3

Automatic Speaker Recognition by Speech Signal

Milan Sigmund
Brno University of Technology
Czech Republic

1. Introduction

Acoustical communication is one of the fundamental prerequisites for the existence of human society. Textual language has become extremely important in modern life, but speech has dimensions of richness that text cannot approximate. From speech alone, fairly accurate guesses can be made as to whether the speaker is male or female, adult or child. In addition, experts can extract from speech information regarding e.g. the speaker's state of mind. As computer power increased and knowledge about speech signals improved, research of speech processing became aimed at automated systems for many purposes.

Speaker recognition is the complement of speech recognition. Both techniques use similar methods of speech signal processing. In automatic speech recognition, the speech processing approach tries to extract linguistic information from the speech signal to the exclusion of personal information. Conversely, speaker recognition is focused on the characteristics unique to the individual, disregarding the current word spoken. The uniqueness of an individual's voice is a consequence of both the physical features of the person vocal tract and the person mental ability to control the muscles in the vocal tract. An ideal speaker recognition system would use only physical features to characterize speakers, since these features cannot be easily changed. However, it is obvious that the physical features as vocal tract dimensions of an unknown speaker cannot be simply measured. Thus, numerical values for physical features or parameters would have to be derived from digital signal processing parameters extracted from the speech signal. Suppose that vocal tracts could be effectively represented by 10 independent physical features, with each feature taking on one of 10 discrete values. In this case, 10^{10} individuals in the population (i.e., 10 billion) could be distinguished whereas today's world population amounts to approximately 7 billion individuals.

People can reliably identify familiar voices. About 2-3 seconds of speech is sufficient to identify a voice, although performance decreases for unfamiliar voices. One review of human speaker recognition (Lancker et al., 1985) notes that many studies of 8-10 speakers (work colleagues) yield in excess of 97% accuracy if a sentence or more of the test speech is heard. Performance falls to about 54% when duration is shorter than 1 second and/or distorted e.g., severely highpass or lowpass filtered. Performance also falls significantly if training and test utterances are processed through different transmission systems. A study

using voices of 45 famous people in 2 seconds test utterances found only 27% recognition in an open-choice test, but 70% recognition if listeners could select from six choices (Lancker et al., 1985). If the utterances were increased to 4 seconds, but played backward (which distorts timing and articulatory cues), the accuracy resulted to 57%. Widely varying performance on this backward task suggested that cues to voice recognition vary from voice to voice and that voice patterns may consist of a set of acoustic cues from which listeners select a subset to use in identifying individual voices. Recognition often falls sharply when speakers attempt to disguise their voices e.g., 59-81% accuracy depending on the disguise vs. 92% for normal voices (Reich & Duke, 1979). This is reflected in machines, where accuracy decreases when mimics act as impostors. Humans appear to handle mimics better than machines do, easily perceiving when a voice is being mimicked. If the target (intended) voice is familiar to the listener, he often associates the mimic voice with it. Certain voices are more easily mimicked than others, which lends further evidence to the theory that different acoustic cues are used to distinguish different voices.

From the performance point of view, automatic speaker recognition by speech signal can be seen as an application of artificial intelligence, in which machine performance can exceed human performance e.g., using short test utterances and a large number of speakers. This is especially true for unfamiliar speakers, where the training time for humans to learn a new voice well is very long compared with that for machines. Constraints on how many unfamiliar voices a person can retain in short-term memory usually limit studies of speaker recognition by humans to about 10 speakers.

2. Verification and Identification of Speakers

Speaker recognition covers two main areas: speaker verification and speaker identification. Speaker verification is concerned with the classification into two classes, genuine person and impostor. In verification, an identity claim is made by an unknown speaker, and an utterance of the unknown speaker is compared with the model for the speaker whose identity is claimed. If the match is above a certain threshold, the identity claim is verified. Figure 1 shows the basic structure of a speaker verification system. A high threshold makes it difficult for impostors to be accepted by the system, but at the risk of rejecting the genuine person. Conversely, a low threshold ensures that the genuine person is accepted consistently, but at the risk of accepting impostors. In order to set a threshold at a desired level of user acceptance and impostor rejection, it is necessary to know the distribution of customer and impostor scores.

