1. Introduction

In western society there are no absolute differences in language between men and women (men use variant x, women use variant y), only gradual ones (men use more x than y, women more y than x). This holds for all linguistic levels: syntactic, lexical, morphological, and phonological. Similarly, there is no rigid distinction between male and female voice qualities. Most of us are familiar with the sensation of hearing a person speak, and noticing voice characteristics that do not conform to our vocal stereotypes, e.g. a man with a high and shrill voice, or a woman with a deep and sonorous voice. Men can have voices with feminine characteristics, and women can have voices with masculine characteristics. There is overlap in voice characteristics between the sexes and there is variation within the groups of men and women. Some men have a more masculine (or feminine) voice than other men and some women have a more feminine (or masculine) voice than other women. Voice characteristics that are marked with regard to gender (e.g. high vs. low pitch) can be present to a greater or lesser extent. Together they make up a picture of a gendered voice.

The first determinant of gender-related voice quality differences is, of course, biological gender. A direct relationship exists between some acoustic /perceptual voice parameters and physiological characteristics of the speaker’s vocal organs. In general, women have shorter and thinner vocal cords than men, so the cords vibrate at higher speeds, causing higher pitch. There are also differences in form and dimensions of the vocal tract, causing differences in formant frequencies, affecting vocal timbre. However, other factors besides physiological factors may play a role. Pitch, formant frequencies, and voice quality in general are culturally variable (Van Bezooijen 1993); the voices of boys and girls differ before they develop sex-specifically in puberty (Graddol and Swann 1989). Notions about gender in society can influence the way people speak, as was shown by an experimental study of pitch differences between Japanese and Dutch female speakers (Van Bezooijen 1995).

The aim of the present research is to explain overlap in pitch between and
pitch variation within the sexes by operationalizing gender in two ways: sex (biological gender) and gender identity (social gender). Gender identity is defined as the extent to which cultural notions about masculinity and femininity are part of a person's identity (Willemsen and Fischer 1996).

In most research to date it is biological gender that has been related to differences in voice quality (Kraayeveld 1997; Tielen 1992), also because it is easy to operationalize. Overlap of voice characteristics between and variation of voice quality within the groups of men and women is explained by variables other than sex, such as age and status, or individual physiological differences. Little research, however, has been done that related gender identity to speech. One of the few exceptions is Smith (1985), who recorded eight subjects with widely varying gender identities. He then had listeners attribute gendered personality traits to the speakers, solely on the basis of their speech. The measured and attributed gender identities correlated substantially (between .65 and .81). However, Smith did not study specific voice parameters, only the general concept of 'voice'. The present research is an attempt to further specify the relation between gender identity and voice, by looking at a specific voice parameter, i.e. pitch.

2. Method

2.1 Recording procedure

Since we aimed at investigating the relation between stable personal characteristics and speech, the influence of situational factors on the behavior of the speakers had to be minimized. Therefore conversations were recorded between acquaintances and not between strangers. This reduced contextualised stereotypical gender behavior (especially relevant in male-female conversations). It also contributed to the informality of the setting. The conversations mostly took place at the house of one of the speakers. Separate microphones were used for the two speakers.

The conversation topics were given, to prevent people from talking about gender-specific topics, which might influence their voice quality. The topics ranged from deciding which of four patients was to receive a donor kidney to planning a holiday together. To ensure a real dialogue in stead of two separate monologues, the speakers were to reach a consensus opinion on all topics. After finishing the conversation, the speakers filled out questionnaires about gender identity, their relation to each other, and personal characteristics such as age and education. All participants were paid.
2.2 Speakers

Fifty-seven conversations were recorded, 19 between two women, 19 between two men, and 19 between a woman and a man. This makes a total of 114 speakers, 57 women and 57 men. The speakers were adults between 19 and 48 years old. In this age range, voices are not influenced by puberty or ageing (e.g. Hollien 1991). The speakers had an average to high education level in diverse fields (medical, economical, social, etc.). They came from all over the Netherlands, but with an emphasis on the area around Nijmegen. The two speakers in each dyad had known each other between one and five years on average.

