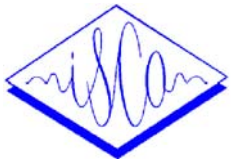

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

SPEECH SIGNAL PARAMETRIZATION FOR SPEAKER RECOGNITION UNDER VOICE DISGUISE CONDITIONS

Wojciech Majewski and Grażyna Mazur-Majewska

Institute of Telecommunications and Acoustics, Wroclaw University of Technology,
Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
majewski@aipsa.ita.pwr.wroc.pl

ABSTRACT

An experiment was performed to find out, if any of commonly applied techniques of speech signal parametrization is particularly resistant to voice disguise. As experimental material three vowels extracted from the word "logarytm" /logarɨtm/ spoken 10 times by each of 10 speakers under seven different speaking conditions were used. Three methods of parametrization were tested: FFT, LPC and ZCR. The results of the experiments indicated that the smallest intraspeaker variations were obtained for ZCR parameters, LPC provided reasonably good results, while FFT parameters were very seensitive to voice disguise and provided the worst results. Generally, however, it has to be stated that the experiments performed did not indicate explicitly which method of parametrization is particularly resistant to voice disguise.

Keywords: voice disguise, signal parametrization, intraspeaker variations

10.21437/Eurospeech.1999-287

1. INTRODUCTION

Recognition of speakers, based on their utterances, have many different applications. Among the most important are the forensic applications. In this type of application, there are, however, some special problems concerning the speaker; often they try to disguise their voices or mimic a voice of some other speaker [1, 2]. Disguised speech may be typically found in situations, when a blackmailer or kidnapper wants to cover his identity and thinks his voice is being recorded.

Regardless of the method utilized for speaker recognition, i.e. whether it is aural-perceptual, visual or automatic one [3], voice disguise has a deteriorating effect on speaker recognition scores. To cope with this problem in case of automatic speaker recognition, an experiment was performed to find out, if any of commonly applied techniques of speech signal parametrization is particularly resistant to voice disguise. This feature may be evaluated on the basis of intraspeaker distances between parameter vectors for a variety of disguise techniques utilized by a given speaker. The smaller the intraspeaker distances, the more resistant to voice disguise given parametrical representation of speech should be.

2. EXPERIMENTAL PROCEDURE

As phonetic material a key word "logarytm" /logarɨtm/ (Eng. logarithm) was selected, since it contains three vowels considered as good carriers of individual voice features and has already proven its usefulness in speaker recognition experiments [4]. This word was spoken 10 times by each of 10 speakers (eight males and two females; age 24-25 years) under seven different conditions, i.e. in natural mode and under following voice disguise conditions:

- with a pinched nose,
- through a handkerchief,
- utilizing whisper,
- with a pencil between the front teeth,
- with a small ball in the mouth,
- using free manner of disguise.

The recordings were made in a computer laboratory by means of a recording system consisting of a microphone and an IBM PC with AD/DA card. The signal from the microphone was low-pass filtered with 3.5 kHz cut-off frequency (what simulated the telephone band), sampled at a rate of 10 kHz and digitized with 12-bit resolution. From each utterance a stable segment of each of three vowels of 51.2 ms duration was taken. The window length was 256 samples with 128 samples shift. The parameter vectors were taken from 25.6 ms frames.

Three method of parametrization were tested:

- FFT (amplitude spectra in 16 one-third octave bands),
- LPC (12 linear prediction coefficients),
- ZCR (distribution of time intervals in 16 time channels).

To evaluate the intraspeaker variations under different speaking conditions the distances between

the parametrical representations of particular vowels were taken. For FFT and ZCR vectors Euclidean and Camber distances were utilized, while for LPC vectors Euclidean and Itakura distances were used. Before the distances were calculated, the parameter values were normalized to the maximal value equal to 1. As a reference vector the mean vector averaged over 10 repetitions of the utterance for normal speaking mode of a particular speaker was taken. As test vectors for a given speaker the mean vectors for particular manners of voice disguise were applied. Besides of distance calculations between the mean vectors, the distributions of parameters within particular speaking mode of particular speakers were evaluated on the basis of the distances between the mean vector for particular speaking mode and the vectors representing each of the 10 repetitions.
The experimental procedure was executed within the diploma work [5] supervised by the first author of the present paper.

