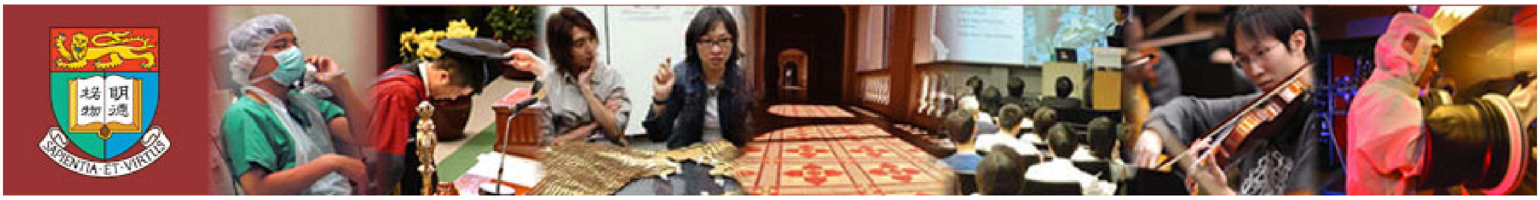

Contribution of voice fundamental frequency and formants to the identification of speaker's gender

Poon Siu Fung Morris

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2011.

Abstract

Identification of gender from speech sounds has been found to rely on speakers' voice fundamental frequency (F_0) and formant frequencies. The present study aims at examining the contribution of F_0 and formants to the correct detection of speaker's gender. Based on the vowel sustained by a male and female speaker, 200 vowels were synthesized with a range of F_0-formant combinations. The synthesized vowels were presented to 28 native Cantonese-speaking listeners to judge the perceived speakers' gender for each of the synthesized stimuli. Results revealed that F_0 was the primary cue for speakers' gender perception while formants contributed little. The cutoff F_0 values for male and female identification were found to be 162.01 Hz and 204.97 Hz, respectively. When F_0 was below 162.01 Hz or above 204.97 Hz, listeners reliably and correctly identified the speakers as male or female, respectively.

Introduction

Human speech production can be conceptualized by using the source-filter theory. According to the theory, human speech output is a product of the sound source (vocal fold vibrations during phonation) and the filter (vocal tract resonances). As an important attribute of sound source, voice fundamental frequency (F_0) of any speech signal is the physical measure of vocal fold vibratory rate, and it is perceptually correlated with the perceived pitch (Kent & Read, 1992). During speech production, the source signal is modified by vocal tract resonance, resulting in some frequencies being amplified (formants), while other frequencies being suppressed (Kent & Read, 1992).

During puberty, significant growths in vocal folds (source) and vocal tract (filter) are evident in both males and females, but males' laryngeal size and vocal tract length increase to a much greater extent than females' (e.g., Whiteside, 2001). The different laryngeal and vocal tract developmental trajectories observed in the two genders lead to divergence in F_0 and formant frequencies observed between adult males' and females' voices. In general, male vocal folds are larger than female ones by approximately 60% (Titze, 1989), resulting in male F_0 falling to about 50% to 60% of female F_0 (Owren, Berkowitz, & Bachorowski, 2007). In addition, male vocal tract is approximately 15% to 20% longer than female one. As a consequence, male voice formant frequencies are about 80% to 90% of female ones (Bachorowski & Owren, 1999).

In view of such sexual dimorphism of voices, some researchers have attempted to separate males' and females' voices by using statistical pattern classification method (e.g., Childers & Wu, 1991; Bachorowski & Owren, 1999; Hillenbrand & Clark, 2009). Childers and Wu (1991) analyzed 10 vowels naturally produced by 27 males and 25 females using various acoustic parameters including F_0, formant frequencies, bandwidths and amplitudes. It was reported that male vowels show significantly lower F_0 and formant frequencies (F1 - F4) than female vowels. Based on the complex pattern recognition model described in their earlier work (Wu & Childers, 1991), they utilized the different acoustic parameters, either single or combined, to recognize speakers' gender. Combining F1 to F4, they achieved 100% gender recognition rate. But when using F_0 alone, the average gender recognition rate dropped to 96.2%. They concluded that F_0 and formant characteristics are nearly reliable for gender recognition, with formant characteristics showing a slight advantage. Hillenbrand and Clark (2007) also carried out pattern recognition tests. They employed a quadratic discriminant analysis technique to classify 12 vowels produced by 45 males and 48 females based on F_0 and formant frequencies (F1 - F3). Together with the findings in Bachorowski and Owren (1999), they concluded that both F_0 and formant frequencies could differentiate male and female vowels accurately, with F_0 having a slightly higher accuracy than formants (about 96% and 92% accuracies for F_0 and formant frequencies, respectively). In addition, they found that when both F_0 and formant frequencies were used, the gender recognition accuracy further improved. Although both research teams came up with slightly discrepant results, both of them highlighted the role of F_0 and formants in distinguishing males' and female's voices.