2.3 Speech material

Initially there were 57 conversations of at least 31 minutes between two speakers. The conversations were recorded on Digital Audio Tape (Sony TCD-D7) with a sampling frequency of 48 kHz. The first minute of each conversation was deleted, because it contained a relatively large amount of 'start up' speech. From the next 30 minutes of the conversation one minute of speech for each speaker was selected in the following way. First, by means of a computer program written especially for this purpose, all speech was segmented into short utterances. The criterion for segmentation was a silent pause of at least 200 ms. All speech files were downsampled to 16 kHz. The program also indicated whether there was sound from only one microphone (one speaker) or two microphones (simultaneous speech from two speakers). Speech files with simultaneous speech were discarded. Next, all speech files with a duration ≥ 1 second and ≤ 5 seconds were selected. Relatively short (<1 second) and relatively long (>5 seconds) utterances were excluded. A random selection of speech files with a total duration of two minutes per speaker was subsequently checked auditorily for background noise, laughter, coughing, etc. From the resulting files 90 seconds were selected to be judged with respect to neutrality of content. On the basis of written-out texts, ten subjects (four women and six men aged 20 — 24) judged the content of the utterances on degree of masculinity and femininity. This was necessary because of future perception experiments with the same speech material. From the speech files which had received the highest neutrality ratings one minute of speech was compiled for each speaker for pitch analysis.
2.4 Biological gender variables (sex of speaker and sex of conversation partner) and social gender (gender identity)

There were two biological and one social gender variables. The first two were the sex of the speaker and the sex of the conversation partner. The latter was added because the difference between same-sex and mixed-sex dyads might influence the voice characteristics of the speakers. Social gender was represented by gender identity. Gender identity was measured with a questionnaire developed for the Netherlands in the early nineties (Willemsen and Fischer, 1996). It is important that the questionnaire be culture specific and not just a translation of e.g. an American questionnaire, as different cultures may have different notions of masculinity and femininity.

The gender identity questionnaire consists of two parts: a personality part with 30 questions about personality traits and a behavior part with 28 questions about appearances, interests, skills, social relations, and sex role behavior. Most questions have to be answered by indicating on a five-point scale the extent to which a person has a particular personality trait or how often he or she engages in a particular behavior. There are separate (but randomized) questions for feminine traits and behavior and masculine traits and behavior. All speakers filled out the questionnaire about themselves. In addition they filled out the traits part of the questionnaire about their conversation partner. Therefore, for each speaker six scores could be computed, on feminine traits, masculine traits, feminine behavior, and masculine behavior as judged by the speaker him or herself, and on feminine traits and masculine traits as judged by his or her conversation partner. Gender identity of the conversation partner was not used as an independent variable in order to keep the research design within manageable proportions.

2.5 Pitch variables: median, minimum, range (in semitones)

The acoustic analyses were done with PRAAT, a speech signal processing software package developed by Boersma and Weenink (1996). Pitch was determined by an autocorrelation method (Boersma 1993). As a first step six pitch measures were calculated: average pitch, median pitch, minimum pitch (lowest value measured, not including the bottom 2.5% because of possible measurement errors, e.g. octave jumps), maximum pitch (highest value, not including the top 2.5%), range (highest — lowest value, not including bottom and top 2.5%), and standard deviation. All values were computed in semitones (a musical scale, 12 st = 1 octave or a 2:1 ratio in frequency in Hz). The reference frequency was 16.35 Hz (C₀).
To reduce the number of pitch measures, Pearson product-moment correlations were computed between all measures. Average pitch and median pitch correlated .99. Median pitch was chosen because it is the more stable measure with regard to measurement errors. Median pitch correlated .83 with minimum pitch, .93 with maximum pitch, .06 with range, and .04 with standard deviation. Because range and standard deviation correlated fairly high (.89), and maximum pitch can be derived from minimum pitch plus range, median pitch was supplemented with minimum pitch and range.

3. Results

3.1 Sex of speaker

Not surprisingly, there is a significant difference in median pitch between men and women; women have a higher median pitch than men (t=24.73, df=112, p=.000). The mean value is 42.31 st (189 Hz) for women and 32.97 st (111 Hz) for men. There is no overlap in the distributions of the median pitch values for men and women. For men the values range from 28.31 st (84 Hz) to 39.12 st (157 Hz), for women from 39.26 st (158 Hz) to 44.90 st (219 Hz). The distributions can be found in Figure 1. The slight overlap between the values for men and women in this Figure is caused by the visual representation in a limited number of categories. The correlation between sex of speaker and median pitch is high: -.92 (Table 1). This means that as much as 84% of the variance in median pitch is explained by the sex of the speaker.

Table 1. Correlations between sex of speaker and the pitch measures. *=sign.01.