3. RESULTS

An example of the influence of voice disguise on the waveforms of three vowels under investigation is presented in *Figure 1*.
An example of intraspeaker distances, both within a given method of voice disguise and between the mean vector for given method of voice disguise and the mean vector for normal speaking mode, is presented graphically in *Figure 2*. As may be seen from the data presented in this figure the smallest intraspeaker distances for speaker no 1 and the vowel /a/ were found for LPC parameters, good results were obtained for ZCR parameters and the worst results were obtained for FFT parameters.

Table 1. Evaluation of intraspeaker distances

Parameter	LPC			ZCR			FFT		
Vowel	a	o	ɨ	a	o	ɨ	a	o	ɨ
Distance	E	I	E	I	E	I	E	C	E	C	E	C
Speaker 1	0	+	–	+	0	0	–	0	–	+	–	+	–	–	0	0	0	–
Speaker 2	+	+	+	+	+	0	–	+	–	0	–	+	0	–	–	–	0	–
Speaker 3	0	+	0	–	–	0	–	+	–	+	0	+	+	–	–	–	+	–
Speaker 4	–	+	–	+	–	–	0	+	0	+	+	+	+	–	+	–	+	–
Speaker 5	0	+	0	0	0	0	0	0	0	+	0	+	–	–	–	–	–	–
Speaker 6	0	+	–	0	–	–	0	+	–	+	0	+	0	–	–	–	+	–
Speaker 7	0	–	0	0	0	0	0	+	+	+	0	+	–	–	–	–	–	–
Speaker 8	0	+	0	0	0	0	0	+	–	+	+	+	–	–	–	–	–	–
Speaker 9	–	0	0	0	–	0	–	+	–	+	–	–	+	–	–	–	+	–
Speaker 10	0	0	0	0	0	+	+	+	–	0	+	+	+	–	+	–	+	–

Distances: E – Euclidean, I – Itakura, C – Camber

The intraspeaker distances presented graphically for each method of parametrization under given measuring conditions were visually evaluated by the authors of this paper as small ("+"), moderate ("0") or large ("-"). The summarized results of these judgements are presented in Table 1. The number of pluses in particular columns of this table indicates how given parametric representation of examined vowels is resistant to voice disguise by examined group of speakers.

4. CONCLUSIONS

The smallest intraspeaker variations were obtained for ZCR parameters, what indicates that this method of parametrization is well suited to represent the speakers under voice disguise conditions. It is, however, to remember that this type of parameters is sensitive to the influence of noise, what was evident in case of voice disguise by means of whisper. LPC parameters with Itakura distance measure provided good results for the vowel /a/ and moderate results for the other two vowels. FFT parametrization provided the worst results, what demonstrates that these parameters are particularly sensitive to voice disguise.
Generally, however, it has to be stated, that the experiments performed did not indicate explicitly, which method of parametrization is particularly resistant to voice disguise. To practically evaluate which parametrical representation of speech is the best for automatic speaker recognition under voice disguise conditions, the experiments of speaker recognition should be carried out. In such a case, however, the final results would depend not only on the selection of the parameters, but also on many other factors (e.g. recognition algorithm) and the influence of the parameters would be obscured.

5. REFERENCES

[1] Hollien, H. (1990), *The Acoustics of Crime: The New Science of Forensic Phonetics*, New York: Plenum Press.
[2] Masthoff, H. (1996), A report on voice disguise experiment, *Forensic Linguistics*, 3(1), pp. 160-5.
[3] Majewski, W. and Basztura, C. (1996), Integrated approach to speaker recognition in forensic applications, *Forensic Linguistics*, 3(1), pp. 50-64.
[4] Majewski, W. and Basztura, C. (1998), Parametrical patterns of voices in forensic applications, *Archives of Acoustics*, 23(4), pp. 463-79.
[5] Suś, D. (1998), *Parametrization of speech signal for recognition of disguised voices*, Master's thesis, Wroclaw Univer. of Technology (in Polish).

