Acoustic and perceptual implications of the transsexual voice.

by Deborah Gunzburger

INTRODUCTION

Speech therapists who counsel transsexuals often report that voice characteristics are resistant to convincing change, especially in the case of male-to-female transsexuals, in whom hormone therapy does not have a pitch-raising effect. The need for voice change and adaption of speech habits has obviously been recognized by transsexuals themselves, and many of them make spontaneous efforts to alter their manner of speaking (Edgerton, 1974). Coleman (1983) remarked that although it seems that simply raising the voice pitch to a level appropriate for female speakers would be effective, it turns out that a distinct male voice quality often persists in spite of such efforts.

There is a proliferating interest in, and therefore, an increasing amount of literature on general gender differences in speech (e.g., Tanner, 1990) and more specific issues of the male-female voice distinction (see Karlsson, 1992, or Tielen, 1992, for recent publications in this area). Extensive assessment of the differences between "normal" male and female speakers may very well lead to better voice adjustment strategies and therapies for transsexual persons.

Assessment of differences must necessarily encompass biological as well as sociocultural aspects of speech, a realm with which I initially deal. I then describe speech production data of a limited number of transsexuals and finally a small-scale perceptual evaluation of the production data analyzed.

BIOLOGICAL VERSUS SOCIOCULTURAL ASPECTS OF VOICE AND SPEECH

Two main differences between male and female voices are well documented and can be explained on anatomical-physiological grounds. The first and, in perceptual terms, most important difference can be explained by the fact that men's vocal cords are larger and thicker than women's. As a consequence, the fundamental frequency or pitch, is lower in a man's voice. Second, because of men's overall larger vocal tract, resonant frequencies of the various cavities are lower. These resonant frequencies are called formants and are mainly distinguishable in vocalic portions of speech. Biologically based voice and speech differences are secondary sex characteristics and are caused by major hormonal influences during puberty.

However, various speech characteristics cannot be explained by biological factors. Due to the influence of cultural patterns, social pressure, and mass media, certain vocal images develop that are shared by groups of people. In terms of male-female differences researchers have found evidence that sex-dependent vocal and articulatory habits take root at an age well before puberty (Gunzburger et al., 1987; see also Sachs et al., 1973; and Meditch, 1975).

Differences in speech habits between men and women on a segmental level have been investigated in a number of studies. We mention only a few: Labov (1966) found a clear difference in the pronunciation of the voiceless fricative / [Theta]/ (as in thin); women pronounced the sound in a "correct" way, whereas men often replaced it by another sound, such as the stop consonant in "tin." The same pattern was found with the voiced dental fricative (as in "this"). These findings were corroborated by Anshen (1969) and Wolfram (1969) for the United States and by Milroy (1980) for Northern Ireland. In addition, the postvocalic /r/ in words like "far" and "fare" was pronounced more frequently by women than by men in the United States. Fischer (1958) found that girls in New England pronounced the progressive verb ending "-ing" more frequently in the standard way than boys do, who instead produced /in/ more often. With respect to vowel sounds, the pattern of a more standard pronunciation by women as compared to men is repeated as shown by a number of studies, see Smith (1985).

Apart from and independent of pronunciation (and of the frequently mentioned lexical and stylistic features), a number of other vocal features can be modified - either consciously or unconsciously - during an utterance. Key (1975) provided cross-cultural data on sex-associated paralinguistic features, including, for example, the Mexican Mazateco "whistle speech" which is realized almost exclusively by men. Women in this community pretend not to understand this communication system based on whistles of varying pitch and duration. On a more mundane level we might mention Scandinavian countries where women may express their agreement by means of an ingressively articulated "ja," whereas men do not.

The best investigated sex-associated prosodic parameter that does not depend strictly on anatomic differences is undoubtedly pitch range. Various studies have shown that the standard deviation of women's F0 from the female mean is much greater than is the case for men. Moreover, women's pitch changes seem to have a sharper gradient over time than men's do. In other words, there is general agreement that women's speech shows more intonational dynamics. For an extensive bibliography on this subject matter see Thorne and Henley (1975) and Thorne et al. (1983).

On the perceptual level, different vocal features are used for male and female speakers in a personality attribution task. Addington (1968) and Aronovitch

(1976) provided data from which it can be concluded that judgments on masculinity- and femininity-related scales of male speakers were based on the range of intensity and pitch, whereas judgments on female speakers correlated with absolute intensity and temporal rate of fluency. Those men who spoke more monotonously were associated with masculinity, particularly in the Addington (1968) study. Women who spoke more slowly, quietly, disfluently and with a relatively high pitch were judged to be more feminine.

