THE RELATIVE INTELLIGIBILITY OF MALE AND FEMALE SPEAKERS: IN SEARCH OF A METHOD

Mirjam T. J. Tielen

1. INTRODUCTION

A number of professions in the service industries like telephonist, receptionist or announcer at the railway station can be observed to be typically occupied by women. Intelligibility is of crucial importance in these professions. The main task of people in the service industries is to inform the public and sometimes there is a lot of noise which troubles the messages. Are women preferred in these professions because their voices can better resist several types of noise? Do women use a more exact pronunciation? Or is the reason a more psychologically or culturally determined one? Perhaps these professions are indeed feminine, not because of the demand for intelligibility but because of the lack of status connected to them.
On the other hand female voices are often excluded from applications in the speech technology because of inadequate methods for speech analysis of female voices. Most systems on speech synthesis and speech analysis are evaluated without having taken into account any female voices.
In view of these phenomena we wanted to find out whether or not (fe)male speakers are more intelligible. Furthermore, from a phonetic point of view it would also be interesting to point out the voice characteristics that are influential or responsible with regard to the relative intelligibility of male and female voices.
In this paper I will first describe the acoustical and perceptual differences between male and female voices and the possible role these characteristics play in determining the intelligibility.
Secondly, several methods for measuring the intelligibility are discussed in order to choose a suitable test method for our purpose.
Finally, the experimental procedure that has been used is described. The experiment is still in progress at the moment of writing this paper. Consequently no results can be given; they will be published in a subsequent paper.

2. MALE AND FEMALE VOICES

The voices of women and men have distinct acoustical and perceptual qualities. Physiological differences between the speech organs of the two sexes account for some of these distinct qualities.
The most important one is the difference in fundamental frequency (Fo) between male and female voices, which is a consequence of a difference in length and thickness of the vocal cords of men and women. Women have smaller vocal cords and hence they are speaking with a higher Fo. The average Fo of male speakers is about 120 Hz, of female speakers about 225 Hz (Fry, 1979). The fundamental frequency is the most important clue for perceptual measures like pitch and intonation. One may argue that the Fo is also the most important clue for sex identification (Coleman, 1976). Yet, the possible influence of this parameter on intelligibility has not been revealed.

A physical consequence of the lower Fo of male speakers is that the spectrum of a male speech signal shows a greater density of the successive harmonics. A greater density of the harmonics entails a closer approximation of the optimal resonance frequencies of the vocal tract; i.e. a better acoustical definition of the formant frequencies. As regards vowel intelligibility this fact could be an advantage to the intelligibility of male voices.
The position of the formant frequencies is another acoustical difference between male and female voices. This difference is due to the fact that in general the vocal tract of a man is larger than the vocal tract of a woman. Several studies have been devoted to male and female differences in formant frequencies in various vowels (Peterson and Barney, 1952; van Nierop et al.,1973; Klein et al., 1970). Perceptually, differences in formant frequencies give rise to differences in timbre. The vowel quality is determined by the formant frequencies. However, it is not clear beforehand that formant frequency is an important feature with respect to intelligibility.
Some researchers have reported about some other voice and speech differences between men and women. Koopmans- van Beinum (1980:71) put the case that female speakers produced larger acoustical contrasts than did male speakers. She observed in several studies a more careful articulation of female speakers. It is possible that this female speech behaviour would be advantageous with respect to the intelligibility.
Brend (1975) presented a listing of intonation patterns that would be different for men and women. She observed that "...some very definite preferences in the general usage and avoidance of some of the patterns by men versus women" (Brend, 1975:84) existed. Yet it is not clear whether or not these characteristics have an influence upon intelligibility.

3. METHODS FOR MEASURING INTELLIGIBILITY

While looking for intelligibility tests it appears that most of the tests are developed and used in the field of audiology (speech audiometry). Generally, the purpose of those tests is to determine the amount of hearing loss of a listener. The methods differ in the way the stimuli are presented, or in the linguistic level at which intelligibility is measured (e.g. phoneme or sentence level). The standard articulation test (French and Steinberg, 1947; Asher, 1958) is well-known. Usually, monosyllabic words are presented to a listener. The listener has to recall or to write down the word he/she hears. The percentage correctly identified phonemes or words is called the articulation score.
The variant in which the lists are Phonetically Balanced (PB lists) can be regarded as a development of the articulation test. The phonemes in a Phonetically Balanced list are claimed to have a frequency of occurrence that represents the frequency of occurrence of the phonemes in a particular language. One can ask in what way this counting of relative phoneme frequencies is achieved (see for a discussion: Gil-Günzburger and Vingerling, 1980). Furthermore, we think it is interesting to measure the intelligibility of all phonemes, whether or not they have a high frequency of occurrence. Some less frequent phonemes are excluded in PB lists.
Another sort of test is the Rhyme test and the variants are tests with closed and open response sets (House et al., 1965; Nye and Gaitenby, 1973). Different CVC combinations are presented to the subjects; the subjects are forced to choose one of several response alternatives. A disadvantage of Rhyme tests is that only meaningful words are possible responses. This fact can influence the scores because of the limited number of alternatives.
Tests used for measuring the intelligibility on word and sentence level (e.g. Nye and Gaitenby, 1974; Plomp and Mimpen, 1979) are also available. Although intelligibility measurements on these higher linguistic levels are more representative for everyday

speech occurrences, the lack of insight in the precise significance that context and predictability have in the intelligibility forms a problem connected to these tests.
Apart from the use of the tests for speech audiometry purposes (topic on the perception), nowadays variants of these tests are also used for evaluation of different speech transmission systems (topic on the medium) or different speakers (topic on the production). In his overview of evaluation methods for different speech systems, Pols (1987) sums up the type of independent variables one could deal with in measuring word or sentence intelligibility of different synthesis systems.
Nakatani and Dukes (1973) wanted to evaluate the quality of several degraded speech types like telephone speech and low pass filtered speech relative to a reference speech. For this purpose they needed an intelligibility test that would be sensitive enough to discriminate among rather high-quality speech samples. Nakatani and Dukes developed the so called speech interference test, which is a kind of articulation test with interfering speech. The idea behind the use of an interfering signal is that by obstructing the speech reception task small differences in quality are magnified. Background speech is often an interfering element in speech communication. Besides, speech as a masker would yield less bias to speech than noise (Nakatani and Dukes, 1973). Moreover, interfering speech taken from the same speaker has the same long term average spectrum as the stimuli and therefore the signal-to-noise ratio is defined better. Nonsense sentences of the pattern "The 'ADJECTIVE' 'NOUN' 'PAST TENSE' the 'NOUN'" (e.g. The blue tire held the king) were presented to subjects. The subjects had to write down the four content words. The interfering speech was produced by the same speaker, but was taken from a different speech context.
By this method Nakatani and Dukes were able to differentiate in quality between several degraded speech types. Vogten (1980) adopted this method to examine the quality differences between some resynthesized speech types. He too showed that the speech interference test could differentiate between speech types of rather high quality.
Eggen (1987) increased the efficiency of the speech interference test by determining the speech interference threshold in an adaptive method. The signal-to-noise ratio of the stimuli was increased if the preceding response of the listener was wrong and decreased if that response was correct. In this way most of the stimuli are presented around the 50% intelligibility threshold. Another difference with (and we think an advantage to) the Nakatani and Dukes' method is the fact that in changing the signal-to-noise ratio Eggen varied the level of the interfering speech instead of the level of the stimuli.

In our study we looked for a method to measure the relative intelligibility of male and female voices. For this purpose we needed a very sensitive test that would be able to differentiate between a number of rather high quality speech samples. The idea of the interfering speech seemed very attractive for our design in which female voices had to be compared to male voices. Most other types of noise (like white noise) can cause trouble in interpreting the results because some types mask female voices better and some mask male voices better. The signal-to-noise ratio of all stimuli is defined more precisely if the speech of a speaker is also used as interfering speech for the same speaker.
Therefore we decided to use the Monosyllabic Adaptive Speech Interference Test of Eggen (1987) to measure the 50% word intelligibility threshold of all speakers. Afterwards an articulation test is performed to get insight into the type of phoneme confusions as well, while listening to male and female speakers.

4. EXPERIMENTAL SET-UP

In the present study the relative intelligibility of male and female voices is measured in two experimental conditions. Unlike in most experiments on intelligibility the outcomes of two different groups, male and female speakers, will be compared. Firstly, the word intelligibility threshold per listener is determined by the speech interference test with an adaptive method. Secondly, a standard articulation test is performed, while masking the stimuli with interfering speech. All speakers produced the same phonemes, but in different CVC combinations. The experiment was carried out partly with the help of computer programs developed at the Institute of Perception Research in Eindhoven (Eggen, 1987).

4.1 Stimuli

We opted for isolated CVC words as stimuli for the experiment just like is done in many existing intelligibility tests. The advantage of the usage of CVC words is that these words can be reproduced easily by the listeners. Moreover these words do not lead to ambiguous interpretations by listeners caused by differences between speakers. Besides, the intelligibility of a number of consonants can be measured in two positions, word-initial and word-final. A disadvantage of CVC words is that a lot of nonsense words are included because many different CVC combinations are necessary. Moreover, the intelligibility of very small speech units is measured which is not representative for everyday speech events.

Table 1. Distribution of the phonemes over a CVC list

C initial		V		C final	
/b/	3	/ɑ/	4	/f/	4
/d/	3	/a/	3	/x/	4
/f/	3	/ɛ/	4	/k/	4
/x/	3	/e/	3	/l/	4
/h/	3	/ɪ/	4	/m/	4
/j/	2	/i/	3	/p/	4
/k/	3	/ɔ/	4	/r/	4
/l/	3	/o/	3	/s/	4
/m/	3	/œ/	4	/t/	4
/n/	3	/y/	3	/n/	4
/p/	3	/ø/	3	/ŋ/	4
/r/	3	/u/	3	/w/	3
/s/	3	/œi/	3	/j/	3
/t/	3	/ɛi/	3		
/v/	3	/ɑu/	3		
/w/	3				
/z/	3				

Lists of fifty CVC words were generated at random by a computerprogram. Several limitations with respect to the sort of possible combinations were included (Cohen et al., 1959). These limitations are intended to produce only phonotactically correct CVC words (see appendix). Certain consonants cannot occur in word-initial or in word-final

position. A phoneme distribution as shown in table 1 was designed while taking these phonotactic constraints into account. We also tried to keep a same number of identical phonemes in each list. The lists are not phonetically balanced, because we intended to incorporate the less frequent Dutch phonemes as well in testing the phoneme intelligibility. The phonemes have the same frequency of occurrence in all lists, so the different stimulus lists will be comparable.

4.2 Recordings

Twenty persons (ten males and ten females) participated as speakers. They have been selected according to the following criteria:
a. All speakers had to be native speakers of the Dutch language.
b. Speakers of a dialect were excluded; all were living or working around Amsterdam.
c. The voices had to be normal; e.g. not marked or deviant.
d. The speakers had to be of approximately the same age for the sake of homogeneity.
Therefore we chose a group of speakers age between forty and fifty years old. All speakers volunteered on an unpaid basis.
In an anechoic room the speakers read aloud three lists of fifty CVC words. All CVC words followed the string "noteer het woord" ("write the word"). The first list was the same for all speakers; this list was meant to make the speaker familiar with the task and to provide an opportunity for the experimenter to determine the optimal level of the recording (this is the level with the highest possible signal-to-noise ratio). The second and third CVC list were different for each speaker; only these lists were used in the listening sessions. Afterwards the speakers produced five carrier sentences like the following: "Je krijgt het woord x te horen" ("You will hear the word x").
All word lists were equalized in intensity using the EPL (Equivalent Peak Level) method of Brady (1968). The EPL level was measured over one list.
Each of the stimuli was placed into one of the carrier sentences. The stimuli were placed at those positions where the speakers had produced "x". This was done on a VAX computer by segmentation and concatenation of the different speech files. Each carrier sentence was present ten times in every list. The distribution of the sentences over the stimuli was at random. These lists are called henceforth stimulus lists.
All speech material of a speaker was also used as interfering speech for that same speaker. It was decided to create the interfering speech from the stimulus lists in order to make the interfering speech as similar as possible to the stimuli. Every sentence was concatenated to the preceding sentence. This resulted in a long speech fragment. Five versions, which only differed in the order of the sentences were generated and added on top of each other with delays of five seconds each. The resulting speech babble was reversed and repeatedly recorded on a Revox tape recorder. This interfering speech was generated for each of the twenty speakers.

4.3 Listening procedure

Two males and two females participated as listeners. All listeners had Dutch as their native language. All had normal hearing according to a pure tone audiometric test. The listeners were between twenty and thirty years of age and they were supposed to be unfamiliar with the voices. As each listener had to respond to all stimuli, the listening sessions required much time. Therefore the subjects were paid for their cooperation.
During the listening sessions the speech was played from the digital to analogue converter of a VAX computer; the interfering speech was played from a Revox tape recorder. The speech was mixed with the interfering speech of that same speaker and

the resulting signal was presented binaurally to the listener. The listener was seated in an anechoic room and heard the stimuli and the interfering speech through headphones. The responses had to be typed on a terminal keyboard.
In the first condition (word intelligibility in an adaptive method) the speech interference threshold was determined by amplifying or attenuating the level of the interfering speech, depending upon the response of the listener on the preceding stimulus. Each session started with a high signal-to-noise ratio. After each correct response the level of the interfering speech was amplified with 2dB in order to decrease the signal-to-noise ratio. The signal-to-noise ratio was increased with 2dB by attenuating the interfering speech if the response of the listener was wrong.
In the second condition (articulation test) the output levels of the stimuli and the interfering speech were kept constant during the listening sessions. The ratio was determined per speaker by using the thresholds of word intelligibility of the first condition.
In both conditions the listener has been presented with one stimulus list of each speaker. Per listening session five lists were scored by a listener. Thus each listener had to perform eight sessions. One session lasted ca. fifty minutes, depending on the scoring time a listener needed.
Two try-out sessions were presented to the listeners before the sessions started. In this way they could accustom to the response task.

5. CONCLUDING REMARKS

Our aim is to know whether any differences between male and female voices play a role in or are responsible for a better intelligibility of either of these two. Therefore we looked for a method that would be appropriate to test the relative intelligibility of twenty different voices. Hopefully the adaptive speech interference test is indeed sensitive enough to discriminate between the various voices and possibly to indicate a significantly different intelligibility of male and female voices. The articulation test will show the relative phoneme intelligibility and perhaps give insight into differences in phoneme confusions in male and female speech.

ACKNOWLEDGEMENT

I wish to thank Florien Koopmans-van Beinum, Leo van Herpt and Louis Pols for reviewing this paper.

Appendix.

The following rules hold for all CVC words:

The /j/ in word-final position only after /a/, /o/ and /u/.
The /w/ in word-final position only after /e/, /i/ and /y/.
The /ŋ/ (always in wordfinal position) only after /ɑ/, /ɛ/, /i/ and /ɔ/.
The /r/ cannot be preceded by diphthongs and /œ/.
The /p/ cannot be preceded by /ɑu/.
The /f/ cannot be preceded by /ɑu/ or/ /y/.
The /x/ cannot be preceded by /ɑu/.
The /m/ cannot be preceded by /ɑu/ or /y/.

6. REFERENCES

Asher, J.W. (1958) Intelligibility tests: a review of their standardization, some experiments, and a new test. Speech Monographs 25, 14- 28.

Brady, P.T. (1968) Equivalent peak level: a threshold-independent speech-level measure. J. Acoust. Soc. Am. 44, 695- 699.

Brend, R.M. (1975) Male-Female intonation patterns in American English. in: Thorne, B. and N. Henley (eds). Language and Sex: difference and dominance, 84- 87.

Cohen, A., Ebeling, C.L., Eringa, P., Fokkema, K. and Holk, A.G.F. van (1959) Fonologie van het Nederlands en het Fries. M. Nijhoff/ 's- Gravenhage.

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.

Eggen, J.H. (1987) Start evaluation of available methods for analysis, manipulation and resynthesis of speech project EVA1. IPO Report 612, 2- 16.

French, N.R. and Steinberg, J.C. (1947) Factors governing the intelligibility of speech sounds. J. Acoust. Soc. Am. 19, 90- 119.

Fry, D.B. (1979) The Physics of Speech. Cambridge Univ. Press/ Cambridge.

Gil-Günzburger, D. and Vingerling, M. (1980) Examination of word lists currently used in speech audiometry. Progress Report Institute of Phonetics, Univ of Utrecht 5;1, 48- 59.

House, A.S., Williams, C.E., Hecher, H.L. and Kryter, K.D. (1965) Articulation-testing methods: consonantal differentiation with a closed response set. J Acoust. Soc. Am. 37, 158- 166.

Klein, W., Plomp, R. and Pols, L.C.W. (1970) Vowel spectra, vowel spaces and vowel identification. J. Acoust. Soc. Am. 48, 999- 1009.

Koopmans–van Beinum, F.J. (1980) Vowel contrast reduction: an acoustic and perceptual study of Dutch vowels in various speech conditions. Diss. Academic Press, A'dam.

Nakatani, L.H. and Dukes, K.D. (1973) A sensitive test of speech communication quality. J. Acoust. Soc. Am. 53, 1083- 1092.

Nierop, D.J.P.J. van, Pols, L.C.W. and Plomp, R. (1973) Frequency analysis of Dutch vowels from 25 female speakers. Acoustica 29, 110- 118.

Nye, P.W. and Gaitenby, J.H. (1973) Consonant intelligibility in synthetic speech and in a natural speech control (MRT results). Haskins Laboratories: Status Report on Speech Research 33, 77- 91.

Nye, P.W. and Gaitenby, J.H. (1974) The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences. Haskins Laboratories: Status Report on Speech Research- 37/ 38, 169- 190.

Peterson, G.E. and Barney, H.L. (1952) Control methods used in a study of the vowels. J. Acoust. Soc. Am. 24, 175–184.

Plomp, R. and Mimpen, A. M. (1979) Improving the reliability of testing the Speech Reception Threshold for sentences. Audiology 18, 43- 52.

Pols, L.C.W. (1987) Quality evaluation of text- to speech synthesis systems. IFA Report 94.

Vogten, L.L.M. (1980) Evaluation of LPC formant coded speech with a speech interference test. IPO Annual Progress Report 15, 33- 41.