ARCHIVES OF ACOUSTICS
31, 4, 445–449 (2006)

MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC) OF ORIGINAL SPEAKERS AND THEIR IMITATORS

W. MAJEWSKI

Wrocław University of Technology
Institute of Telecommunications, Teleinformatics and Acoustics
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: wojciech.majewski

(*received June 15, 2006; accepted September 30, 2006*)

The results of intra- and interspeaker distances between MFCC vectors obtained from speech samples of eight well-known Polish personalities and their imitations performed by cabaret entertainers are presented and discussed. The intraspeaker distances between MFCC vectors representing normal and disguised speech samples of 10 speakers are also presented. The analysis of the measurement results indicated that, utilizing Euclidean distance between MFCC vectors, it is possible to differentiate the original speakers from the imitators. On the other hand, the MFCC vectors cannot be used to confirm the speaker's identity in the case of voice disguise.

Key words: MFCC, voice disguise, voice imitation.

1. Introduction

Recognition of speakers based on their utterances have many different applications, among the most important is a forensic application. In this type of application there are, however, some special problems concerning the speaker; often they try to disguise their voices or imitate a voice of some other speaker. In the case of voice disguise speakers attempt to change their voices in such a way that they are not recognized. Imitation can be considered an extreme form of speaker disguise in which speakers not only attempt to alter their voices in such a manner that they can no longer be recognized, but deliberately manipulate their voice so that they will be mistaken as the voices of some other people.

In forensic applications voice disguise and voice imitation constitute a problem for speaker recognition, since they may reduce its accuracy and usefulness. There is a number of reports on speaker recognition under voice disguise conditions [3, 7], but only single reports are available that describe experiments under voice imitation conditions [2, 8]. Some experiments on aural-perceptual speaker recognition under voice imitation conditions have been recently carried out by the author [5]. In that paper it has been shown that the imitators are able to "fool" the listeners, i.e. to convince them that they hear the original speaker. It was therefore interesting to find out if this ability of the imitators will be reflected in the acoustical parameters of speech of original speakers and their imitators. The measurements of speaking fundamental frequency (F0) have shown that the imitators are able to repeat F0 of another speaker [6] and that the professional imitators are the best in realization of this task [2]. Thus, F0 is not the suitable parametric representation of a speech signal that would permit to differentiate the original speakers from their imitators. In the present study mel frequency cepstral coefficients (MFCC) [1] shall be examined. MFCC are widely used in present speech and speaker recognition systems and it was hoped that this parameter will constitute a better tool to differentiate the original speakers from the imitators because the imitators cannot directly influence the MFCC values.

2. Experimental procedure

In the experiments two data bases were used. The first data base consisted of speech samples of 10 speakers recorded during four recording sessions which were two months apart. In addition, some speech samples of these speakers were recorded under voice disguise conditions.

The second data base consisted of speech samples of well-known Polish personalities and their imitations performed by cabaret entertainers. As original voices the speech samples of Władysław Bartoszewski (the former Foreign Office minister), Władysław Gomułka (the former first secretary of the Polish United Workers Party), Wojciech Jaruzelski (the former chairman of the State Council and president), Tadeusz Mazowiecki (the former prime minister), Leszek Miller (the former prime minister), Janusz Rewiński (actor and cabaret performer), Jan Rokita (politician) and Lech Wałęsa (the former president) were utilized. The imitations of the voices of the above named speakers were performed by the following artists: Bolesław Gromnicki (imitations of Gomułka, Jaruzelski and Wałęsa), Jerzy Kryszak (imitations of Jaruzelski, Mazowiecki and Wałęsa), Adam Łabuński (imitations of Bartoszewski and Rokita), Waldemar Ochnia (imitation of Rewiński), Andrzej Zaorski (imitations of Jaruzelski, Mazowiecki and Wałęsa) and one anonymous imitator (imitation of Miller).The original speech samples and their imitations were obtained from the recordings available in the archives of radio and television.

The extraction of MFCC parameters was carried for single phonemes – vowels. The number of extracted MFCC was 12. The comparison of MFCC vectors was performed for the vowels in the same phonemic context. As a measure of the differences between two vectors under comparison, the Euclidean distance was applied.

The first data base was utilized to obtain some data on intraspeaker distances between MFCC vectors under normal and disguised conditions of speech production. As a test material a steady segment of the vowel [a] extracted from the word "san" was utilized.

The second data base was used to calculate the intraspeaker distances between MFCC vectors for the original speakers and the imitators and the interspeaker distances between the original speakers and their imitators. In this case different vowels were examined, but for a given pair: speaker – imitator the same vowel was used.

The experimental procedure was executed within the diploma work [4] supervised by the author of the present paper.