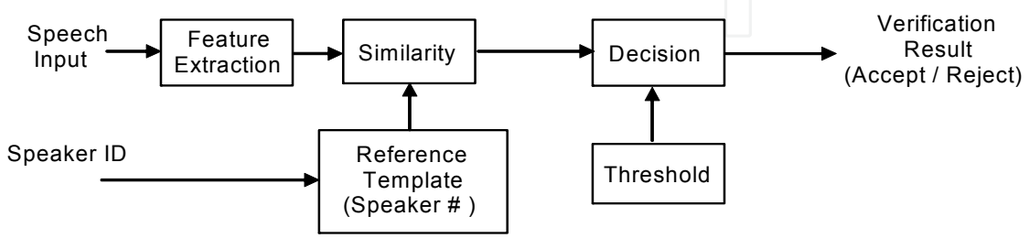

Fig. 1. Basic structure of speaker verification system.

There are two corresponding types of errors, namely the rejection of genuine speakers, often called false rejection, and the acceptance of impostors, often called false acceptance. The most common performance measure used for comparing speaker verification systems is the equal error rate. The equal error rate is found by adjusting the threshold value until the false acceptance rate is equal to the false rejection rate. In most cases, this value must be determined experimentally by collecting the recognition scores for a large number of both accepting and rejecting comparisons. This involves applying an a-posteriory threshold. An illustration of an error rate graph is shown in Figure 2. The use of an equal error rate implies a perfect choice of threshold, which is not possible in a real application since the threshold would have to be determined a-priory. This problem can be solved using probability theory. The threshold for speaker verification must be updated with long-term voice variability (Matsui et al., 1996).

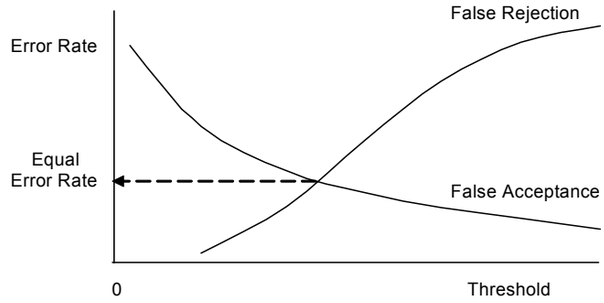

Fig. 2. False rejection rate and false acceptance rate as a function of the decision threshold.

In speaker identification, a speech utterance from an unknown speaker is analysed and compared with models of known speakers. The unknown speaker is identified as the speaker whose model best matches the input utterance. Figure 3 shows the basic structure of a speaker identification system.

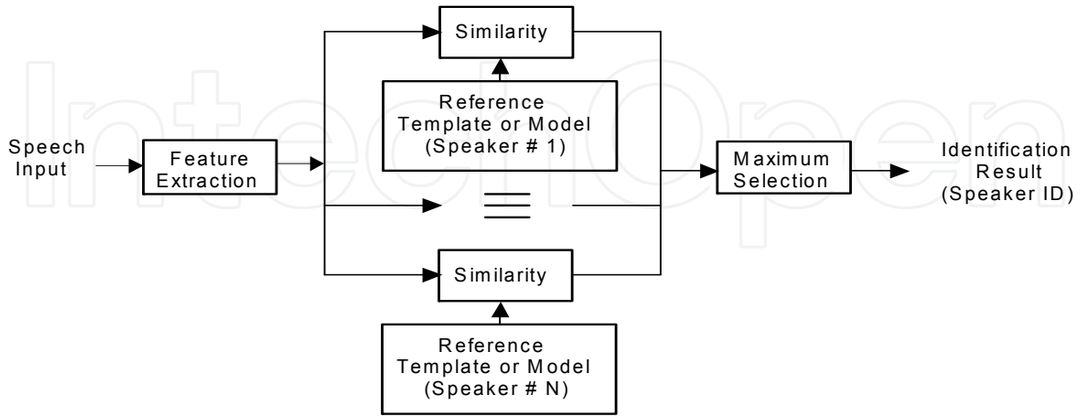

Fig. 3. Basic structure of speaker identification system.

There is also the case called “open set“ identification, in which a model for the unknown speaker may not exist. In this case, an additional decision alternative, “the speaker does not

match any of the models", is required. The fundamental difference between identification and verification is the number of decision alternatives. In identification, the number of decision alternatives is equal to the size of the population, whereas in verification there are only two decision alternatives (accept or reject).

3. Text-Dependent Speaker Recognition

Speaker recognition methods can also be divided into text-dependent and text-independent methods. The former require the speaker to provide utterances of the key words or sentences having the same text for both training and recognition trials, whereas the latter do not rely on a specific text being spoken. The text-dependent methods are usually based on template matching techniques in which the time axes of an input speech sample and each reference template or reference model of registered speakers are aligned, and the similarity between them accumulated from the beginning to the end of the utterance is calculated. The structure of text-dependent recognition systems is, therefore, rather simple. Since this method can directly exploit the voice individuality associated with each phoneme or syllable, it generally achieves higher recognition performance than the text-independent method.