<table>
<thead>
<tr>
<th>sex of speaker</th>
<th>median pitch</th>
<th>minimum pitch</th>
<th>pitch range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.92*</td>
<td>-.84*</td>
<td>.05</td>
</tr>
</tbody>
</table>
Figure 1. Distribution of median pitch values

Figure 2. Distribution of minimum pitch values
The minimum pitch values for men and women also differ significantly; men have a lower minimum pitch than women ($t=16.43$, $df=112$, $p=.000$). The mean value for men is 27.07 st (80 Hz) and for women 36.85 st (139 Hz). The distributions of the minimum pitch values can be found in Figure 2. It can be seen that there is a small overlap between the values for the male and those for the female speakers. The correlation between sex of speaker and minimum pitch is relatively high: .84, which means that 71% of the variance in minimum pitch can be attributed to the sex of the speaker.

The pitch ranges in semitones of men and women do not differ significantly: 12.09 st (80 Hz) and 11.69 st (133 Hz) respectively ($t=-.543$, $df=112$, $p=.588$). The distributions are given in Figure 3. Male and female values overlap substantially. The correlation between pitch range and sex of speaker is very low: .05.

![Figure 3. Distribution of pitch range values](image)

3.2 **Sex of conversation partner**

Three groups of dyads were recorded: female-female, male-male, and female-male. The aim was to see if the sex of the conversation partner influenced pitch, in a possible interaction with the sex of the speaker. For the analyses in this section therefore male and female speakers will be examined separately.
Table 2. Correlations between sex of conversation partner and median pitch, minimum pitch, and pitch range. * = sign. 01

<table>
<thead>
<tr>
<th>Sex conv. partn.</th>
<th>Female speakers</th>
<th>Male speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median</td>
<td>minimum range</td>
</tr>
<tr>
<td>.35*</td>
<td>-.03</td>
<td>.28*</td>
</tr>
</tbody>
</table>

In Table 2 the correlations between the sex of the conversation partner and the three pitch measures are given. The correlation between sex of conversation partner and median pitch is significant, both for male and female speakers. For female speakers the correlation is positive, for male speakers negative. This means that speakers use a higher median pitch when speaking to a person of the opposite sex (female speakers to male listeners, male speakers to female listeners) than when speaking to a person of the same sex. In Figure 4 the mean values of minimum pitch, median pitch and maximum pitch (minimum pitch + pitch range) are given. Figure 4 shows that when a speaker addresses a member of the opposite sex, the bottom pitch remains relatively stable but the pitch peaks are increased by ca. 2 st. This causes an asymmetrical expansion of the pitch range (a significant effect for female speakers, see Table 2).

Semitones

Figure 4. Mean values of minimum pitch, median pitch, and maximum pitch for the speakers in all types of dyads: female speaker to female conversation partner (F to F), female to male (F to M), male to male (M to M), and male to female (M to F).
3.3 Gender identity

By means of a t-test it was checked first whether female and male speakers differed in their scores on the gender identity questionnaire. As was explained in section 2.4, there are six scores: feminine traits (FT), masculine traits (MT), feminine behavior (FB), and masculine behavior (MB) as judged by the speaker him/herself, and feminine traits (FTP) and masculine traits (MTP) as judged by his or her conversation partner. There was a significant difference in the scores for men and women on two of the six scores: FB (t=3.65, df=112, p<.001), and FTP (t=2.62, df=112, p=.010). In both cases, the women’s scores were higher than the men’s scores. In other words, female speakers attribute more feminine behavior to themselves than male speakers do. In addition, female speakers get attributed more feminine traits by their conversation partners than male speakers.

Because of the significant difference in the scores on part of the questionnaire, the statistical analyses with regard to the pitch measures were done for female and male speakers separately.

In Table 3 the correlations are given between the six gender identity scores and the three pitch measures. All of them are low. There are only two significant correlations, both involving masculinity. For female speakers there is a significant relation between masculine traits and minimum pitch. For male speakers there is a significant relation between masculine behavior and minimum pitch. Thus, for male and female speakers alike, a higher score on masculinity is linked to a higher minimum pitch. However, median pitch and pitch range do not differ significantly between speakers with high and low masculinity scores. Therefore, the relation between minimum pitch, median pitch and pitch range can be described best as a compression of the bottom half of the pitch range of more masculine speakers when compared to the pitch range of less masculine speakers.

Table 3. Correlations between the six gender identity scores and the three pitch measures for female and male speakers separately. *=sign.05. See text for abbreviations.