The first experiment described here deals with articulatory-acoustic parameters. My purpose was to obtain descriptive data of some possibly systematic changes of voice and speaking characteristics as a function of changed sex and gender identity of male-to-female transsexuals. In this experimental situation subjects on the one hand had obvious anatomical constraints as to their vocal cords and vocal tracts, but on the other hand tried to intuitively realize maximal differences in acquired speech behavior. Analysis is focused on intraindividual comparison.

SPEECH PRODUCTION EXPERIMENT

Speakers, Stimulus Material, Recordings

Speakers were invited to participate in the experiment at an informal meeting for transsexuals organized by the Dutch Association for Sexual Reform. Speech samples of six speakers were suitable for further acoustic analysis. (Four speakers dropped out when it actually came to the point of recording them in the female speaking mode.) Speakers' ages ranged from 22 to 59 years; all had received hormone therapy for at least 18 months and been living in their new gender role for 1.5 to 10 years. Speaker 1 and Speaker 2 had undergone transsexual surgery, and for Speaker 2 this was combined with surgical laryngeal modification. Some of the speakers had been seeking help from a speech therapist and all admitted to having made a conscious effort to alter their prefemale way of speaking without being able to state exactly what their alterations consisted of.

The stimulus material consisted of a list of 56 "ordinary" Dutch words. In addition, these words were also combined into a coherent and, as regards content, neutral piece of running prose.

Subjects were taperecorded individually. They had a general idea that the research involved sex-dependent differential voice and speaking habits and they knew what their task would consist of in the experiment. They received financial compensation for participation.

Every session started with a - sometimes lengthy - piece of casual conversation so as to make the speaker feel at ease and ample time was allocated to get acquainted with the stimulus material to be read. The actual

recordings were made in a sound-treated booth using high-quality recording equipment. After having read the text and the isolated words in the female manner, the same stimulus material was to be read in the former, male way. No time pressure was exerted and these consecutive recordings were made only after subjects had had ample practice time and declared themselves to be prepared for the task.

Analysis

Isolated words and phrases were analyzed separately. As to the latter, the text was subdivided at syntactically natural points into 25 phrases with an average length of 7.4 words. The purpose of the analysis was to gain insight into possible differences between the male and female realizations of durational aspects, pitch and pitch range, loudness and loudness range, and various formant characteristics. The exact nature of the acoustic analysis and all the resulting parameters are described and discussed in detail elsewhere (viz., for isolated words, Gunzburger, 1989; for phrases, Gunzburger, 1993). Here, we restrict ourselves to presenting the most relevant data in terms of being interpretable for the phonetically untrained reader.

Results and Discussion

Results of the acoustic measurements were checked as to their statistic significance by means of a paired t test. Table I shows these values.

We draw attention to the following points:

Mean duration of isolated words is for all but one speaker and for the pooled values significantly longer in the female version; pooled (and three of the six individual) mean phrase duration values are significantly higher in the female version. The absence of data in the literature about durational aspects of the male-female speech distinction is conspicuous. To the best of our knowledge the only investigation that attempted to deal with temporal cues on a suprasegmental level gives some data on utterance rate in terms of words per minute (Terrango, 1966). It appeared that male speakers who were judged to exhibit effeminate speech had a lower speaking rate than speakers who were judged to use "normal" masculine speech (185 words/min as compared to 194 words/min). Subjects in the current study frankly admitted having made some extra effort to read in the female mode, which might have resulted in a greater amount of overall utterance time in some cases. Note the contrast of these data with the popular belief of a higher female speaking rate.

As to pitch, in the isolated word condition all but one subject use a higher F0 in the female version. Speaker 2, who had undergone surgical vocal cord construction, realizes an extremely high F0 in the female mode (309 Hz!), which sounds meager, unnatural, and falsetto-like. In the phrase condition,

four subjects used a significantly higher pitch for the female version. Speaker 5 had the same mean F0 value for the male and female version; pooled data show a significantly higher value for the female version. An increase in F0 is the most obvious parameter to adapt to achieve a changed gender-dependent phonation pattern. Such increased vocal cord tension might be the indirect result of a continuous overall shift of the tongue towards a higher front position (Fant, 1968).

[TABULAR DATA FOR TABLE I OMITTED]

Pitch range in the isolated word condition is significantly greater for two speakers and pooled data on the female speaking mode. Values of the other speakers show a tendency in the same direction. For phrases, where of course the notion of pitch range has inherently more importance, all but one speaker's values reach the level of significance. As mentioned in "Biological Versus Sociocultural Aspects of Voice and Speech," these data corroborate earlier findings that intonational dynamism is typical of female speech. Our data bear out McConnell-Ginet's (1975) claim that since both actual and perceived femaleness correlate with changing fundamental frequency, i.e., nonmonotonicity, rapid pitch shifts, and especially a wide pitch range are the primary characteristics in mimicry of feminine speech by male speakers.

In addition to different intonational characteristics, subjects also have clearly adopted other prosodic habits as regards loudness level and loudness range to make their speech match their changed gender role. Measurements of isolated words show that, with the exception of Speaker 2 (who had undergone vocal surgery), all subjects speak at a lower intensity level in the female version; differences reach a level of significance in three speakers and for pooled data. For phrases, the same picture emerges. This strategy makes sense in the light of intuitive perceptual notions of soft and gentle voices being clearly associated with feminine stereotypes such as tenderness, affection, and submissiveness and loud and strong voices conveying masculine stereotypes such as ambition, strength, and dominance. The wider intensity range, generally associated with male speaking characteristics (e.g., McConnell-Ginet, 1983) is positively enhanced by the current data: For isolated words all male versions and for phrases all but one male version have a wider intensity range than the female versions; however, not all of these differences reach the level of significance.

Values of the first and second formant location, bandwidth and their respective standard deviations fail to indicate a systematic relationship between male and female realization. Central frequency of F3, however, is systematically higher in the female version. Although, as stated by Fant (1960), quantitative expressions for the relative role of any particular part of the vocal tract as a determinant of the formants has to be specified per vowel, or per group of vowels, a general pattern is worth mentioning in this context:

A decreased mouth cavity length results in overall increased F3 values. This systematic upward shift in the third formant is the more interesting, given the obvious anatomical constraints of the subjects (viz., a male vocal tract in terms of dimensions) and the fact that any average speaker is totally unconscious of his or her formant frequencies, let alone able to change them voluntarily. The first two formants are mainly responsible for the phonetic quality of the segments, in this case vowels, whereas the third and higher formants primarily influence the timbre of a voice. Fant (1960) mentioned average F3 as one of the keys to identifying speaker type. An admittedly very speculative but, in the context of this highly particular population, attractive attempt at venturing some articulatory explanation is the following: By decreasing the degree of lip-rounding, the mouth cavity becomes shorter, and this results, as mentioned previously, in an increased F3 value. Differential use of facial expressions during speech, including greater retraction of the mouth corners, has been considered characteristic of female speakers in a cross-cultural context (Ohala, 1984). Another possible explanation can be found in the literature on singing (e.g., Sundberg, 1974, 1975): The length of the vocal tract can be altered by raising or lowering the larynx - an effect known to differentiate between female trained and untrained singers.

PERCEPTION TEST

Method

Twenty five male and 25 female phrase utterances of Speaker 1 and Speaker 4 were used for perceptual evaluation. Speaker 1 conforms to the global pattern of generally accepted and - in our acoustic measurements confirmed - differential voice characteristics of higher F0 and lower intensity values for female speakers, whereas Speaker 4 came up with an insignificant F0 difference and atypical intensity values. Since for practical reasons the number of speakers used for perceptual evaluation had to be limited anyway, the selection of these two speakers seems justified in the light of possible perceptual repercussions of their differential acoustic data.

Utterances were presented pairwise to a total number of 31 (17 male and 14 female) naive listeners, who were between 18 and 20 years of age and had no self-reported hearing impairment. Listeners' task consisted of sex identification, to be indicated on an answer sheet. They were not aware of being asked to assess transsexuals.

Results

As concerns Speaker 1, 10 responses of a total of 775 (25 items x 31 listeners) were incorrect, which corresponds to 1.3%. (The term "incorrect" is applied to an item that was scored male-like while uttered female-like and vice versa.) With Speaker 4, 200 items, or 25.8%, were scored incorrectly.

The cogency of Speaker 1's scores hardly needs any comment; statistically Speaker 4's results also reach the level of significance: Since responses are either correct or incorrect, they are considered to be binomially distributed. The number of 200 incorrect scores does not fall within the range of random scores (360-415) and is therefore significant. An analysis of variance shows that scores on Speakers 1 and 4 differ significantly: $F(1, 24) = 87.1$, p [less than or equal to] 0.001. There is no significant effect of the factor item (utterance presented): $F(24, 24) = 1.04$, ns.

CONCLUSION

Production data of isolated words as well as of utterances on the phrase level provide evidence of the interesting fact that, in spite of the given anatomical constraints, but probably due to very high motivation, subjects are able to intuitively adopt a number of vocal characteristics that are known to add to a feminine voice quality. On the basis of this significant finding, professional speech therapists should concentrate on enhancing these characteristics to further develop their effectiveness. Surgical intervention can be considered an adjunct to voice therapy, but speech and voice therapy should be included in the rehabilitation of the transsexual and should also take care of preventing adoption of an effeminate male quality resorted to by some transsexuals instead of the female quality that is desired. In addition, care should be taken to prevent the possibility of vocal abuse in the new mode of phonation.

