms (both wide and narrow band)

An automatic approach to formant analysis based on forced alignment using single emitting state phone-level HMMs to detect the vowel centres and ESFS for formant frequency measurement (described in<sup>29</sup>) was implemented. However, due to many outliers resulting from the use of whispered speech, the more time consuming manual methods were preferred: thus all results have been verified manually one-by-one.

Finally, the decisions about formant frequencies were determined by the outcome of the methods mentioned, as well as by comparing the results to

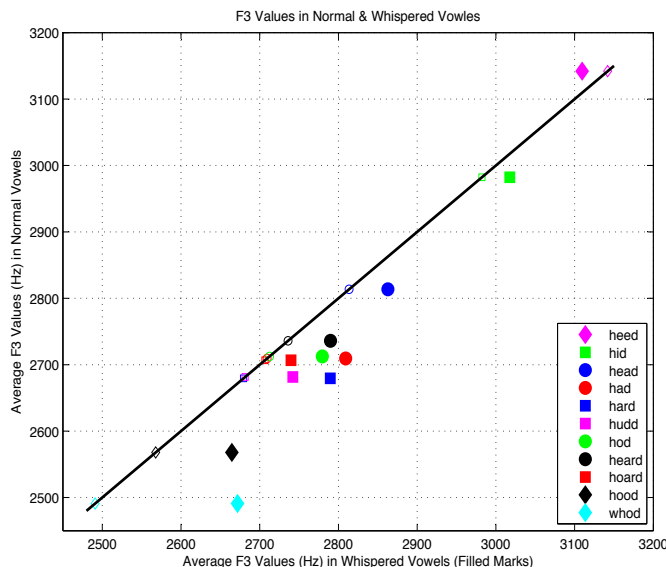


Figure 6: Average values of  $F3$  for normal and whispered vowels. The solid line indicates the  $F3$  values for normal vowels while filled marks shows the corresponding average  $F3$  values from whispered samples across the horizontal axis. heed, hid, head, had, hard, hudd, hod, heard, hoard, hood, and who'd are /i,i,ε,æ,ɑ,ʌ,ɒ,ə,ɔ,ʊ,u/ respectively.

select the most accurate representation (a general knowledge of acoustic phonetics, such as the close proximity of  $F2$  and  $F3$  in /i/ and /ε/, also played a role in this process). Figure 2 shows the individual data points of the measured first and second formants through this approach for a) normal samples and b) whispered data while a few redundant points have been omitted to be clearer.

### 3. Results and Discussion

#### 3.1. Results

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 the West Midlands, from speakers with particular accent characteristics<sup>35</sup>, the amount of vowel variation in Birmingham accent, 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<sup>35</sup>.

Figure 3 shows the average frequencies for  $F1$  and  $F2 - F1$  along with ellipses showing the standard deviation within each vowel category. The variations between Birmingham accent and average formant frequencies in RP accent have been shown. As can be seen, /ɪ/ and /i/ as well as /ɑ/ and /ɒ/ are showing a high degree of overlap in this accent while shift of vowels have been identified. The detailed discussion and the verification of these data in comparison with RP is presented in 3.2.1.

In figure 4, the variations between normal and whispered vowels are separately illustrated for male and female speakers, while the corresponding acoustic vowel diagrams on a  $F1 \times F2$  space are presented in figure 5 based on average formant frequency. Again, this shows normal and whisper samples for a)men, and b)women.

Figure 6 compares average values of third formant in normal samples with whispered ones. Overall, the  $F3$  values from both normal and whisper samples are quite similar, with all vowels except /u/, averaging less than 100 Hz shifts from normal to whispered data. These slight shifts on  $F3$  values average 2.0% ( $\sigma : 2.4\%$ ) for men and 2.2% ( $\sigma : 1.6\%$ ) for women. Thus, third formants have not been much affected by whispering compared to first and second formants.

