Development of a Femininity Estimator for Voice Therapy of Gender Identity Disorder Clients

Nobuaki Minematsu† and Kyoko Sakuraba‡

†The University of Tokyo
‡Kiyose-shi Welfare Center for the Handicapped
mine@k.u-tokyo.ac.jp, sakuraba@mtd.biglobe.ne.jp

Abstract. This work describes the development of an automatic estimator of perceptual femininity (PF) of an input utterance using speaker verification techniques. The estimator was designed for its clinical use and the target speakers are Gender Identity Disorder (GID) clients, especially MtF (Male to Female) transsexuals. The voice therapy for MtFs, which is conducted by the second author, comprises three kinds of training; 1) raising the baseline *F0* range, 2) changing the baseline voice quality, and 3) enhancing *F0* dynamics to produce an exaggerated intonation pattern. The first two focus on static acoustic properties of speech and the voice quality is mainly controlled by size and shape of the articulators, which can be acoustically characterized by the spectral envelope. Gaussian Mixture Models (GMM) of *F0* values and spectrums were built separately for biologically male speakers and female ones. Using the four models, PF was estimated automatically for each of 142 utterances of 111 MtFs. The estimated values were compared with the PF values obtained through listening tests with 3 female and 6 male novice raters. Results showed very high correlation (R=0.86) between the two, which is comparable to the intra- and inter-rater correlation.

1 Introduction

Advanced speech technologies are applied not only for man-machine interface and entertainment but also for medical treatment [1] and pronunciation training of foreign language education [2]. Many works were done for developing cochlea implants [3–5] and artificial larynxes [6–8] and, recently, the technologies have been applied to realize an on-line screening test of laryngeal cancer [9] as well as an on-line test of pronunciation proficiency of foreign languages [10]. The present work examines the use of the technologies for another medical treatment; voice therapy for GID clients.

A GID individual is one who strongly believes that his or her true psychological gender identity is not his or her biological or physical gender, i.e., sex. In most of the cases, GID individuals live for years trying to conform to the social role required by their biological gender, but eventually seek medical and surgical help as well as other forms of therapy in order to achieve the physical characteristics and the social role of the gender which they feel to be their true one.

In both cases of FtMs (Female-to-Male) and MtFs, many of them take hormone treatment in order to make physical change of their bodies and the treatment is certainly effective for both cases. However, it is known that the hormone treatment brings about sufficient change of the voice quality only for FtMs [11, 12]. Considering that the voice quality is controlled by physical conditions of the articulators, the vocal folds and the vocal tract are presumed to retain their pretreatment size and shape in the case of MtFs. To overcome this hardship and mainly to shift up the baseline $F0$ range, some MtFs take surgical treatment. Although the $F0$ range is certainly raised in the new voice, as far as the second author knows, it is a pity that the naturalness is decreased in the new voice instead. Further, many clinical papers and engineering papers on speech synthesis claim that raising the $F0$ range alone does not produce good femininity [13–15]. Since shape of the vocal tract has a strong effect on the voice quality, good control of the articulators has to be learned to achieve good femininity. Considering small effects of the hormone treatment and the surgical treatment on MtF clients, we can say that the most effective and least risky method to obtain good femininity is taking voice training from speech therapists or pathologists with good knowledge of GID.

2 Why Femininity Estimator?

In the typical therapy conducted by the second author, the following three methods are used based on [16]. 1) raising the baseline $F0$ range, 2) changing the baseline voice quality, and 3) enhancing $F0$ dynamics to produce an exaggerated intonation pattern. One of the most difficult things in the voice therapy lies not on a client's side but on a therapist's side, i.e., accurate and objective evaluation of the client's voice. It is often said that as synthetic speech samples are presented repeatedly, even expert speech engineers tend to perceive better naturalness in the samples, known as habituation effect. This is the case with good therapists. To avoid this effect and evaluate the femininity unbiasedly, listening tests with novice listeners are desirable. But the tests take a long time and a large cost because a new test has to be done whenever some acoustic change happens in the client's voice through the therapy.