Of special interest is the fact that, whereas the first and second formant locations undergo no systematic change in the two speaking modes, there is a systematic upward shift in the central frequency of the third formant, which may be the result of consciously or unconsciously shortening the mouth cavity length. As mentioned previously, retracting the mouth corners ("the ever smiling female"?) shortens the mouth cavity and raises its resonances, signaling (on a global ethological - including human? - scale) smallness, nonthreatening attitude, goodwill of the receiver, etc., in short, a number of so-called stereotypical female characteristics.

REFERENCES

Addington, D. W. (1968). The relationship of selected vocal characteristics of personality perception. Speech Monogr. 35(4): 492-508.

Anshen, F. (1969). Speech variation among Negroes in a small Southern community. Unpublished doctoral dissertation, New York University.

Aronovitch, C. D. (1976). The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. J. Soc. Psychol. 99: 207-220.

Coleman, R. O. (1983). Acoustic correlates of speaker sex identification: Implications for the transsexual voice. J. Sex Res. 19: 293-306.

Edgerton, M. T. (1974). The surgical treatment of transsexuals. Clin. Plastic Surg. 1.

Fant, G. (1960). Acoustic Theory of Speech Production, Mouton, The Hague.

Fant, G. (1968). Analysis and synthesis of speech processes. In Malmberg, B. (ed.), Manual of Phonetics, North Holland, Amsterdam, pp. 173-277.

Fischer, J. L. (1958). Social influences on the choice of a linguistic variant. Word 14: 47-56.

Gunzburger, D. (1989). Voice adaptation by transsexuals. Clin. Linguist. Phonet. 3.2: 163-172.

Gunzburger, D. (1993). An acoustic analysis and some perceptual data concerning voice change in male-female transsexuals. Eur. J. Disorders of Communication 28.1: 13-21.

Gunzburger, D., Bresser, A., and ter Keurs, M. (1987). Voice identification of prepubertal boys and girls by normally sighted and visually handicapped subjects. Lang. Speech 30: 47-57.

Karlsson, I. (1992). Analysis and synthesis of different voices with emphasis on female speech. Doctoral dissertation, University of Stockholm, Sweden.

Key, M. R. (1975). Male/Female Language, Scarecrow Press, Metuchen, NJ.

Labov, W. (1966). The Social Stratification of English in New York City, Center for Applied Linguistics, Washington, DC.

McConnell-Ginet, S. (1975). Intonation in the social context: Language and sex. Paper presented at the Ninth International Congress in Sociology, Uppsala, Sweden.

McConnell-Ginet, S. (1983). Intonation in a man's world. In Thorne, B., Kramarae, C., and Henley, N. (eds.), Language, Gender and Society.

Meditch, A. (1975). The development of sex-specific speech patterns in young children. Anthropol. Linguist. 17(9): 421-465.

Milroy, L. (1980). Language and Social Networks, Basil Blackwell, Oxford.

Ohala, J. J. (1984). Ethological perspective on common cross-language

utilization of F0 of voice. Phonetica 41: 1-16.

Sachs, J., Lieberman, P., and Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. In Shuy, R. W., and Fasold, R. W. (eds.), Language Attitudes: Current Trends and Prospects, Georgetown University Press, Washington, DC.

Smith, P. N. (1985). Language, the Sexes and Society, Blackwell, Oxford.

Sundberg, J. (1974). Articulatory interpretation of the "singing formant." J. Acoust. Soc. Am. 55: 838-844.

Sundberg, J. (1975). Formant technique in a professional female singer. Acustica 32: 89-96.

Tanner, D. (1990). You Just Don't Understand. Women and Men in Conversation, Morrow, New York.

Terrango, L. (1966). Pitch and duration characteristics of the oral reading of males on a masculinity-femininity dimension. J. Speech Heating Res. 9: 590-595.

Tielen, M. T. J. (1992). Male and female speech. An experimental study of sex-related voice and pronunciation characteristics. Doctoral dissertation, University of Amsterdam, The Netherlands.

Thorne, B., and Henley, N. (eds.). (1975). Language and Sex, Newbury House, Rowley, MA. Thorne, B., Kramarae, C., and Henley, B. (eds.). (1983). Language, Gender and Society, Newbury House, Rowley, MA.

Wolfram, W. (1969). A Sociolinguistic Description of Detroit Negro Speech, Center for Applied Linguistics, Washington, DC.