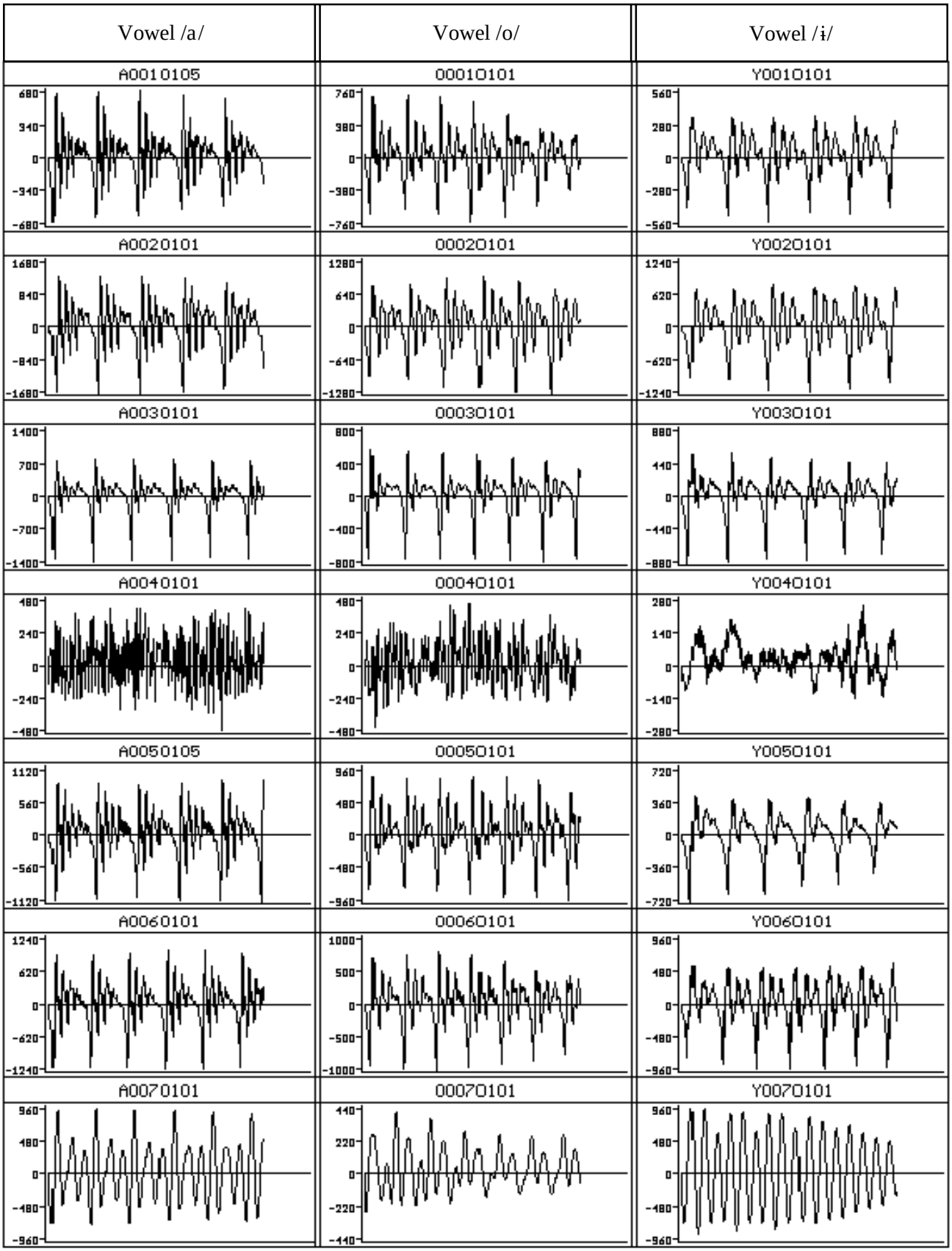

Figure 1. An example of the influence of voice disguise on waveforms of three vowels under investigation for speaker no 4 (from the top to the bottom: normal speech, pinched nose, through handkerchief, whisper, pencil between teeth, ball in the mouth, free manner of disguise – in this case a resonant chamber in front of the mouth).