Further, in most of the cases, GID clients are very eager to know how they sound to novice listeners, not experts. Some clients, not so many, claim that they sound feminine enough although they sound less feminine to anybody else. The objective evaluation of their voices is very effective to let these clients know the truth. For these two reasons, in this study, a listening test simulator was developed by automatically estimating the femininity which novice listeners would perceive if they heard the samples.

Among the above three methods, the first two ones focus on static acoustic properties and the last one deals with dynamic $F0$ control. The dynamic control of $F0$ for various speaking styles is a very challenging task in speech synthesis research and, therefore, we only focused on the femininity controlled by the $F0$ range and the voice quality. In medical and educational applications, unlike entertainment applications, technologies should not be used easily if they are not mature enough. In this work, good discussion was done in advance about what should be done by machines and what should be done by humans in the therapy. Only when there are some things difficult for humans and easy for machines, then, those things should be treated by machines.

GMM modeling of *F0* values and that of spectrums were done separately for biologically male speakers and female ones. By using the four models, the estimator was developed. In addition to the experimental results of the femininity estimation, some merits and demerits of using the estimator in actual voice therapy are described.

3 GMM-based Modeling of Femininity

3.1 Modeling Femininity with Isolated Vowel Utterances

Questions of acoustic cues of good femininity were often raised in previous studies [17–21]. Acoustic and perceptual analysis of speech samples of biologically male and female speakers and those of MtF ones were done and the findings lead to the three kinds of methods in the previous section. About the voice quality, as far as the authors know, all the studies focused on isolated vowel utterances and formant frequencies were extracted to estimate the femininity. It is true that, even from a single /a/ utterance, it is possible to estimate vocal tract length [22] and then, the femininity. It is also true, however, that even if a client can produce very feminine isolated vowels with careful articulation, it does not necessarily mean that the client can produce continuous speech with good femininity. This is the case with foreign language pronunciation. Even if a learner can produce very good isolated vowels, the learner is not always a good speaker of the target language in normal speech communication. This is partly because good control of speech dynamics including prosody is required in continuous speech. We can consider another reason that so much attention cannot be paid to every step of producing vowels in a sentence. We can say that the desired tool for MtF voice therapy is an estimator of the femininity from naturally-spoken continuous speech. In this case, we have a fundamental problem. With the analysis methods used in the previous studies, it is difficult to estimate the femininity from continuous speech because formant frequencies change not only due to the femininity but also due to phonemic contexts of the target vowels.

3.2 Modeling Femininity with Continuous Speech

This problem can be solved by using GMM-based speaker recognition/verification techniques. In continuous speech, various phonemes are found and the phonemes naturally cause spectral changes. If the utterance has sufficient spectral variations, averaging the spectrum slices over time can effectively cancel the spectral changes caused by the phonemic variation. The resulting average pattern of spectrum comes to have a statically biased form of spectrum, which is considered to characterize the speaker identity and the stationary channel. In GMM-based speaker recognition/verification, the average pattern is modeled not as a single spectrum slice but as a mixture of Gaussian distributions, where the spectrum is often represented as cepstrum vector.

What kind of phone does the averaged spectrum correspond to? If a continuous speech includes vowels only, it is possible to give a clear phonetic interpretation of the averaged spectrum [23]. Figure 1 shows the vowel chart of American English, where the 10 monophthongs are plotted. This figure clearly shows the articulatory center (average) of the vowels corresponds to /ə/, schwa sound. The other figure in Figure 1 is a result of MDS (Multi-Dimensional Scaling) analysis of vowel samples from an American female speaker. Here, a single Gaussian distribution was used to model each vowel acoustically and a distance matrix was calculated from the vowel examples using Bhattacharyya distance measure. In the figure, we can say that the acoustic center of the vowels also corresponds to /ə/. It is known that schwa is generated with a sound tube of a uniform cross-sectional area, which implies that schwa is produced with the least articulatory effort. In continuous speech, most of the unstressed vowels are reduced to be schwa sounds, meaning that the schwa is the most frequent sound observed in naturally-spoken sentences. The averaging operation not only can cancel the spectral changes caused by the phonemic variation but also can represent the acoustic quality of the most frequent phone (vowel) of that speaker.