3.1 Effectiveness of Various Phonemes for Speaker Recognition

The speaker-specific information contained in short-term spectra was used in the initial experiments. Twelve male speakers read the same text twice. The signal was sampled at 22 kHz with 16-bit linear coding. The speech signals were labeled using own tool (Sigmund & Jelinek, 2005), and a log-power spectrum (128 point FFT) was calculated in the centre of continuant sounds. The spectral channel containing maximum intensity was then set at 0 dB. The reference samples were created by averaging three spectra. Finally, the spectra were compared by a distance measure derived from a correlation based similarity measure

$$d = 1 - \frac{< \mathbf{x}, \mathbf{y} >}{\|\mathbf{x}\| \, \|\mathbf{y}\|} \tag{1}$$

where **x** and **y** represent the spectral vectors. Each phoneme in the test was compared with each of the corresponding reference phonemes. The reference sample with the minimal distance was considered to be identified. The identification rate varies from 11% to 72%. The results obtained indicate that an individual analysis of each phoneme is impossible but that the data can be reasonably grouped into phonetically defined classes. Table 1 gives average identification rates. Thus, in terms of speaker-recognition power, the following ranking of phoneme classes results:

vowels, nasals > liquids > fricatives, plosives.

As expected, vowels and nasals are the best phonemes for speaker identification. They are relatively easy to identify in speech signal and their spectra contain features that reliably distinguish speakers. Nasals are of particular interest because the nasal cavities of different speakers are distinctive and are not easily modified (except when nasal congestion). For our purposes, Table 1 gives a preliminary general overview of the results.

Phoneme Class	Identification Rate (in %)
Vowels (a, e, i, o, u)	68
Nasals (m, n)	67
Liquids (l, r)	53
Fricatives (f, s, sh, z)	46
Plosives (p, t, b, d, g)	32

Table 1. Speaker identification rate by phoneme classes.

Two experiments were then performed on the data set within the vowel class. Because of the formant (i.e. local maxima) structure of vowel spectra the identification rate for each vowel can be estimated. In Table 2, individual vowels are compared in terms of speaker-recognition power.

Vowel	No. of Vowels in Test	Identification Rate (in %)
i	117	52.7
o	106	61.4
u	85	68.2
a	121	74.8
e	122	76.2

Table 2. Speaker identification rate by individual vowels.

It is known that different speakers show not only different formant values but exhibit different arrangements in their vowel systems. The general distribution patterns of vowels in formant planes can be used to build up a feature matrix for the vowel system of individual speakers. Table 3 shows the identification rate for various numbers of different vowels. The test started with only one vowel (the most effective) and successively other vowels were added one by one according to their individual effectiveness as ranked in Table 2. The identification rate increased almost logarithmically from 76.2% using the one individually best vowel "e" up to 97.4% using all the five vowels simultaneously.

No. of Vowels	Vowels	Identification Rate (in %)
1	e	76.2
2	e, a	88.7
3	e, a, u	93.8
4	e, a, u, o	95.6
5	e, a, u, o, i	97.4

Table 3. Speaker identification rate depending on number of vowels used.

3.2 Effectiveness of Speech Features in Speaker Recognition

In order to see which features are effective for speaker recognition, we studied here the following six parametric representations: 1) autocorrelation coefficients; 2) linear prediction (LP) coefficients; 3) log area ratios; 4) cepstral coefficients; 5) mel-cepstral coefficients; 6) line spectral-pair (LSP) frequencies. More details how to compute these parameters could be found in (Rabiner & Juang, 1993). Although all of these representations provide equivalent information about the LP power spectrum, it is only the LSP representation that has the localized spectral sensitivity property. As can be seen in Section 3.1, the vowel phonemes result the best in recognition performance regarding the speaker identification rate. Thus, the vowels as speech data used for this purpose were derived from utterances spoken by nine male speakers. These utterances were low-pass filtered at 4 kHz and sampled at 10 kHz. The steady-state part of the vowel segment was located manually.

For each speaker and for each feature set the first ten coefficients were used. The Euclidean distance was obtained by comparing a test vector against a template. A match was detected based on the minimum distance criterion, if the intra-speaker distance was shorter than all the inter-speaker distances. Otherwise a mismatch was declared. These matches and mismatches were registered in the confusion matrices for each parametric representation. Table 4 shows recognition rates for all six parametric representations mentioned above. From these results, it can be seen that for text-dependent speaker recognition the autocorrelation coefficients are not very effective, the log area ratio coefficients set generally surpasses any other feature sets, and mel-cepstral coefficients are comparable with LSP frequencies in recognition performance. In order to compute the text-dependent speaker recognition performance for each feature set, the following procedure was used. For each vowel, five repeats were used as the training set and about thirty randomly chosen vowels were used as the test set; all this for a given speaker. The training set and the test set were disjunct.