<table>
<thead>
<tr>
<th></th>
<th>female speakers</th>
<th>male speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median</td>
<td>minimum</td>
</tr>
<tr>
<td>MT</td>
<td>.10</td>
<td>.28*</td>
</tr>
<tr>
<td>MB</td>
<td>-.18</td>
<td>.22</td>
</tr>
<tr>
<td>MTP</td>
<td>-.04</td>
<td>.06</td>
</tr>
<tr>
<td>FT</td>
<td>-.12</td>
<td>.07</td>
</tr>
<tr>
<td>FB</td>
<td>.03</td>
<td>-.01</td>
</tr>
<tr>
<td>FTP</td>
<td>.00</td>
<td>.05</td>
</tr>
</tbody>
</table>
4. Discussion

The aim of the research was to provide an alternative explanation for overlap in pitch between and pitch variation within the sexes. Therefore, in addition to operationalizing gender as biological gender (sex of speaker and sex of conversation partner), as is customary, gender also was operationalized as social gender (gender identity measured with a masculinity/femininity questionnaire). Fifty-seven same-sex and mixed-sex conversations were recorded and one minute of speech was selected for each of the 114 speakers. Median pitch, minimum pitch, and pitch range were measured.

Not surprisingly, women had a higher median pitch and a higher minimum pitch than men. The first result is documented extensively (see Krook 1988), and the second result is in a sense a logical consequence of the first. However, the pitch ranges of male and female speakers did not differ significantly. There is a fair amount of research that reports similar results. Henton (1989), in an overview of 18 mostly English studies on mostly vowels and read speech, concludes that female and male pitch ranges are essentially the same. For Dutch, Tielen (1992) also reaches this conclusion for read and spontaneous speech.

The sex of the conversation partner turned out to have a significant effect on the median pitch of the speaker. Both female and male speakers used a higher median pitch when talking to a person of the opposite sex than when talking to a person of the same sex. This was caused by an asymmetrical expansion of the pitch range: higher pitches were increased while lower pitches were unaffected. According to the ‘frequency code’ of Ohala (1983) a low pitch is associated with physical and psychological power (e.g. tall, strong, assertive, independent, dominant), and high pitch with powerlessness (e.g. small, weak, not assertive, dependent, submissive). These associations hold for both men and women. A higher pitch for male speakers is, among other things, associated with less ambitious, less intelligent, and less confident (Brown, et al. 1974), and for female speakers with modest, low prestige, and dependent (Van Bezooijen 1995). It may be that in mixed-sex conversations speakers aim at appearing less dominant and competent than in same-sex conversations. In addition, alternative explanations are possible, different for the two sexes. For both sexes, high pitch is associated with femininity (Addington 1968; Van Bezooijen 1995). In mixed-sex conversations female speakers may use a higher pitch to sound more feminine, whereas this strategy is not to be expected of male speakers. Another strategy is accommodation. Male speakers may aim at reducing the interpersonal distance between them and their female conversation partners by reducing the distance between the male and female pitch level by adopting a higher pitch. Here this strategy is not
used by female speakers, because then they would have lowered their pitch in mixed-sex conversations and not raised it.

The gender identity scores (feminine and masculine traits and behavior judged by the speaker him/herself, feminine and masculine traits judged by his/her conversation partner) correlated poorly with the pitch measures. Two significant correlations were found: (1) female speakers with relatively high scores on masculine traits had a higher minimum pitch than female speakers with relatively low scores and (2) male speakers with relatively high scores on masculine behavior had a higher minimum pitch than male speakers who scored relatively low. A significant relation between masculinity and high minimum pitch is difficult to explain. The results of the present research indicated a lower minimum pitch for the male speakers in comparison with the female speakers. Therefore, a relation would be expected between low minimum pitch and high masculinity scores. However, the opposite was found; masculine speakers adopted a higher bottom pitch. Because median pitch and pitch range did not differ significantly between speakers with high and low masculinity scores, the difference in minimum pitch points to a compression of the bottom half of the pitch range of masculine speakers. At present, this result cannot be explained; more research is needed.

The lack of a strong significant relation between median pitch and gender identity was unexpected in the light of the results of attribution studies, where it is shown that differences in pitch have a distinct influence on how (gender) identities are perceived (e.g. Brown et al. 1974; Van Bezooijen 1995). One would expect these notions of gendered speech to have an influence on people's actual speaking behavior. Perhaps it is not just the median pitch level per se which reflects gender identity, but the relative position of the median pitch in the total pitch range or even phonational range of a speaker. Studying this will be the next step in the research.

Acknowledgments

I thank Renée van Bezooijen (University of Nijmegen) and Vincent van Heuven (Leiden University) for their comments on a previous version of this article. I thank Jos Pacilly (Leiden University) for developing the segmentation and channel identification program mentioned in paragraph 2.3, and for his help in handling the speech files.