Studies of statistical pattern recognition found that F_0 and formants could distinguish males' and female's voices, but the results may not be directly applied to how listeners perceive gender of voices. Some researchers attempted to investigate the contribution of F_0 and formant frequencies to the perception of speakers' gender based on perceptual testing (e.g., Coleman, 1976; Gelfer & Mikos, 2005; Smith & Patterson, 2005; Whiteside, 1998). In Coleman (1976), Gelfer and Mikos (2005), and Whiteside (1998), stimuli were divided into four types: (1) male F_0 paired with male formants; (2) female F_0 paired with female formants; (3) male F_0 paired with female formants; and (4) female F_0 paired with male formants. Findings in these studies were consistent: when F_0 and formant frequencies pointed to the same gender (i.e., Types 1 & 2), the correct gender identification rate was higher in the male than in the female case. However, when F_0 and formant frequencies conflicted with each other (i.e., Types 3 & 4), discrepant findings regarding gender identification were observed. Coleman found that such stimuli tended to be perceived as males, thus suggesting that male cues (both F_0 and formant frequencies) were stronger than female cues. Gelfer and Mikos, and Whiteside suggested that listeners relied on F_0 more than formants for speakers' gender identification tasks.

Despite the interesting findings reported by Coleman (1976), questions are raised regarding the methodology. In particular, connected speech produced by normal speakers using laryngeal vibrators (an equipment with which voice F_0 can be manipulated) as sound source was used in Coleman's study. Gelfer and Mikos (2005) commented that the connected speech materials could have provided additional cues other than F_0 and formant frequencies, such as intonation and stress pattern, to aid listeners in gender identification. In addition, the flat intonation contour associated with the laryngeal vibrators may have favored the perception of maleness.

Unlike Coleman (1976), Gelfer and Mikos (2005), and Whiteside (1998) made use of synthesized vowels for their perceptual judgment tasks, and both studies yielded similar conclusions: listeners tended to make use of F_0 more than formants for gender identification. Yet, Gelfer and Mikos used speech stimuli that appeared to be better designed. Whiteside synthesized the speech stimuli using averaged formant data extracted from vowel segments of sentences produced by six speakers. The use of short synthesized vowels (100 ms for long vowels and 50 ms for short vowels) in the study with varying F_0 contours could have affected listeners' judgment of gender (Whiteside, 1998). Gelfer and Mikos modified the way speech stimuli were synthesized. They used individual formant data extracted from the sustained vowels to synthesize one set of speech stimuli. Moreover, the synthesized vowels had a longer duration (three seconds) and flat F_0 contour.

In contrast to the previous studies, based on five vowels produced by a single male speaker, Smith and Patterson (2005) synthesized 245 speech stimuli over a wide range of F_0-formants combinations for their perceptual experiment. Listeners judged the size and age/sex (i.e., man, woman, boy, or girl) for each stimulus. They concluded that both F_0 and formant frequencies contributed to the perception of speaker's sex and age. Such a design allowed researchers to better reveal the changes of speakers' gender perception along variations of F_0 and formants.

Studies of English-speaking populations have demonstrated evidence supporting the roles of F_0 and formant frequencies in speaker's gender perception. Yet, there is no consensus on the relative importance of the two acoustic cues to gender identification. In addition, studies on the Cantonese-speaking population are lacking. There has been evidence that Mandarin speakers produced significantly higher F_0 for all vowels tested and higher F3 for some vowels than do Caucasian, African-American, and Hindi Indian speakers (see Andrianopoulos, Darrow & Chen, 2001a; 2001b). It is thus possible that the range of F_0 and formant frequencies perceived as typical to either gender is not the same in individuals speaking different languages or from different cultural groups. Yet, there is no definite conclusion regarding if and how language and culture affect the way speakers' gender is perceived by listeners. The present study utilized perceptual experiments with synthesized vowels to find out: (1) *the relative contributions of F_0 and formant frequencies to the perception of speakers' gender in Cantonese-speaking population, and (2) the cutoff F_0-formant frequencies combinations corresponding to the perception of both genders*. In the study, speech stimuli were synthesized in a way similar to that used by Smith and Patterson (2005).

Method

Participants

Ten male and 10 female adult native Cantonese speakers (age: *M* = 21.36 years, S.D. = 1.19, range = 20 - 26 years) were recruited as speakers to record speech samples for analysis and synthesis. For the perceptual experiment, 28 (11 male and 17 female) adult native Cantonese speakers (age: *M* = 19.79 years, S.D. = 1.03, range = 19 - 24 years) were recruited as listeners. All participants were university students who reported having no known speech, language and hearing problems and volunteered to participate in the study.

Speech Task and Recording Procedure

The recordings took place in a sound treated room of the Speech Science Laboratory of the University of Hong Kong, which had an ambient noise level of less than 50 dB SPL. In the speech task, each speaker was seated in the room individually and instructed to sustain the syllable /a/ at the high-level tone, as in the Cantonese word "丫", for approximately five seconds at a comfortable loudness level. The speech samples were recorded by using a high-quality microphone (SM58, Shure) via a preamplification unit (PreMobile USB, M-Audio). During the recording, a mouth-to-microphone distance of approximately 8 cm was maintained. Prior to the recording, a brief practice period was provided to participants to familiarize themselves with the recording environment and procedure. Audio signals were digitized with a sampling rate of 20 kHz and quantization rate of 16 bits/sample by using the Praat software (Boersma & Weenink, 2009). The digitized signals were stored in a computer for later analyses.

Acoustic Analysis

A three-second segment was extracted from the medial portion of each recorded vowel which was used for later analysis by using Praat (Boersma & Weenink, 2009) for F_0 and formant frequencies (F1 to F3). This vocalic portion was selected as it represented the most stable production of the steady state vowel.

Results of the analysis are shown in Table 1. F_0 of vowels produced by male speakers was lower than that produced by female speakers. The F_0 values of both groups appeared high when compared to the average speaking F_0 values reported in English literature (e.g., Baken & Orlikoff, 2000). This is probably because the speakers were instructed to sustain the vowel at the high-level tone. Results also showed that the mean F1, F2, F3 and overall average formant frequency (averaged across F1 to F3) of vowels produced by male speakers were all lower than those produced by female speakers. The analysis results were consistent with the prediction that male voice generally had lower F_0 and formant frequencies. On the other hand, there was no overlap between F_0 ranges of male and female speakers. In contrast, the ranges of F1, F2, F3 and overall average formant frequency showed overlapping between the two gender groups.

Table 1. Mean (*M*), standard deviation (S.D.) and range of vowel fundamental frequency (F_0) and formant frequencies (in Hertz) for male (N=10) and female (N=10) speakers.

	Male speakers	Female speakers
F_0	*M* = 139.53, S.D. = 15.34,	*M* = 233.16, S.D. = 11.36,
	Range = 118.36 - 159.02	Range = 210.38 - 245.86
F1	*M* = 741.19, S.D. = 101.62,	*M* = 874.05, S.D. = 94.85,
	Range = 576.53 - 874.41	Range = 647.17 - 986.68
F2	*M* = 1264.70, S.D. = 102.64,	*M* = 1446.69, S.D. = 136.38,
	Range = 1075.66 - 1398.92	Range = 1246.54 - 1603.87

F3	*M* = 2690.96, S.D. = 229.74, Range = 2341.56 - 3054.32	*M* = 3227.33, S.D. = 193.35, Range = 3000.31 - 3607.25
F1-F3	*M* = 1565.62, S.D. = 119.97, Range = 1373.76 - 1754.13	*M* = 1849.35, S.D. = 115.64, Range = 1694.34 - 2043.67

One male and one female speaker who had the overall average formant frequency (averaged across F1 to F3) that was closest to the corresponding gender group mean formant frequencies were selected. Their data were used for the formant scaling across gender and synthesis of stimuli for subsequent perceptual experiment. The ratios of mean F1, F2 and F3 between these two speakers were calculated, respectively. Then a composite formant frequency scale factor (female/male) was calculated by averaging the three ratios, which was found to be 1.20. This scale factor was consistent with those suggested by other researchers (e.g., Bachorowski & Owren, 1999; Fant, 1966).

Stimuli Synthesis

Formant data of the selected male and female speakers were used as the basis for creating two sets of synthesized vowels. Vowels of one male and one female were used to synthesize the stimuli because it is not known if gender of the original vowels can affect results of gender perception. Using Praat (Boersma & Weenink, 2009), the male speaker's F1 to F3 were multiplied by 10 scale factors evenly derived from 1.00 to 1.20 and coupled with F_0 that was scaled to 10 values within 100 to 250 Hz, creating 100 synthesized vowels (10 formant frequency values x 10 F_0 values) each of which was three-second long. The vowels were then duplicated to form a set of 200 stimuli (the "male stimuli"). Similarly, another set of 200 stimuli (the "female stimuli") was synthesized by using the same method, except that the female vowel was used as basis of synthesis and the formant frequencies were multiplied by 10 scale factors from 1.00 to 0.83 (i.e., 1/1.20). The scaling procedure aimed to simulate a range of F_0 and formant frequencies that are male-appropriate, female-appropriate or gender-ambiguous. Upon completion of such process, 400 stimuli with different combinations of F_0 and formant frequencies were prepared.

Perceptual Experiment

The two sets of synthesized vowels were presented to the listeners in two separate sessions. During each session, the listeners were seated in groups in a sound-treated room with an ambient noise level less than 60 dB SPL. The stimuli were presented to the listeners in a randomized order via high-quality loudspeakers. For each stimulus, listeners were instructed to judge whether the speaker was a male or female. In the case of ambiguity, they were asked to guess. Upon listening to a stimulus, the listeners circled the gender they perceived on an answer sheet provided. An inter-stimulus pause of about five seconds was introduced to provide sufficient time for the listeners to complete the judgment task. A short break was provided after every 50 stimuli were presented to minimize possible fatigue effect.

Intra-Listener Reliability

All synthesized vowels were presented twice and the results obtained from the first and second presentations from all listeners were used to calculate the intra-listener reliability. Seven listeners demonstrated a reliability of less than 75% and their data were excluded from analysis. For the remaining 28 listeners yielded an average reliability of 83.13%, indicating that results from listeners' perception were consistent and reliable.

Results

Results of gender identification of male and female stimuli are shown in Figures 1 and 2 respectively. Specifically, Figures 1 and 2 show the percentages of stimuli being perceived as male's and female's voices respectively over all F_0-formant combinations. Figure 1 reveals a general trend of decreasing rate of the stimuli being perceived as male's (male identification) with increasing F_0. The opposite pattern is observed for female identification as shown in Figure 2: Female identification increased with increasing F_0. Yet, identification rate of either gender showed little or inconsistent changes as formants changed.

The identification contours appear to be smooth towards the two ends of F_0 (i.e., near 100 Hz and 250 Hz) as compared to medial F_0. This pattern indicates that at high or low F_0, gender identification did not change much with formants; whereas when F_0 was set to the middle range, the gender identification rate fluctuated with formant frequencies.

For both male and female stimuli, a clear cutoff along the F_0 axis could be identified, but not along the formant axis. By using extrapolation, the cutoff F_0 for 75% correct gender identification was found to be 162.01 Hz and 204.97 Hz for male and female stimuli respectively. An identification rate of 75% was used as the cutoff criterion because this value represented the mid-point between chance response (wild guessing) and perfect discrimination in experiments involving two choices (Gescheider, as cited in Owren, Berkowitz & Bachorowski, 2007). In other words, when voice F_0 fell to 162.01 Hz or below, the voices were mainly (more than 75% chance) identified as males'. When voice F_0 rose to 204.97 Hz or above, the voices were mainly identified as females'. When voice F_0 was between 162.01 and 204.97 Hz, the speakers' gender could not be identified reliably, and there seemed to lack a consistent pattern.

In addition, results also showed that the accuracy of male identification was slightly higher than female identification. For the male stimuli, when F_0 at 100 Hz was paired with formant scale factor 1.00 (i.e., both F_0 and formants were male-appropriate), a male identification rate of 100% was achieved. For the female stimuli, when F_0 at 250 Hz was paired with formant scale factor 1.00 (i.e., both F_0 and formants were female-appropriate), the female identification rate was 94.64%.

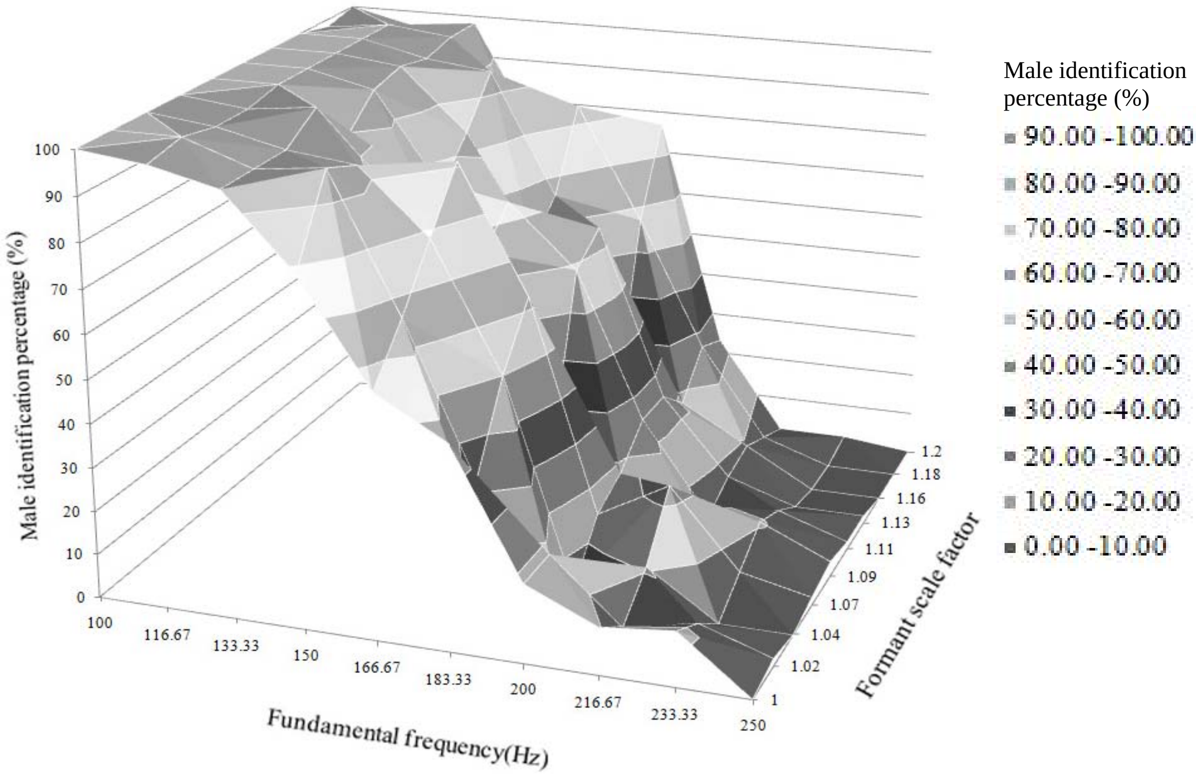

Figure 1. Percent male identification over different fundamental frequency (F_0)-formant combinations.

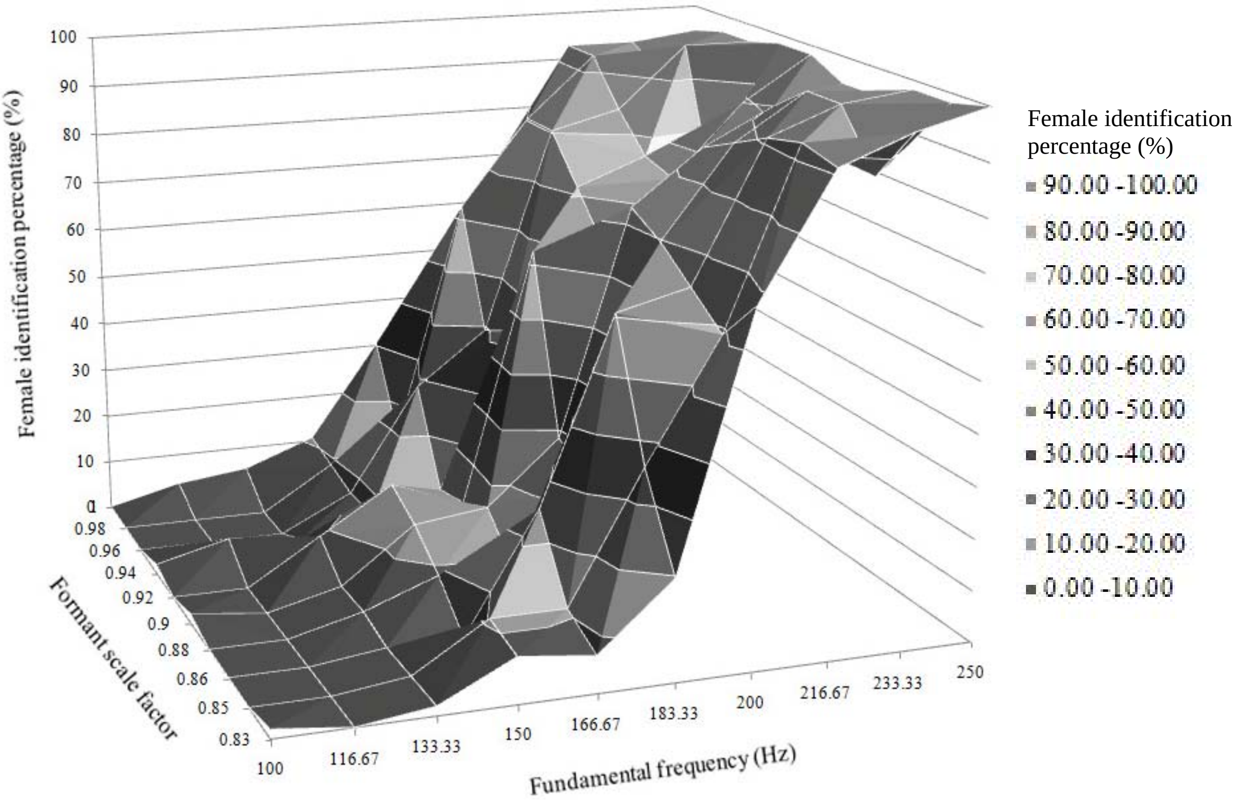

Figure 2. Percent female identification over different fundamental frequency (F_0)-formant combinations.

Discussion

The present study attempted to examine the contribution of F_0 and formant frequencies to speakers' gender identification. Results showed a clear slope along the F_0 axis, indicating that the chance of stimuli being perceived as a male's voice decreased with increasing F_0 (see Figures 1 and 2). Yet, the effect of formants on gender perception appeared to be smaller and unclear, as indicated by the lack of a clear contour along the formant axis in the Figures. This pattern provides a straightforward answer to the research question about which acoustic cue, F_0 or formant frequencies, contributes more to speakers' gender perception. Results of the current study support that listeners mainly depend on F_0, but not formants, to perceive speaker's gender.

The phenomenon that speakers' gender perception depended mainly on F_0 but not formants is more obvious when F_0 is high or low. In both Figures 1 and 2, listeners showed highly consistent gender judgment at both ends of F_0. According to Figure 1, for the stimuli perceived as male's voices, male identification rate ranged narrowly from 98.21% - 100% and 0% - 3.57% across different formants when F_0 was 100 Hz and 250 Hz, respectively. However, when F_0 was at the medial portion (183.33 Hz), male identification rate fluctuated across different formants to a much larger extent from 19.64% - 76.79%. Similar results were found in the female stimuli. As shown in Figure 2, for the stimuli perceived as female's voices, female identification rate also ranged narrowly from 0% - 1.79% and 94.64% - 100% across different formants when F_0 was 100 Hz and 250 Hz, respectively. When F_0 was at the medial portion (183.33 Hz), female identification rate changed variably across different formants from 19.64% - 62.50%. This pattern of identification contour suggests that cues from F_0 alone are strong enough for stimuli to be perceived as a particular gender, regardless of formants. Coleman (1976) argued that when F_0 was at a range that was neither typical to male nor female, formant frequencies would take up the role of guiding speakers' gender perception. Results of the present study do not support this argument. In this study, even when F_0 fell within a gender-ambiguous range, changes in formants did not necessarily lead to consistent changes to speakers' gender perception. It follows that F_0 is the primary cue for gender perception, with formants providing very little or no cues. This finding is consistent with those reported in the studies of Coleman (1976), Gelfer and Mikos (2005), and Whiteside (1998).

However, findings in Smith and Patterson (2005) do not seem to agree with the current conclusion that F_0 is the major cue for gender perception. They reported that both F_0 and formants contributed to perception of gender and age. The discrepancy may be accounted for by the different experimental design used in their study and the current study. In Smith and Patterson's study, the range of formants used for synthesizing stimuli extended to children's range, which was significantly wider than that used in this study. Yet, only seven data points along this large range of formants were used to synthesize the stimuli. The limited data points and large range of formants used in their study might have over-simplified the relationship between formants and gender perception. In the current study, where only the formants within normal adult male and female range were investigated in greater details, it was found that the relationship between formants and perceived gender was not as clear.

Since both age and gender were investigated together in their study (i.e., listeners were instructed to label each stimulus as man, woman, boy or girl), it is not known whether the role of formants in identifying speakers' gender was still be as important if only the range of adults was studied. In fact, by inspecting their results, when formants were at adult range, perception of male and female voices mainly depended on F_0 but did not change much with formants.

The present results of the acoustic analysis of vowels may give some hints on why voice F_0 contributes more to speaker's gender perception than do formants. Referring to Table 1, the male and female ranges of F_0 were distinctive with no overlapping region. In contrast, the male and female ranges of F1, F2, F3 and overall average formant frequency did show some overlap. As male and female F_0 ranges are discrete while their formant ranges are overlapping to a certain extent, listeners may find it easier and more reliable to use F_0 rather than formants as the major cue for classifying gender. This is a possible explanation for why the correct gender identification rates change mainly with F0 changes but fluctuated across different formants.

Another aim of the present study is to identify the cutoff F_0-formant frequency combinations that are associated with the perception of male's and female's voices. As discussed previously, the effect of formants on the perception of speakers' gender appeared to be ambiguous. Sensible discussion was restricted only to cutoff F_0 values. The 75% gender identification cutoff F_0 for male and female stimuli are found to be 162.01 Hz and 204.97 Hz respectively. This indicates that speakers' gender perception depending on voice F_0 was categorical in nature. It was found the F_0 ranging from 100.00 Hz to 162.01 Hz was the male-appropriate range, in which listeners could reliably and correctly identify the speakers as males. F_0 ranging from 204.97 Hz to 250.00 Hz defines the female-appropriate range, in which listeners could reliably and correctly identify the speakers as females. On the contrary, F_0 between 162.01 Hz and 204.97 Hz was the gender-ambiguous range. When F_0 fell within this range, listeners failed to judge speakers' gender reliably. In this case, it is suspected that listeners might have made use of other suprasegmental cues or just random strategy to judge speakers' gender.

The present results also indicate that perception of male's stimuli was more accurate than that of female's stimuli, when both F_0 and formant cues were not conflicting with each other. For the male stimuli with F_0 at 100 Hz coupled with male-appropriate formant, 100% correct identification rate was yielded. However, for the female stimuli with F_0 at 250 Hz coupled with female-appropriate formant, only 94.64% correct identification rate was obtained. This "male advantage" pattern of gender identification has also been reported in Coleman (1976), Gelfer and Mikos (2005), and Whiteside (1998). The current results lend credence to the hypothesis proposed by Owren, Berkowitz and Bachorowski (2007), who adopted a developmental account for explaining the phenomenon of male identification being more accurate than female identification. As a male adolescent progresses to puberty, his male voice drops distinctively in terms of F_0 and formant frequencies due to significant growths in vocal folds and vocal tract. F_0 and formant frequencies of female voices also change during puberty but the extent is much smaller than the change in male's voices.

Owren et al. thus argued that the low F_0 with low formant frequencies were distinctive features of adult male's voices but the F_0-formant frequency combinations for adult females were more diverse and closer to children's. Following this logic, the low F_0 and low formants combination renders a template for male's voices, allowing listeners to identify voices readily as male's voices. But for female identification, listeners lack a specific template that defines female's voices. Owren et al. (2007) concluded that listeners identified male's voices more efficiently than female's voices. However, the asymmetrical identification results may also reveal the limitation of using synthesized vowels as perceptual stimuli. In natural speech, additional suprasegmental or metalinguistic cues (e.g., breathiness) may also affect the perception of gender. According to Klatt and Klatt (1990), female voices are generally more breathy than male voices. When this difference is absent in synthesized vowels, perception of male's voices may be favored, yielding a low female identification. Apparently, this issue is yet to be investigated. As suggested by Hillenbrand and Clark (2009), the relative strength of these suprasegmental cues (e.g., voice qualities) and acoustic cues (e.g., F_0 and formants) in distinguishing male's and female's voices was still unknown and worth further investigation.

In the present study, as well as other previous studies, the method of synthesizing stimuli involves shifting the formant frequencies as a whole. In other words, F1, F2 and F3 are multiplied by the same scale factor equally. The underlying assumption of such method of manipulating formants is that formant scaling between male's and female's voices is uniform. However, in reality, the relationship between male and female formants is more complex than a uniform one. Fant (1966) pointed out that this simple scaling method cannot accurately demonstrate male-female formant relationship because males tend to have a larger ratio of pharyngeal length to mouth cavity length and larger larynx than females. The use of simple scaling method may partially contribute to the unclear effect of formants on gender identification in this study. To improve, further research may target to scale the formant frequencies independently so that the scaling can be closer to reality. Otherwise, more sophisticated source-filter synthesizer is needed overcome this difficulty.

The present study is an initial attempt to investigate speakers' gender identification in Cantonese-speaking population. Results clearly show that, for an isolated vowel /a/, F_0 is the most salient cue for perceiving speakers' gender. The present findings agree with some previous studies in English-speaking populations. However, as only one vowel has been used in the present study, it is difficult to conclude whether language and/or culture affect gender perception. In addition, the findings only lay the foundations but are not conclusive for perception of speakers' gender from speech sounds. Human speech is complex in that it contains a great variety of sounds, such as consonants and vowels. Extra suprasgemental information is also present in our daily speech, including tones, intonation, voice quality, etc. Further research may employ words and sentences to investigate the contributions of these parameters to speakers' gender identification. Research on the relative importance of acoustic cues and suprasegmental cues is also needed. Understanding the relationship between these parameters and perceived speakers' gender will bring about clinical implications. With such research endeavor, speech therapists can better facilitate clients with puberphonia or transgendered clients to alter their voice or speech style, so as to enhance or change the gender perceived by others.

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Dr Manwa L. Ng, for his invaluable guidance and insightful advice throughout the course of the study. I would also like to express my heartfelt thanks to all the participants in the study, who have kindly devoted their precious time and effort to this study. Appreciation also goes to my clinical supervisor, Ms Karen Yu, for her kind assistance in conducting a perceptual experiment. In addition, I want to extend my thanks to my classmates who have offered stimulating opinions to this study. Finally, I wish to take this opportunity to express my gratefulness to my family, friends and classmates for their continuous encouragement and unconditional support throughout the years.

References

Andrianopoulos, M. V., Darrow, K., & Chen J. (2001a). Multimodal standardization of voice among four multicultural populations formant structures. *Journal of Voice, 15*(1), 61-77.

Andrianopoulos, M. V., Darrow, K., & Chen J. (2001b). Multimodal standardization of voice among four multicultural populations: Fundamental frequency and spectral characteristics. *Journal of Voice, 15*(2), 194-219.

Bachorowski, J. A., & Owren M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America, 106*(2), 1054-1063.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical management of speech and voice* (2nd ed.). San Diego, CA: Singular Publishing Group.

Boersma, P, & Weenink, D. (2009). *Praat: doing phonetics by computer*. Retrieved March 18, 2011, from http://www.fon.hum.uva.nl/praat/

Childers, D. G., & Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis. *Journal of the Acoustical Society of America, 90*(4), 1841-1856.

Coleman, R. O. (1976). A comparison of the contribution of two voice quality characteristics to the perception of maleness and femaleness in the voice. *Journal of Speech and Hearing Research, 19*, 168-180.

Fant, G. (1966). *A note on vocal tract size factors and non-uniform F-pattern scaling.* Retrieved February 22, 2011, from http://www.speech.kth.se/qpsr

Gelfer, M. P., & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of Voice*, *19*(4), 544-554.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f_0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics, 71*(5), 1150-1166.

Kent, R. D., & Read, C. (1992). *The acoustic analysis of speech*. San Diego, CA: Singular Publishing Group, Inc.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*, 820-857.

Owren, M. J., Berkowitz, M., & Bachorowski, J.-A. (2007). Listeners judge talker sex more efficiently from male than from female vowels. *Perception & Psychophysics, 69*(6), 930-941.

Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgments of speaker size, sex and age. *Journal of the Acoustical Society of America, 118*, 3177-3186

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America, 85*(4), 1699-1707.

Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *Journal of the Acoustical Society of America, 110*, 464-478.

Wu, K., & Childers, D. G. (1991). Gender recognition from speech. Part I: Coarse analysis. *Journal of the Acoustical Society of America, 90*(4), 1828-1840.