Parameters	Test Patterns	Recognition Rate (in %)
Autocor. coeffs.	270	61.3
LP coeffs.	268	83.7
Log area ratios	254	94.1
Cepstral coeffs.	262	87.5
Mel-Cepstal coeffs.	249	91.2
LSP frequencies	241	90.8

Table 4. Performance of vowel-dependent speaker recognizer using various parametric representations.

4. Text-Independent Speaker Recognition

There are several applications in which predetermined key words cannot be used. In addition, human beings can recognize speakers irrespective of the content of the utterance. Therefore, text-independent methods have recently been actively investigated. Another advantage of text-independent recognition is that it can be done sequentially, until a desired

significance level is reached, without the annoyance of repeating the key words again and again. In text-independent speaker recognition, the words or sentences used in recognition trials cannot generally be predicted. For this recognition, it is important to remove silence/noise frames from both the training and testing signal to avoid modeling and detecting the environment rather than the speaker.

4.1 Long-Term Based Methods

As text-independent features, long-term sample statistics of various spectral features, such as the mean and variance of spectral features over a series of utterances, are used (see Fig. 4). However, long-term spectral averages are extreme condensations of the spectral characteristics of a speaker's utterances and, as such, lack the discriminating power included in the sequences of short-term spectral features used as models in text-dependent methods. The accuracy of the long-term averaging methods is highly dependent on the duration of the training and test utterances, which must be sufficiently long and varied.

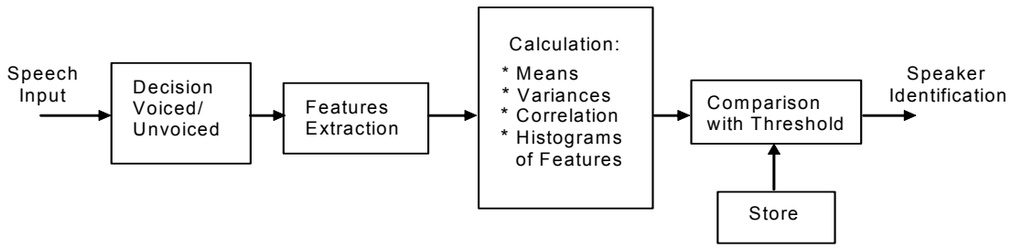

Fig. 4. Typical structure of the long-term averaging system.

4.2 Average Vocal Tract Spectrum

In a long-time average spectrum of a speech signal the linguistic information (coded as frequency variation with time) is lost while the speaker specific information is retained. In this study, a speaker analysis approach based on linear predictive coding (LPC) is presented. The basic idea of the approach is to evaluate an average long-time spectrum corresponding to the anatomy of the speaker's vocal tract independent of the actually pronounced phoneme. First, we compute the short-time autocorrelation coefficients R_j (k), k=1,...,K for the j-th frame (20 msec) of speech signal s (n)

$$R_j(k) = \sum_{n=1}^{N-k} s(n)\, s(n+k) \tag{2}$$

where N is the number of samples in each frame, and then we compute the K average autocorrelation coefficients

$$\overline{R}(k) = \frac{1}{J} \sum_{j=1}^{J} R_j(k) \tag{3}$$

corresponding to the whole utterance formed by J frames. Thus, from the average autocorrelation coefficients, we get the average predictor coefficients $\overline{a}_m$ e. g., via the Durbin algorithm (Rabiner & Juang, 1993) and finally the normalised average LPC-based spectrum using

$$S(f) = \left| \frac{1}{1 - \sum_m a_m z^{-m}} \right|^2 \qquad \text{for } m = 1,...,M \tag{4}$$

where $z = \exp(j\, 2\pi f / f_s)$, f_s is the sampling frequency and M is order of the LPC model equal to the highest autocorrelation order K. More details how to compute the LPC coefficients and corresponding spectra on short frame of speech signal can be found in (Rabiner & Juang, 1993). The speech signal was sampled at 22 kHz using a 16-bit A/D converter under laboratory conditions over a period of five months. A group of 26 speakers (19 male, 7 female) aged 20 to 25 years took part in the tests.

A comparison between intra- and inter-speaker variability in long-time spectrum is shown in Figures 5 and 6. Figure 5 illustrates two vocal tract spectra of the same speaker corresponding to two different texts. The difference between both curves is 12%.