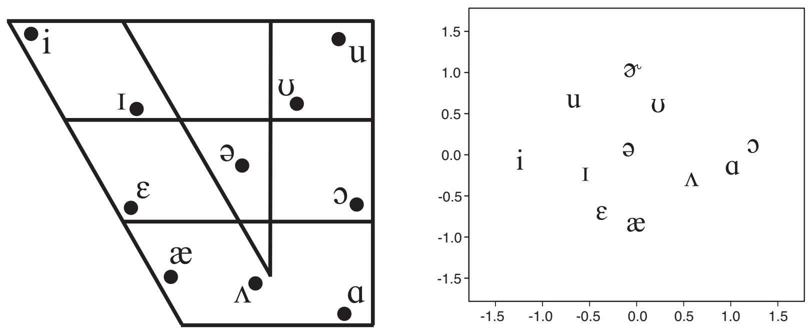

Fig. 1. The vowel chart of American English and an MDS chart of its vowel examples

With speech samples of any text spoken by a large number of female speakers, a GMM was trained to characterize the spectrum-based femininity, M^s_F. With male speakers, a GMM for the masculinity was trained, M^s_M. Using both models, the eventual spectrum-based femininity for a given cepstrum vector o, $F^s(o)$, was defined as the following formula [24, 25];

$$F^s(o) = \log \mathcal{L}(o|M^s_F) - \log \mathcal{L}(o|M^s_M). \tag{1}$$

Similar models are trained for the F_0-based femininity and masculinity;

$$F^f(o) = \log \mathcal{L}(o|M_F^f) - \log \mathcal{L}(o|M_M^f). \tag{2}$$

Integration of the four models, M_F^s, M_M^s, M_F^f, and M_M^f can be done through generalizing the above formulae by linear regression analysis.

$$\begin{aligned} F(o) = \alpha \log \mathcal{L}(o|M_F^s) + \beta \log \mathcal{L}(o|M_M^s) + \\ \gamma \log \mathcal{L}(o|M_F^f) + \varepsilon \log \mathcal{L}(o|M_M^f) + C, \end{aligned} \tag{3}$$

where α, β, γ, ε, and C are calculated so that the $F(o)$ can predict perceptual femininity (PF) of o the best. The PF scores were obtained in advance through listening tests with novice listeners.

4 Femininity Labeling of MtF Speech Corpus

4.1 MtF Speech Corpus

A speech corpus of 111 Japanese MtF speakers was built, some of whom sounded very feminine and others sounded less feminine and needed additional therapy. Each speaker read the beginning two sentences of "Jack and the beanstalk" with natural speaking rate and produced isolated Japanese vowels of /a, i, u, e, o/. The two sentences had 39 words. All the speech samples were recorded and digitized with 16 bit and 16 kHz AD conversion. Some clients joined the recording twice; before and after the voice therapy. Then, the total number of recordings was 142. For reference, 17 biologically female Japanese read the same sentences and produced the vowels.

4.2 Perceptual Femininity Labeling of the Corpus

All the sentence utterances were randomly presented to 6 male and 3 female adult Japanese listeners through headphones. All of them were in their 20s with normal hearing and they were very unfamiliar with GID. The listeners were asked to rate subjectively how feminine each utterance sounded and write down a score using a 7-degree scale, where $+3$ corresponded to the most feminine and -3 did to the most masculine. Some speech samples of biological female speakers were used as dummy samples. Every rater joined the test twice and 18 femininity judgments were obtained for each utterance. Figure 2 shows histogram of the averaged PF scores for the individual MtF utterances. Although some utterances still sounded rather masculine, a good variance of PF was found in the corpus. The averaged PF of biological female speakers was 2.74. While, in Figure 2, the averaging operation was done over all the raters, in the following section, it will be done dependently on the rater's biological gender.

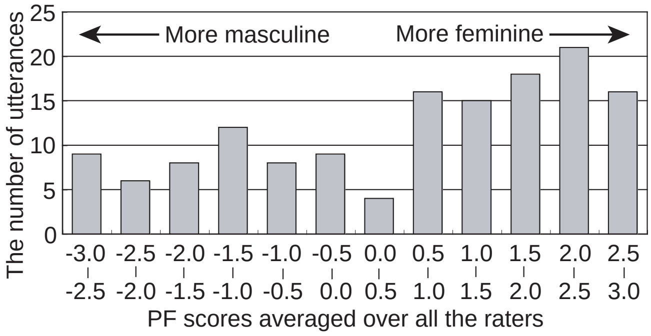

Fig. 2. Histogram of the averaged PF scores of the 142 utterances

4.3 Intra-rater and Inter-rater Judgment Agreement

Agreement of the judgments within a rater was examined. Every rater joined the test twice and the correlation between the two sessions was calculated for each. The averaged correlation over the raters was 0.80, ranging from 0.48 to 0.91. If the rater with the lowest correlation can be ignored, the average was recalculated as 0.84 (0.79 to 0.91).

PF scores by a rater were defined as the scores averaged over the two sessions. Using these scores, the judgement agreement between two raters was analyzed. The agreement between a female and another female was averaged to be 0.76, ranging from 0.71 to 0.83. In the case of the male raters, the agreement was averaged to be 0.75, ranging from 0.59 to 0.89. The agreement between a female and a male was 0.71 on average, ranging from 0.60 to 0.79. Some strategic differences in the judgment may be found between the two sexes.

PF scores by the female were defined as the averaged scores over the three female raters. Similarly, *PF scores by the male* were defined. The correlation between the two sexes was 0.87, which is very high compared to the averaged inter-rater correlation between the two sexes (0.71). This is because of the double averaging operations, which could reduce inevitable variations in the judgments effectively.

Now, we have 12 different kinds of PF scores; 9 from the 9 raters, 2 as the scores by the male and the female, and the other one obtained by averaging the male score and the female one. In the following sections, the correlations between the original PF scores and the automatically estimated PF scores, defined in Section 3.2, are investigated.

5 Training of M_F^s, M_M^s, M_F^f and M_M^f

As described in Section 3.2, automatic estimation of the femininity is examined based on GMM-based modeling. As speech samples for training, JNAS (Japanese

Newspaper Article Sentences) speech database, 114 biological males and 114 biological females, was used [26]. The number of sentence utterances was 3,420 for each sex. Table 1 shows acoustic conditions used in the analysis. For the spectrum-based GMMs of M_F^s and M_M^s, silence removal was carried out from the speech files and 12 dimensional MFCCs with their Δs and $\Delta\Delta$s of the remaining speech segments were calculated. For the $F0$-based GMMs of M_F^f and M_M^f, log$F0$ values were utilized.

Table 1. Acoustic conditions of the analysis

sampling	16bit / 16kHz
window	25 ms length and 10 ms shift
parameters	MFCC with its Δ and $\Delta\Delta$ for M_F^s and M_M^s
	log$F0$ for M_F^f and M_M^f
GMM	mixture of 16 Gaussian distributions

6 Automatic Estimation of Femininity

6.1 Simple Estimation Based on F^s and F^f

For each of the 142 MtF utterances, their femininity scores were estimated using F^s and F^f. The estimated scores were compared with the 12 different PF scores and the correlation was calculated separately. The averaged correlation over the first 9 PF scores is 0.64 for F^s and 0.66 for F^f. Table 2 shows the correlation of F^s and F^f with the other three PF scores; the male and the female scores and their averaged one. While the female PF is more highly correlated with F^s than F^f, the male PF is more highly correlated with F^f than F^s. This may imply different strategies of judging the femininity between the male raters and the female ones. It seems that the male tend to perceive the femininity more in high pitch of the voice. This finding is accordant with the results obtained in a previous study done by the second author [27]. In the study, it was shown that male listeners tend to assign higher femininity scores to speech samples with higher pitch.

Table 2. Correlation of F^s and F^f with the three PF scores

	female PF	male PF	averaged PF
F^s	0.71	0.70	0.73
F^f	0.67	0.76	0.74

6.2 Integrated Estimation with Weighting Factors

Linear regression analysis was done to predict the 12 PF scores using the four models, where the PF scores were converted to have a range from 0 (the most masculine) to 100 (the most feminine). As shown in Equation 3, four weighting factors and one constant term were calculated to minimize the prediction error.