3. Results

The distribution of intraspeaker Euclidean distances between the MFCC vectors representing normal and disguised speech samples of 10 speakers included in the first data base are presented in Table 1. The intraspeaker distances were calculated between normal speech samples recorded in the first and second recording session (1–2), in the first and the third (1-33) and in the first and the fourth session (1–4). The intraspeaker distances between normal and disguised speech samples (n–d) are also presented. The mean values for particular speakers and the mean value averaged over speakers are also included. From the data presented in this table it may be seen that the intraspeaker distances exhibit quite a large dispersion depending upon the speaker and speech samples under comparison. The overall mean intraspeaker distance averaged over the speakers is 154 for the normal mode of speech production and 325 for the distances between normal and disguised speech samples. These data provide some information as to what range of intra- and interspeaker distances may be expected while comparing speech samples produced by original speakers and their imitators.

Table 1. Intraspeaker distances between MFCC vectors for normal speech samples (1–2, 1–3, 1–4) and for normal versus disguised speech samples (n–d).

Speaker number	1–2	1–3	1–4	Mean	n–d
1	159	157	164	160	321
2	137	109	213	153	287
3	75	155	202	144	275
4	215	116	128	153	451
5	90	160	194	148	293
6	80	60	120	87	404
7	108	151	–	129	313
8	264	13	–	138	97
9	77	152	–	114	–
10	228	408	–	318	484
Overall mean	143	148	170	154	325

Table 2. Intra- and interspeaker distances between MFCC vectors for original speakers and their imitators.

Speaker	Speech sample	Intraspeaker distance Original	Intraspeaker distance Imitation	Interspeaker distance Orig.–Imit.
Władysław Bartoszewski	był [bIw]	178	200	368
Władysław Gomułka	ień [jen’]	40	24	300
Władysław Gomułka	lat [lat]	27	52	313
Wojciech Jaruzelski	wat [vat]	148	85	367
Wojciech Jaruzelski	tel [tel]	92	97	426
Tadeusz Mazowiecki	jes [jes]	93	115	351
Leszek Miller	jes [jes]	123	88	502
Leszek Miller	rząd [Zo~t]	81	82	450
Janusz Rewiński	pan [pan]	188	231	299
Janusz Rewiński	tel [tel]	85	262	377
Jan Rokita	tak [tak]	143	68	311
Lech Wałęsa	wam [vam]	193	204	323
Lech Wałęsa	nas [nas]	210	227	335
Overall mean		123	133	363

In Table 2 the intra- and interspeaker distances between MFCC vectors for the original speakers and their imitators included in the second data base are presented. The mean intraspeaker distances for the original speakers and for the imitators are similar to the intraspeaker distances presented in Table 1. It is interesting to note that the mean values of interspeaker distances between the original speakers and their imitators are similar to the values of intraspeaker distances between normal and disguised speech samples presented in Table 1 (325 and 363, respectively), in spite of the fact that in the first case we are dealing with the same speakers, while in the second case the speakers are different.

4. Discussion and conclusions

The experiments performed provided some insight into the problem of the influence of voice disguise and voice imitation on intra- and interspeaker variations of MFCC vectors. The comparison of the intraspeaker distances for normal speech samples (Table 1) and for original speakers and imitators (Table 2) indicates that these distances are roughly in the same range and that they are almost three times smaller than the mean interspeaker distances between the original speakers and their imitators (Table 2) and the mean intraspeaker distances between normal and disguised speech (Table 1).

Thus, it is concluded that, utilizing Euclidean distance between MFCC vectors, it is possible to differentiate the original speakers from their imitators, in spite of the fact that the imitators were able to "fool" the listeners, i.e. to convince them that they were listening to the original speakers [5]. On the other hand, the MFCC vectors cannot be used to confirm the speaker's identity in the case of voice disguise because in such a case the intraspeaker distances between normal and disguised speech samples are much larger than the intraspeaker distances for the normal mode of speech production.

References

[1] BECCHETTI G., RICTTI L., *Speech recognition – theory and C++ implementation*, John Wiley & Sons, Chichester 1999.

[2] ERIKSSON A., WRETLING P., *How flexible is the human voice? A case study in mimicry*, Proc. Eurospeech, Rhodos, **2**, 1043–1046 (1997).

[3] KÜNZEL H. J., *Effects of voice disguise on speaking fundamental frequency*, Forensic Linguistics – Int. J. Speech, Language and the Law, **7**, 2, 149–179 (2000).

[4] MACIEJKO W., *Intra- and interspeaker differences in acoustical parameters of original speakers and their imitators* [in Polish], Master's Thesis, Wrocław University of Technology, 2005.

[5] MAJEWSKI W., *Aural-perceptual voice recognition of original speakers and their imitators*, Archives of Acoustics, **30**, 4 (Supplement), 183–186 (2005).

[6] MAJEWSKI W., *Speaking fundamental frequency of original speakers and their imitators*, Speech Analysis, Synthesis and Recognition, SASR Workshop, Kraków 2005 (CD Rom).

[7] MASTHOFF H., *A report on a voice disguise experiment*, Forensic Linguistics – Int. J. Speech, Language and the Law, **3**, 1, 50–64 (1996).

[8] SCHLICHTING F., SULLIVAN K. P. H., *The imitated voice – a problem for voice line-ups?* Forensic Linguistics – Int. J. Speech, Language and the Law, **4**, 1, 148–165 (1997).