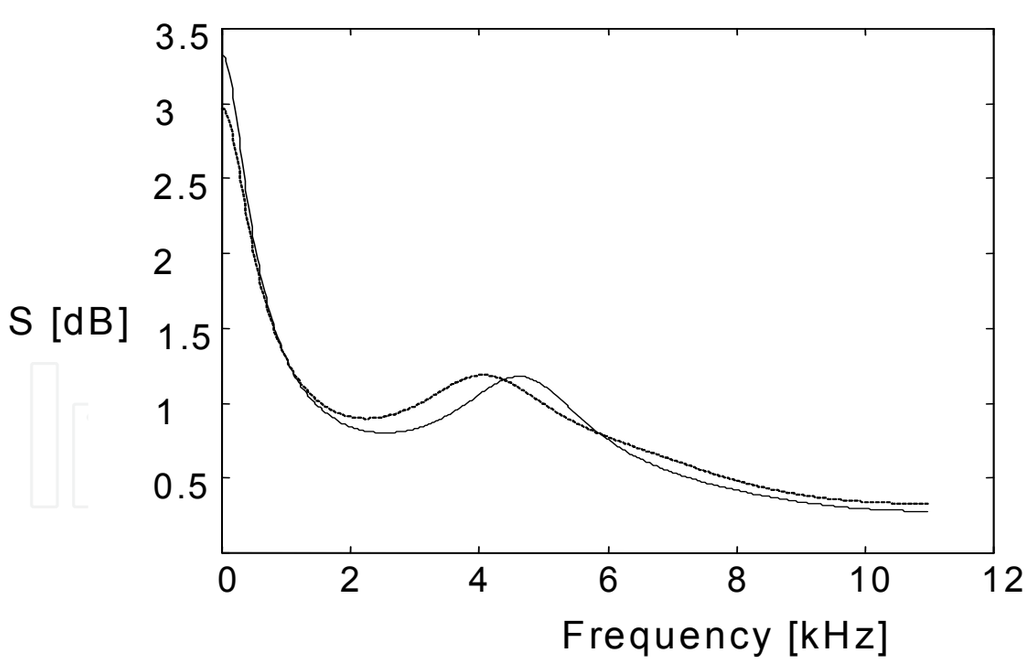

Fig. 5. Long-time spectrum difference of one and the same speaker (LPC order 6, speech duration 100 sec).

Vocal tract spectra obtained from two different speakers saying the same text is shown in Figure 6. The difference between both curves increased to 22% in this case. The average

intra-speaker difference over all speakers was 12.6%, while the average inter-speaker difference (gender-specific) reached 23.4%. In accordance with the inter-gender differences, the estimated difference between the two groups of speakers (male and female) was more apparent (29.6%) than within the groups (Sigmund & Mensik, 1998).

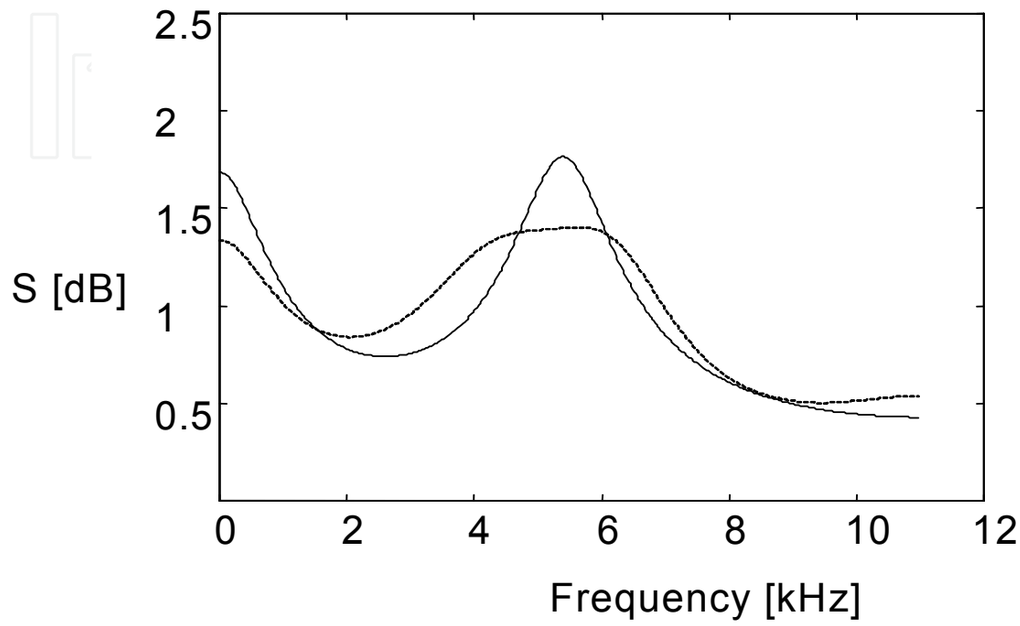

Fig. 6. Long-time spectrum variability between speakers (speech duration 100 sec).