The averaged multiple correlation over the first 9 PF scores was increased up to 0.76. Table 3 shows the multiple correlation coefficients with the other three PF scores. Figure 3 graphically shows the correlation with the female and male scores. Here, for utterance(s) of an MtF speaker, the weighting factors were calculated using utterances of the other MtF speakers and then, the femininity score(s) of utterance(s) of that MtF speaker were estimated. Namely, the estimation was done in a speaker-open mode. Considering magnitude of the intra- and inter-rater correlations of PF, we can say that $F(o)$ is a very good estimator of PF.

Table 3. Correlation of F with the three PF scores

	female PF	male PF	averaged PF
F	0.78	0.86	0.84

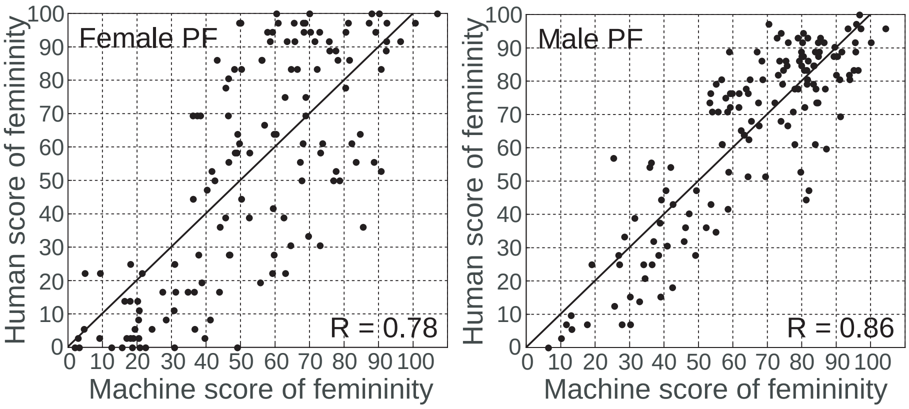

Fig. 3. Correlation between the original and the estimated PF scores

6.3 Discussions

The four weighting factors and the constant term in Equation 3 show different values for the 12 PF scores. The difference in values between two raters characterizes the difference in judging strategies between them. Table 4 shows 12 patterns of α, β, γ, ε, and C with the multiple correlation coefficient (R). As was found in Section 6.1, clear difference was found between the female and the male raters. The female tend to focus on spectral properties (α=0.125 and 0.068 for female PF and male PF), while the male tend to focus on pitch (γ=0.107 and 0.144 for female PF and male PF). In this sense, F3's judgment is very male because she emphasized pitch (γ=0.167) and de-emphasized spectral properties (α=0.085). In Table 3 and Figure 3, the multiple correlation was not so high for the female PF scores (R=0.78) and this can be considered probably because of F3.

For every PF score, the absolute value of α and that of β are similar. α and β of the female PF are 0.125 and -0.127. Those of the male PF are 0.068 and -0.064. This directly means that, with spectral properties of the voice, it is as important to shift the client's voice closer to the female region as to shift the voice away from the male region. On the contrary, the absolute value of γ and that of ε show a large difference. γ and ε of the female PF are 0.107 and -0.013. Those of the male PF are 0.144 and 0.009. In every case, ε takes a very small value, near to zero, compared to γ. This indicates that, as for *F0*, although it is important to shift the voice into the female region, it matters very little if the voice is still located in the male region. This asymmetric effects of spectrum and *F0* can be summarized as follows by using terms of bonus and penalty. If the voice is closer to the female region, larger bonus is given and if the voice is closer to the male region, larger penalty is given. Although both bonus and penalty should be considered with spectral properties of the voice, only bonus is good enough with its *F0* properties.

Table 4. Values of the weighting factors and the constant term

	α M_F^s	β M_M^s	γ M_F^f	ε M_M^f	C	R
F1	0.154	-0.143	0.068	-0.017	1.301	0.77
F2	0.133	-0.126	0.086	-0.012	1.005	0.68
F3	0.087	-0.112	0.167	-0.011	-0.742	0.72
M1	0.074	-0.069	0.120	-0.009	1.091	0.82
M2	0.035	-0.024	0.176	-0.030	1.479	0.73
M3	0.076	-0.085	0.150	0.028	0.446	0.80
M4	0.034	-0.035	0.172	-0.005	0.941	0.76
M5	0.090	-0.084	0.141	0.023	1.237	0.86
M6	0.100	-0.090	0.107	0.047	1.389	0.70
female PF	0.125	-0.127	0.107	-0.013	0.521	0.78
male PF	0.068	-0.064	0.144	0.009	1.097	0.86
average PF	0.097	-0.096	0.126	-0.002	0.809	0.84

7 Actual Use of the Estimator in Voice Therapy – Merits and Demerits –

The second author has used the estimator in her voice therapy for MtF clients since Feb. 2006. Figure 4 shows an actual scene of using the estimator in the therapy and Figure 5 is the interface of the estimator. It was found that, when biologically male speakers without any special training pretended to be female, it was very difficult to get a score higher than 80. However, it is very interesting that good MtF speakers, who can change their speaking mode voluntarily from male to female, could have the estimator show a very low score (very masculine) and a very high score (very feminine) at their will. Since the estimator is focusing on only static acoustic properties, we consider that these MtFs have two baseline shapes of the vocal tract, which may be realized by different positioning of the tongue, and two baseline ranges of F_0. In this sense, the estimator helped the clients a great deal who were seeking for another baseline of the vocal tract shape and/or that of F_0 range through try-and-errors. Needless to say, quantitative and objective evaluation of their trials motivated the clients very well.

Fig. 4. An actual scene of the voice therapy with the estimator

Only the focus on static acoustic properties naturally caused some problems. As described in Section 1, by producing a rather exaggerated pattern of intonation, listeners tend to perceive higher femininity. Although this exaggeration is a good technique to obtain high PF in the voice, the estimator completely ignores this aspect and then, some clients received unexpectedly high scores or low scores. In actual therapies, the therapist has to use the machine by carefully observing what kind of strategy the client is trying to use. If speech dynam-

Fig. 5. Interface of the estimator

ics is effectively controlled, then, the therapist should not use the machine but give adequate instructions only based on the therapist's ears. Further, especially when the clients evaluate their voices by themselves, what is possible and what is impossible by the machine should be correctly instructed to them.

8 Conclusions

This work described the development of an automatic estimator of the perceptual femininity from continuous speech using speaker verification techniques. Spectrum-based and F_0-based GMMs were separately trained with biological male and female speakers. By integrating these models, the estimator was built. The correlation of the estimated values and the perceptual femininity scores obtained through listening tests was 0.86, comparable to the intra- and inter-rater correlation. Some analyses were done about sexual differences of the femininity judgment and some strategic differences in the use of spectrum-based cues and F_0-based cues were shown. It was indicated that the male tend to give higher scores to the voices with higher pitch. Further, it was shown independently of the rater's sex that the penalty of the F_0 range being still in the male region is remarkably small. As future work, we are planning to take MRI pictures of a good MtF speaker's control of the articulators when producing feminine vowels and masculine ones. We hope that the estimator will help many MtF clients improve the quality of their lives.

References

1. H. Levitt, "Processing of speech signals for physical and sensory disabilities," Proc. National Academy of Sciences, 92(22), 9999–10006, Oct. 1995.
2. F. Ehsani and E. Knodt, "Speech technology in computer-aided language learning: strategies and limitations of a new CALL paradigm," Language Learning & Technology, 2(1), 45–60, July 1998.
3. P. C. Loizou, "Mimicking the human ear," IEEE Signal Process Magazine, 15, 101–130, Sep. 1998.
4. P. C. Loizou, "Signal-processing techniques for cochlear implants – a review of progress in deriving electrical stimuli from the speech signal," IEEE Engineering in Medicine and Biology Magazine, 18(3), 34–46, May/June 1999.

5. Y. Suhail and K. G. Oweiss, "Augmenting information channels in hearing aids and cochlear implants under adverse conditions," Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 889–892, May 2006.
6. H. L. Barney, F. E. Haworth, and H. K. Dunn, "An experimental transistorized artificial larynx," Bell System Technical Journal, 38, 1337–1356, 1959.
7. K. M. Houston, R. E. Hillman, J. B. Kobler, and G. S. Meltzner, "Development of sound source components for a new electrolarynx speech prosthesis," Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 2347–2350, March 1999.
8. G. S. Meltzner, J. B. Kobler, and R. E. Hillman, "Measuring the neck frequency response function of laryngectomy patients: implications for the design of electrolarynx devices," J. Acoustic Society of America, 114(2), 1035–1047, Aug. 2003.
9. H. Mori, H. Otawa, T. Ono, Y. Ito, and H. Kasuya, "Internet-based acoustic voice evaluation system for screening of laryngeal cancer," J. Acoustic Society of Japan, 62(3), 193–198, 2006 (in Japanese).
10. J. D. Jong and J. Bernstein, "Relating phonepass scores overall scores to the council of europe framework level descriptors," Proc. European Conf. Speech Communication and Technology, 2803–2806, Sept. 2001.
11. M. T. Edgerton, "The surgical treatment of male transsexuals," Clinics in Plastic Surgery, 1(2), 285–323, April 1974.
12. F. G. Wolfort and R. G. Parry, "Laryngeal chondroplasty for appearance," Plastic and reconstructive surgery, 56(4), 371–374, Oct. 1975.
13. R. C. Bralley, G. L. Bull, C. H. Gore, and M. T. Edgerton, "Evaluation of vocal pitch in male transsexuals," J. Communication Disorder, 11, 443–449, 1978.
14. L. E. Spencer, "Speech characteristics of MtF transsexuals: a perceptual and acoustic study," Folia phoniat., 40, 31–42, 1988.
15. K. H. Mount and S. J. Salmon, "Changing the vocal characteristics of a postoperative transsexual patient: a longitudinal study," J. Communication Disorder, 21, 229–238, 1988.
16. M. P. Gelfer, "Voice therapy for the male-to-female transgendered client," American J. Speech-Language Pathology, 8, 201–2008 (1999)
17. H. Sato, "Acoustic cues of female voice quality," IECE Transaction, 57(1), 23–30, 1974 (in Japanese)
18. S. Bennett, "Acoustic correlates of perceived sexual identity in preadolescent children's voices," J. Acoustic Society of America, 66(4), 989–1000, 1979.
19. M. L. Andrews and C. P. Schmidt, "Gender presentation: perceptual and acoustical analyses of voice," J. Voice, 11(3), 307–313, 1997.
20. M. P. Gelfer and K. J. Schofield, "Comparison of acoustic and perceptual measures of voice in MtF transsexuals perceived as female vs. those perceived as male," J. Voice, 14(1), 22–33, 2000.
21. V. I. Wolfe, "Intonation and fundamental frequency in MtF TS," J. Speech Hearing Disorders, 55, 43–50, 1990.
22. A. Paige and V. W. Zue, "Calculation of vocal tract length," IEEE Trans. on Audio and Electroacoustics , AU-18(3), 268–270, 1970.
23. N. Minematsu, S. Asakawa, and K. Hirose, "Para-linguistic information represented as distortion of the acoustic universal structure in speech," Proc. IEEE Int. Conf. Acousitcs, Speech, and Signal Processing, 1, 261–264, May 2006.
24. A. E. Rosenberg and S. Parthasarathy, "Speaker background models for connected digit password speaker verification," Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 81–84, May 1996.

25. L. P. Heck and M. Weintraub, "Handset-dependent background models for robust text-independent speaker recognition," Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 1071–1074, April 1997.
26. http://www.mibel.cs.tsukuba.ac.jp/jnas/
27. K. Sakuraba, S. Imaizumi, K. Hirose, and K. Kakehi, "Sexual difference between male and female listeners in the perceptual test with the voice produced by MtF transsexuals," Proc. Spring Meeting of Acoustic Society of Japan, 2-P-4, 337–338, March 2004 (in Japanese).