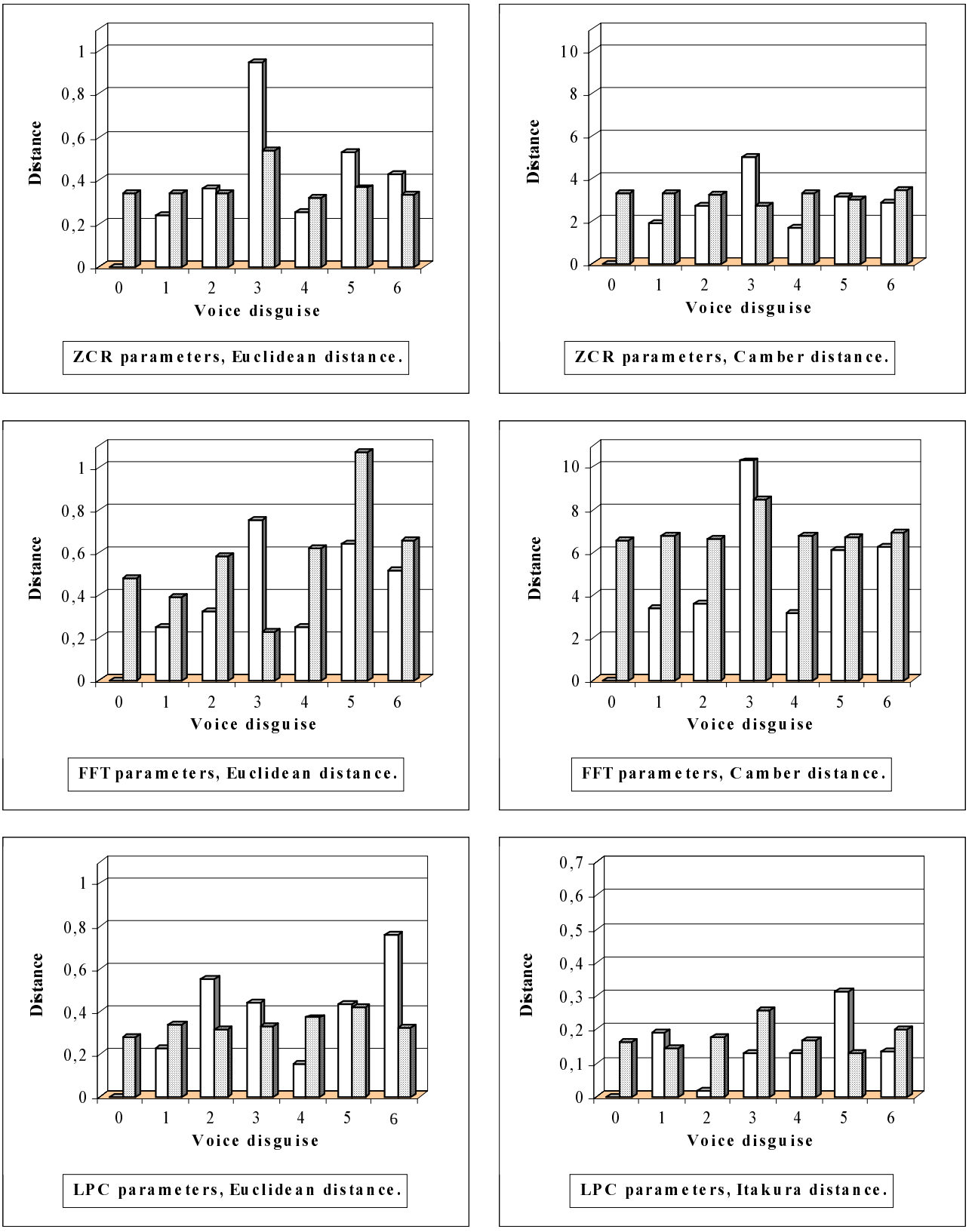

Figure 2. Intraspeaker distances for speaker no 1 and vowel /a/:
□ – distance between the mean vectors for normal and disguised speech.
⁘ – mean distance between the mean and individual vectors for given method of speech productions.
0 – normal speech, 1 – pinched nose, 2 – through handkerchief, 3 – whisper, 4 – pencil between teeth, 5 – ball in the mouth, 6 – free manner of disguise : the mouth covered with hand.