## 3.2. Discussion

### 3.2.1. RP and Birmingham Accents

Hughes and Trudghill<sup>36</sup> lists the lack of distinction made between /ʌ/ and /ʊ/, and /i/ becoming very close (shifted up) toward the vowel space occupied by the /ɪ/ as the main vowel characteristics of the Birmingham accent. Both these characteristics can be observed from figure 2(a) and figure 3. The  $F1$  values of /ʌ/ ('hudd') decrease from 722 Hz (RP) to 567 Hz while the  $F2 - F1$  formant increase to 702 Hz from 514 Hz (RP). These values are now more similar to the  $F1$  and  $F2 - F1$  values for /ʊ/ ('hood') in Birmingham normal samples, 442 Hz and 612 Hz respectively compared to the RP  $F1$  and  $F2 - F1$  values of 376 Hz and 574 Hz respectively.

The distribution for /ʊ/ appears to be a subset of the distribution for /ʌ/. This might be because although the pronunciation of /ʊ/ is relatively unambiguous for a Birmingham speaker, the pronunciation of /ʌ/ is perhaps more ambiguous - the Birmingham speakers may be aware that there is a 'proper' pronunciation of /ʌ/ and they may be trying to approximate it.

According to Wells<sup>35</sup>, the West Midlands pronunciation of the vowels in 'bath' and 'trap' are identical (following the short vowel, /æ/ from 'trap').

The closest (not the same) we have to this is ‘hard’(/ɑ/) versus ‘had’(/æ/). The overlap between these two distributions increases in normal speech, as in figure 2(a) and figure 3. The average  $F1$  value of /æ/ increases from 748 Hz (RP) to 829 Hz while the average  $F2 - F1$  falls significantly down from 998 Hz (RP) to 618 Hz which are now closer to the corresponding amounts of  $F1$  and  $F2 - F1$  in /ɑ/ with 736 Hz and 403 Hz in Birmingham normal samples.

The overlap between the  $F1$  and  $F2 - F1$  values for the vowels in ‘hod’ (/ɒ/) and ‘hard’ (/ɑ/) is very striking in figure 3. The average value for /ɒ/ moves to the RP value for /ɑ/, while the value for Birmingham shows a slightly higher  $F1$  than the RP version.

A closer version of ‘heard’ is mentioned in Wells’ discussion<sup>35</sup> for certain urban accents in the West Midlands. The space occupied by ‘heard’ (/ə/) in figure 3 is indeed more closed than that expected for RP, as the  $F1$  values slightly decreased from 581 Hz (RP) to 558 Hz and  $F2 - F1$  values increase from 800 Hz (RP) to 1023 Hz. Listed by Clark<sup>37</sup> as another characteristic in WM dialect, ‘lot’ is distinguished from ‘thought’; the closest samples we have is ‘hod’ (/ɒ/) and ‘hoard’ (/ɔ/) which can also be identified in figure 3.

By and large, the main characteristics of the WM accent noted in the literature can be seen in figures 2(a) and 3; the consistency of this study with previous ones increases confidence in the reliability of the results.

### 3.2.2. Normal and Whisper Vowels

More convergence of adjacent vowels is evident in the whispered samples both for men and women. As shown in figure 4, vowel groups such as {/ɔ/ and /ʊ/} or {/ʌ/ and /ɑ/ and /ɒ/} or {/ɪ/ and /i/} become similar in terms of formant characteristics while the back vowels show greater amounts of shift compared to front-closed vowels. This means that vowel durations as well as small changes in the shape of the vocal tract for differentiating these vowels in normal speech, are less significant in their whispered counterparts.

Another main difference for whispered speech appears to be that /ʌ/ has moved away from /ʊ/ and now overlaps instead with /ɑ/ and /ɒ/ for both men and women. This is interesting, since one of the Birmingham accent clues (as mentioned in subsection 3.2.1) seems to have disappeared, while the clue for /ɑ/ and /æ/ maintains approximately the same relationship as seen in spoken vowels.

The acoustic vowel diagrams presented in figure 5 show that for both men and women, the effect of whispering is greater for the first formants than the second formants, however  $F2$  values have also been shifted propor-

Table 1: *Average formant values in normal and whispered vowels for men<sup>a</sup>*

		/ɪ/	/i/	/ɛ/	/æ/	/ɑ/	/ʌ/	/ɒ/	/ə/	/ɔ/	/ʊ/	/u/
F1	N	337	399	568	746	685	500	593	523	457	442	380
	W	495	501	751	943	878	756	798	760	601	611	653
	S.A	0.468	0.255	0.321	0.263	0.289	0.510	0.345	0.451	0.315	0.381	0.717
F2	N	2356	2149	1895	1447	1111	1200	1006	1530	841	1019	1548
	W	2412	2257	1955	1595	1294	1328	1249	1643	1035	1165	1696
	S.A	0.023	0.050	0.031	0.102	0.165	0.106	0.240	0.073	0.229	0.143	0.095
F3	N	3035	2876	2643	2586	2505	2537	2536	2606	2605	2434	2348
	W	2979	2894	2717	2688	2628	2639	2587	2682	2550	2502	2583
	S.A	-0.018	0.006	0.028	0.039	0.049	0.040	0.020	0.029	-0.021	0.028	0.100

<sup>a</sup>N: Normal, W: Whisper, S.A: Shift amount in %

tionally. Interestingly, the extreme front-back and open-close vowels show almost consistent shift for both men and women but more significant shifts appear in central open-mid and close-mid vowels in which women show more tendency toward lower tongue positions in whisper mode. For example, first formants of mid-central vowels such as /ə/ and /ʌ/ in women shift up from 593 Hz and 634 Hz to 905 Hz and 949 Hz, respectively while the amount of shifts for the vowels at the extremes are significantly less; for example, from 390 Hz and 463 Hz to 511 Hz and 579 in /ɪ/ and /ɔ/, respectively for women.

Apart from the shifting, the size of quadrilaterals in figure 5 show different changes for men and women in conversion from spoken to whispers. While the area remains almost the same for men, the significant change appears in women’s diagrams particularly on the height of the quadrilateral corresponding to whispers.

The average size change of quadrilateral for men is 9% for both height and width at the extremes while the height shows increase but width decreases by this amount when moving from normal to whispered mode. These figures are 6% decrease in width and the significant 21% increase in height for quadrilaterals of women’s vowels. In fact, the major change appears in height increase of women’s whispers by 21%.

Figure 5 also demonstrates that the vowels for men and women both, by and large, occupy similar relative positions in whispered and normal modes,

Table 2: Average formant values in normal and whispered vowels for women<sup>a</sup>

		/ɪ/	/i/	/ɛ/	/æ/	/ɑ/	/ʌ/	/ɒ/	/ə/	/ɔ/	/ʊ/	/u/
F1	N	390	431	653	912	787	634	743	593	463	441	438
	W	511	522	838	1144	966	949	913	905	579	551	636
	S.A	0.309	0.211	0.283	0.255	0.227	0.497	0.228	0.526	0.250	0.248	0.450
F2	N	2665	2554	2200	1447	1167	1338	1117	1633	877	1090	1669
	W	2738	2653	2241	1730	1346	1532	1318	1774	1056	1130	1772
	S.A	0.027	0.038	0.018	0.195	0.153	0.144	0.180	0.086	0.203	0.036	0.061
F3	N	3248	3087	2984	2832	2853	2825	2888	2865	2807	2701	2633
	W	3239	3140	3007	2929	2950	2844	2971	2897	2928	2826	2759
	S.A	-0.002	0.017	0.007	0.034	0.034	0.006	0.028	0.011	0.043	0.046	0.048

<sup>a</sup>N: Normal, W: Whisper, S.A: Shift amount in %

however, it is difficult to arrive at the simple summary of the differences that are seen in these figures while a few general observations have been mentioned to the extent that conventional articulatory interpretations of formant data are valid.

The diverse amount of shifts in figure 4 show that each vowel has its own variation when converting to whispered speech and this amount also varies in terms of formant number. Tables 1 and 2 summarise these variations for the first three formants in whisper and normal speech for men and women, respectively.

As shown in Table 1 and 2, all first and second formants are shifted up. The shifts amounts are ranging from 25% in /i/ to 71% in /u/ for men and from 21% in /i/ to 52% in /ə/ within the first formants and from 2.3% in /ɪ/ to 24% in /ʌ/ for men and from 2.7% in /ɪ/ to 20% in /ɔ/ for women within the second formants. Furthermore, significant shifts occur in the first formants with averages of 39% ( $\sigma$  : 13%) and 31% ( $\sigma$  : 11%) while these numbers are 11% ( $\sigma$  : 7%) and 10% ( $\sigma$  : 7%) for the second formants for men and women, respectively. As mentioned before, third formants are almost consistent between normal and whisper modes.

Whispered speech *per. se.* is often of substantial intelligibility<sup>38,15</sup>, therefore the overall intelligibility of whispered vowels should be acquired consequently. However, the whispered samples reported in this paper have been

investigated through informal listening tests, with the results indicating that main vowels such as /ɪ,ɛ,æ,ɒ,ɔ,u/ are recognised well without confusion with each other but there are frequent mismatch occurring between vowel groups such as {/ʌ/ and /ɑ/ and /ɒ/} or {/ɪ/ and /i/}; for example, when whispered ‘heed’ was played, the listeners could not identify it from ‘hid’ but it has been simply recognised from other carriers.

#### 4. Conclusion

A study to establish a vowel formant space for whispered speech has been carried out. By comparing whispered vowels with the corresponding phonated samples separately for men and women, a table outlining the amount of shift for each vowel and formant was presented, while distribution of formant values for normal and whispered samples was illustrated.

Acoustic vowel diagrams were presented showing that more shift occurs within central open-mid and close-mid vowels rather than the extreme front-back and open-close vowels both in men and women in whisper mode. In fact, shift amounts in whispered vowels depend on the way of articulation in normal mode.

Since the study was conducted on speakers from Birmingham, the analysis also briefly considered the effect of British West Midlands (WM) accent in comparison with Standard English (RP). Some of these accentual effects in the whispered speech, as observed in the data, were also discussed.

#### Acknowledgment

The authors would like to acknowledge the support of a Singapore National Medical Research Council development grant for this research. The project is partly supported by an NMRC EDG (EDG07MAY002) grant.

#### References

1. Hillenbrand J, Gayvert RT. Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech and Hearing Research* 1993;36:694–700.
2. Miller JD. Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America* 1989;85(5):2114–2134.

3. Bele IV. The speaker's formant. *Journal of Voice* 2006;20(4):555 – 578.
4. Nawka T, Anders LC, Cebulla M, Zurakowski D. The speaker's formant in male voices. *Journal of Voice* 1997;11(4):422 – 428.
5. Syrdal AK. Aspects of a model of the auditory representation of american english vowels. *Speech Communication* 1985;4:121–135.
6. Nearey TM. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America* 1989;85(5):2088–2113.
7. Gelfer MP, Mikos VA. The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of Voice* 2005;19(4):544 – 554.
8. Peterson GE, Barney HL. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 1952;24:175–184.
9. Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of american english vowels. *Journal of the Acoustical Society of America* 1995;97(5):3099–3111.
10. Fox RA, Jacewicz E. Analysis of total vowel space areas in three regional dialects of american english. In: *Acoustics 08 Paris*. 2008:495–500.
11. Solomon NP, McCall GN, Trosset MW, Gray WC. Laryngeal configuration and constriction during two types of whispering. *Journal of Speech and Hearing Research* 1989;32:161–174.
12. Rubin AD, Praneetvatakul V, Gherson S, Moyer CA, Sataloff RT. Laryngeal hyperfunction during whispering: Reality or myth? *Journal of Voice* 2006;20(1):121 – 127.
13. Catford JC. *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press; 1977.
14. Sundberg J, Scherer R, Hess M, Mller F. Whispering—a single-subject study of glottal configuration and aerodynamics. *Journal of Voice* 2010;24(5):574 – 584.
15. Jovicic ST, Saric Z. Acoustic analysis of consonants in whispered speech. *Journal of Voice* 2008;22(3):263 – 274.