5. Automatic Gender Recognition

Some studies show that speaker identification as well as speech recognition would be simpler, if we could automatically recognize a speaker's gender (sex). For example, in the "cocktail party effect", the voices of two or more speakers may be mixed. If the speakers are of opposite sex and if sex identification can be made on short segments of speech, the voices can be at least partially separated. Sex identification was used primarily as a means to improve recognition performance and to reduce the needed computation. Accurate sex identification has different uses in spoken language systems, where it can permit the synthesis module of a system to respond appropriately to an unknown speaker. In languages like French, where formalities are often used, the system acceptance may be easier if greetings such as "Bonjour Madame" are foreseen. In the past, automatic gender identification has been investigated for clean speech by Wu and Childers (Wu & Childers, 1991). Clean speech and speech affected by adverse conditions are evaluated for a variety of gender identification schemes in (Slomka & Sridharan, 1997). Using speech segments with an average duration of 890 msec (after silence removal), the best mentioned accuracy is 98.5% averaged over all clean and adverse conditions. There is some evidence that sex-related speech characteristics are only partly due to physiological and anatomical differences between the sexes; cultural factors and sex-role stereotypes also play an important part.

The main feature which can speaker's sex distinguish is fundamental frequency F_0 with typical values of 110 Hz for male speech and 200 Hz for female speech. The pitch of children is so different that they are often treated as "the third sex". Most values of F_0 among people aged 20 to 70 years lie between 80-170 Hz for men, 150-260 Hz for women while 300-500 Hz for children (Baken & Orlikoff, 2000). There are Gaussian distributions of these ranges, so that dispersion is wide and we often could not categorize the acoustic signal reliably by using this criterion only. Figure 7 (Titze, 1989) illustrates the inverse relationship between fundamental frequency of speech F_0 and length of glottal membrane L_m .

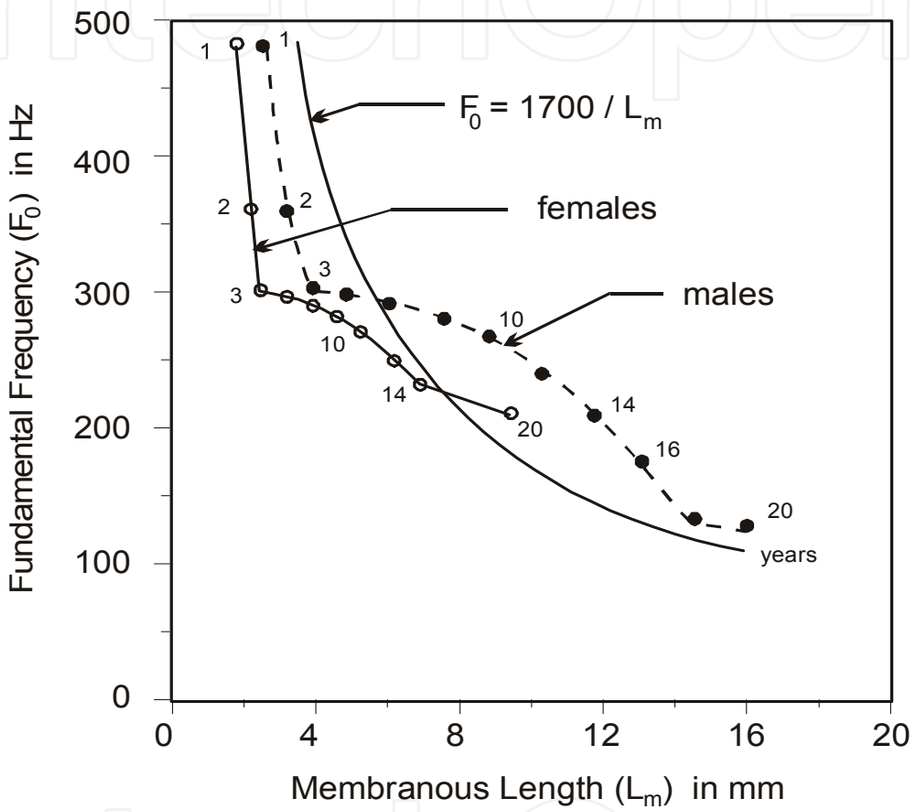

Fig. 7. Mean speaking fundamental frequency F_0 as a function of membranous length L_m.

5.1 Cepstral Analysis

The most commonly used short-term features in speech signal processing are cepstral coefficients and their frequency-warped alternative coefficients. Thus, the mel-frequency warped cepstral coefficients were taken for our experiment to identify the sex of a speaker. First, a Hamming window was applied for each speech frame (20 msec) of the recorded vowels and the FFT spectrum was computed. Then, the spectrum was mel-warped and the inverse Fourier transform of the logarithm of the warped spectrum produced the vector of cepstral coefficients. The mel-frequency scale is linear below 1 kHz and logarithmic above 1 kHz (Rabiner & Juang, 1993). Using a set of 41 mel-cepstral coefficients c_0 through c_{40} and their various differences, the performance of these individual features as identifiers of the sex of a speaker was measured. Table 5 summarizes the selected suitable coefficients which had the lowest variation calculated individually for Czech vowel phonemes and then

averaged for both genders separately. The best individual feature seems to be the coefficient c_{24} followed by c_{26} and c_{25}, respectively. On the other hand, the differences of cepstral coefficient pairs are not reliable for sex identification (Kepesi & Sigmund, 1998).

		c_0	c_1	c_2	c_6	c_9	c_{17}	c_{18}
Male	μ	-828,0	-326,0	338,0	28,0	94,0	21,0	31,0
	σ	181,0	122,0	141,0	73,0	52,0	65,0	39,0
Female	μ	-1150,0	-597,0	164,0	182,0	-20,0	129,0	112,0
	σ	125,0	88,0	137,0	70,0	53,0	45,0	47,0
		c_{22}	c_{23}	c_{24}	c_{25}	c_{26}	c_{35}	c_{36}
Male	μ	20,0	-35,0	-1,2	4,6	-42,0	20,9	60,1
	σ	46,0	51,0	39,0	57,0	47,0	41,0	45,0
Female	μ	109,0	37,0	98,0	158,0	101,0	-119,0	-84,0
	σ	57,0	55,0	29,0	28,0	34,0	80,0	91,0

Table 5. Mean μ and standard deviation σ of selected mel-frequency cepstral coefficients c_i.

5.2 Gender Identifiers

Two sex recognition approaches were used in our test. The first approach was based on an individual cepstral coefficient. Applying an empirical formula to the coefficient c_{24} we get the gender identifier D_{24} in the form

$$D_{24} = |c_{24} - 80| - |c_{24} - 40| - 120 + 2c_{24} \tag{5}$$

This indicator gives a negative value for male and a positive value for female speakers. The second approach used a set of selected cepstral coefficients according to the Table 5. For both sex classes the reference mean vectors were formed as follows:

Male reference: **M** = [-326, 338, 28, 94, 21, 31, 20, -35, -1, 5,…]
Female reference: **F** = [-597, 164, 182, -20, 129, 112, 109, 37,…]

and the Euclidean distances d_1 and d_2 were calculated in each test

$$d_1(\mathbf{X},\mathbf{M}) = [(\mathbf{X}-\mathbf{M})^T (\mathbf{X}-\mathbf{M})]^{1/2} \tag{6}$$

$$d_2(\mathbf{X},\mathbf{F}) = [(\mathbf{X}-\mathbf{F})^T (\mathbf{X}-\mathbf{F})]^{1/2} \quad (7)$$

where **M** and **F** denote the reference vectors mentioned above, **X** is the tested vector formed by the same coefficients c_i as the reference vectors, and T denotes transpose. Computing the difference of the two distances

$$D = d_1(\mathbf{X},\mathbf{M}) - d_2(\mathbf{X},\mathbf{F}) \quad (8)$$

we get a measure which gives similar polarity result as the identifier D_{24} (negative for male, positive for female).
Both procedures described above were evaluated for Czech vowel phonemes, which provided an identification accuracy of more than 90%. Especially for vowel "a" almost no error occurs. Table 6 shows the recognition rate obtained for all individual vowels cut out from a normally spoken speech.

Identifier	Test Vowel				
	a	e	i	o	u
D_{24}	99	92	97	93	91
D	99	94	98	92	94

Table 6. Gender recognition rate in percent testing all individual vowels.

The used speech data consisted of 420 sentences in total, 5 sentences by each of the 84 speakers (53 male and 31 female). All speakers in the database were subjective in good physical and psychical condition and have no speech, language or hearing difficulties. Most of the speakers are student aged 20 to 25 years. All speakers are Czech natives speaking with standard accent. The speakers were not informed of the objectives of the study before the experiment. The speech signal was sampled at 22 kHz using a 16-bit A/D converter under usual conditions in an office room.