16. Kallail KJ, Emanuel FW. Formant-frequency difference between isolated whispered and phonated vowel samples produced by adult female subject. *Journal of Speech and Hearing Research* 1984;27:245–251.
17. Jovicic ST. Formant feature differences between whispered and voiced sustained vowels. *Acta Acustica united with Acustica* 1998;84:739–743.
18. Smith FW. A formant study of whispered vowels. Ph.D. thesis; University of Oklahoma; 1973.
19. Sharifzadeh HR, McLoughlin IV, Ahmadi F. Reconstruction of normal sounding speech for laryngectomy patients through a modified celp codec. *IEEE Transactions on Biomedical Engineering* 2010;57(10):2448–2458.
20. Morris RW, Clements MA. Reconstruction of speech from whispers. *Medical Engineering and & Physics* 2002;24:515 – 520.
21. Weitzman RS, Sawashima M, Hirose H. Devoiced and whispered vowels in japanese. *Annual Bulletin, Research Institute of Logopedics and Phoniatics* 1976;10:61–79.
22. Esling JH. Laryngographic study of phonation type and laryngeal configuration. *Journal of the International Phonetic Association* 1984;14:56–73.
23. Tartter VC. What’s in a whisper? *Journal of the Acoustical Society of America* 1989;86:1678–1683.
24. Fant G. Acoustic Theory of Speech Production. second printing, 1970 ed.; The Hague: Mouton; 1960.
25. Thomas IB. Perceived pitch of whispered vowels. *Journal of the Acoustical Society of America* 1969;46:468–470.
26. Stevens HE. The representation of normally-voiced and whispered speech sounds in the temporal aspects of auditory nerve responses. Ph.D. thesis; University of Illinois; 2003.
27. Lehiste I. Suprasegmentals. Cambridge: MIT Press; 1970.

28. Gao M. Tones in whispered chinese: articulatory features and perceptual cues. Master's thesis; University of Victoria; 2002.
29. D'Arcy S. The effect of age and accent on automatic speech recognition performance. Ph.D. thesis; University of Birmingham; 2007.
30. Wu CH, Chiu YH, Shia CJ, Lin CY. Automatic segmentation and identification of mixed-language speech using delta-bic and lsa-based gmms. *IEEE Transactions on Audio, Speech, and Language Processing* 2006;14(1):266 – 276.
31. Ramus F, Mehler J. Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America* 1999;105(1):512–521.
32. Zissman M. Comparison of four approaches to automatic language identification of telephone speech. *IEEE Transactions on Audio, Speech, and Language Processing* 1996;4(1):31–44.
33. Adami AG, Hermansky H. Segmentation of speech for speaker and language recognition. In: *Eurospeech*. 2003:.
34. Mustafa K, Bruce IC. Robust formant tracking for continuous speech with speaker variability. *IEEE Transactions on Speech and Audio Processing* 2006;14(4):435– 444.
35. Wells JC. Accents of English, Volume 2: The British Isles. Cambridge: Cambridge University Press; 1982.
36. Hughes A, Trudgill P. English Accents and Dialects. London: Edward Arnold; 1987.
37. Clark U. Varieties of English 1: The British Isles; chap. The English West Midlands: phonology. Berlin: Mouton de Gruyter; 2008:145–177.
38. Ito T, Takeda K, Itakura F. Analysis and recognition of whispered speech. *Speech Communication* 2005;45(2):139–152.