6. Applications of Automatic Speaker Recognition

Law enforcement and military security authorities were among the first to make use of speaker recognition technology. The first type of machine speaker recognition using spectrograms of their voices, called voiceprint analysis or visible speech, was begun in the 1960s. The term voiceprint was derived from the more familiar term fingerprint. Voiceprint analysis was only a semiautomatic process. First, a graphical representation of each speaker's voice was created. Then, human experts manually determined whether two graphs represented utterances spoken by the same person. The graphical representations

took one of two forms: a speech spectrogram or a contour voiceprint (Baken & Orlikoff, 2000). The more commonly used form consists of a representation of a spoken utterance in which time is displayed on the horizontal axis, frequency on the vertical axis and spectral energy as the darkness at a given point.
At present, the increase in commercial application opportunities has resulted in increased interest in speaker recognition research. The main commercial application for speaker recognition seems to be speaker verification used to the physical entry of a person into a secured area, or the electronic access to a secured computer file or licensed databases. Such voice-based authorization is often a part of a security system that also includes the use of PIN number, password, and other more conventional means. The most immediate challenge in voice-based authorization is a caller authentication over the telephone network that will be accurate enough so that financial transactions could take place under it aegis. Car access is yet another popular area where voice-based security systems are gained ground. Some automobile manufactures are testing a speaker identification system to control door locks and ignition switches. An interesting twist to this application is that the ignition switch can be programmed not to work if the driver is under the influence of drugs or alcohol, since intoxication is detectable in the speech signal.

7. Acknowledgement

This work was supported by the Czech Ministry of Education in the frame of the Research Plan No. MSM 0021630513 "Advanced Electronic Communication Systems and Technologies".

8. References

Baken, R. J. & Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice,* Singular Publishing Group, ISBN 1-56593-869-0, San Diego

Kepesi, M. & Sigmund, M. (1998). Automatic recognition of gender by voice, *Proceedings of Radioelektronika'98*, pp. 200-203, ISBN 80-214-0983-5, Brno, April 1998, CERM, Brno

Lancker, D.; Kreiman, J. & Emmorey, K. (1985). Familiar Voice Recognition: Patterns and Parameters - Recognition of Backward Voices. *Journal of Phonetics*, Vol. 13, No. 1, (January 1985), pp. 19-38, ISSN 0095-4470

Matsui, T.; Nishitani, T. & Furui, S. (1996). Robust methods of updating model and a-priori threshold in speaker verification, *Proceedings of IEEE Internat. Conf. on Acoustics, Speech and Signal Processing*, pp. 97-100, ISBN 0-7803-3192-3, Atlanta, May 1996, IEEE Computer Society, Washington, DC

Rabiner, L. R. & Juang, B. H. (1993). *Fundamentals of Speech Recognition*, Englewood Cliffs, ISBN 0-13-015157-2, New Jersey

Reich, A. & Duke, J. (1979). Effects of selected vocal disguises upon speaker identification by listening. *Journal of the Acoustical Society of America*, Vol. 66, No. 4, (April 1979), pp. 1023-1028, ISSN 0162-1459

Sigmund, M. & Jelinek, P. (2005). Searching for phoneme boundaries in speech signal, *Proceedings of Radioelektronika 2005*, pp. 471-473, ISBN 80-214-2904-6, Brno, April 2005, MJ Servis, Brno

Sigmund, M. & Mensik, R. (1998). Estimation of vocal tract long-time spectrum, *Proceedings of Elektronische Sprachsignalverarbeitung*, pp. 69-71, ISSN 0940-6832, Dresden, September 1998, w.e.b. Universitätsverlag, Dresden

Slomka, S. & Sridharan, S. (1997). Automatic gender identification under adverse conditions, *Proceedings of Eurospeech'97*, pp. 2307-2310, ISSN 1018-4074, Rhodes, September 1997, Typoffset, Patras

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America*, Vol. 85, No. 4, (April 1989), pp. 1699-1707, ISSN 0162-1459

Wu, K. & Childers, D. G. (1991). Gender recognition from speech. *Journal of the Acoustical Society of America*, Vol. 90, No. 4, (April 1991), pp. 1828-1840, ISSN 0162-1459

Frontiers in Robotics, Automation and Control

Edited by Alexander Zemliak

ISBN 978-953-7619-17-6
Hard cover, 450 pages
Publisher InTech
Published online 01, October, 2008
Published in print edition October, 2008

This book includes 23 chapters introducing basic research, advanced developments and applications. The book covers topics such us modeling and practical realization of robotic control for different applications, researching of the problems of stability and robustness, automation in algorithm and program developments with application in speech signal processing and linguistic research, system's applied control, computations, and control theory application in mechanics and electronics.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen