

SPEAKING FUNDAMENTAL FREQUENCY OF ORIGINAL SPEAKERS AND THEIR IMITATORS

Wojciech MAJEWSKI

Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: wojciech.majewski@pwr.wroc.pl

(received October 16, 2006; accepted November 7, 2006)

The results of SFF measurements of well-known Polish personalities and their imitations performed by cabaret entertainers are presented and discussed. The difference in F0 between the original speakers and their imitators was in some cases few Hz only, but in some other cases it reached more than 100 Hz.

Keywords: pitch, original voice, voice imitation.

1. Introduction

Recognition of speakers based on their utterances has many different applications, among the most important are forensic applications. In this type of application there are, however, some special problems concerning speakers; they often try to disguise their voices or imitate the voice of other speaker [1]. In the case of voice disguise speakers attempt to change their voices in such a way that they are not recognized. According to SCHLICHTING and SULLIVAN [2] imitation can be considered as 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. Thus, more research is needed to find out how serious the problem is. There is a number of reports on speaker recognition under voice disguise conditions [3–6], but only few reports are available that describe experiments under voice imitation conditions [2, 7]. Some experiments on aural-perceptual speaker recognition under voice imitation conditions have been recently carried out by the author and presented elsewhere [8]. 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 is therefore interesting to find out if this ability

of the imitators will be reflected in the acoustical parameters of speech of the original speakers and their imitators. In the present paper the experiments on the comparison of selected acoustical parameters of original speakers and their imitators will be presented. To be more specific, speaking fundamental frequency (SFF or F_0) as one of the most important parameters in aural-perceptual and automatic methods of speaker recognition [9] will be examined.

2. Experimental procedure

The test material consisted of speech samples of well-known Polish personalities and their imitations performed by cabaret entertainers. As the original voices the speech samples of Władysław Bartoszewski (the former Minister of Foreign Affairs), Władysław Gomułka (the former first secretary of the Polish United Workers Party), 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 and Wałęsa), Jerzy Kryszak (imitations of Mazowiecki and Wałęsa), Adam Łabuński (imitations of Bartoszewski and Rokita), Waldemar Ochnia (imitation of Rewiński), Andrzej Zaorski (imitations of Mazowiecki and Wałęsa) and one anonymous imitator (imitation of Miller). Thus, there were three imitations of Wałęsa, two imitations of Mazowiecki and one imitation for each of the remaining speakers. Generally, the semantic contents of the original speech samples and their imitations were different, but some phrases uttered by Bartoszewski, Gomułka and Miller were the same for the original speakers and their imitators.

The original speech samples and their imitations were obtained from the recordings available in the archives of radio and television.

The performed experiments consisted of two parts. In the first part, the values of SFF for selected vowels [u, o, a, e, I] in the middle and at the end of the words were measured. In the second part, mean F_0 values for the same phrases uttered by three original speakers and their imitators were measured.

The experimental procedure was executed within the diploma work [10] supervised by the author of the present paper. The measurements were carried out by Multi-Speech system produced by Kay Elemetrics Corp.

3. Results

The results of SFF measurements for the selected vowels are presented in Table 1. From the data presented in this table it may be seen that the values of F_0 for particular vowels may be quite different for original speakers and their imitators and only in the minority of cases these values are similar. The differences in SFF between the original speakers and their imitators are presented in Table 2. In this table it may be seen that

Table 1. SFF in Hz for original speakers and their imitators for examined vowels.

Speech samples	Vowels (m – in the middle, e – at the end)								
	[a] m	[a] e	[e] m	[e] e	[o] m	[o] e	[ɪ] m	[ɪ] e	[u] m
Bartoszewski – original	189	231	238	195	228	186	199	228	186
Bartoszewski – imitator AŁ	226	183	179	198	211	194	240	205	220
Gomułka – original	208	245	174	198	180	169	88	139	169
Gomułka – imitator BG	186	185	230	163	140	191	199	168	158
Mazowiecki – original	121	95	111	96	130	160	137	85	146
Mazowiecki – imitator JK	252	137	231	189	212	158	210	149	211
Mazowiecki – imitator BG	156	197	123	162	157	195	182	121	155
Miller – original	156	127	84	126	129	100	132	130	132
Miller – anonymous imitator	182	153	187	194	193	98	167	101	195
Rewiński – original	252	215	237	244	255	241	178	168	205
Rewiński – imitator WO	244	208	256	201	156	179	217	185	198
Rokita – original	137	187	132	122	159	154	154	144	212
Rokita – imitator AŁ	152	158	187	154	153	142	168	179	169
Wałęsa – original	191	155	140	146	124	153	198	184	150
Wałęsa – imitator BG	195	174	229	150	209	199	187	168	199
Wałęsa – imitator JK	221	258	230	192	222	213	188	194	198
Wałęsa – imitator AZ	197	201	220	221	130	211	201	200	233

Table 2. Absolute difference in SFF in Hz between original speakers and their imitators for examined vowels.

Speech Samples	Vowels (m – in the middle, e – at the end)								
	[a] m	[a] e	[e] m	[e] e	[o] m	[o] e	[ɪ] m	[ɪ] e	[u] m
Bartoszewski vs. imitator AŁ	37	48	59	3	17	8	41	23	34
Gomułka vs. imitator BG	22	60	56	35	40	22	111	29	11
Mazowiecki vs. imitator JK	131	42	119	93	82	2	74	64	65
Mazowiecki vs. imitator BG	35	102	12	66	28	35	45	36	9
Miller vs. anon. imitator	26	25	103	68	64	2	35	29	63
Rewiński vs. imitator WO	8	7	19	43	99	62	39	17	7
Rokita vs. imitator AŁ	15	29	55	32	6	12	14	35	43
Wałęsa vs. imitator BG	4	19	89	4	85	54	11	16	49
Wałęsa vs. imitator JK	30	103	90	46	98	60	10	10	48
Wałęsa vs. imitator AZ	6	46	80	75	6	58	3	16	83

these differences may reach the values exceeding 100 Hz ([I]m for Gomułka and his imitator, [a]m and [e]m for Mazowiecki and imitator JK, [a]e for Mazowiecki and imitator BG, [e]m for Miller and his imitator, [a]e for Wałęsa and imitator JK). On the other hand, there is a substantial number of cases, where these differences are very small, in the range of few Hz (e.g. [e]e for Bartoszewski and his imitator, [o]e for Mazowiecki and Miller and their imitators, [I]m for Wałęsa and imitator AZ, etc.). Thus, from the data presented in Tables 1 and 2 it is hard to judge if F_0 values for original speakers and their imitators are similar or not. More definite information may be obtained from the data presented in Table 3. In this table the mean values of SFF for all examined vowels for the original speakers and their imitators are presented. The differences between F_0 for the original speakers and their imitators are not large (for Bartoszewski, Gomułka and Rokita it is only a few Hz difference). The differences are generally negative in value, what indicates that the imitators were speaking with higher F_0 than the original speakers. Probably the imitators wanted to achieve this way a proper artistic impression.

In the subsequent part of the experiment, the mean values of F_0 were measured for the same words uttered by three original speakers and their imitators. The test material consisted of the utterances of Leszek Miller (words: “rządu” [Zondu], “rokity” [rokitI], “premier” [premjer], “pola” [pola], “pani” [pani]), Władysław Gomułka (words: “demonstracji” [demonstratsji], “dłatego” [dłatego], “dzień” [dz'en'], “kobiet” [kobjet], “zamętu” [zamentu], “ułatwić” [uwatvits']) and Władysław Bartoszewski (words: “dłaczego” [dlatSego], “ja” [ja]). In the selection of words a care was taken to extract them from the utterances of similar emotional context in order to minimize the influence of additional factors on the values of F_0 . Nevertheless, the results of the measurements shown in Table 4 indicate a substantial dispersion between the F_0 values obtained for original speakers and their imitators. The minimal difference was 10.8 Hz for the word “zamętu” uttered by Gomułka and his imitator. The maximal difference was 72.9 Hz for the word “rokity” produced by Miller and his imitator.

Table 3. Mean values of SFF in Hz for original speakers and their imitators averaged over the examined vowels.

Speech samples	Original	Imitator	Difference
Bartoszewski vs. imitator AŁ	208.9	206.2	2.7
Gomułka vs. imitator BG	174.4	180.1	-5.7
Mazowiecki vs. imitator JK	120.1	194.3	-74.2
Mazowiecki vs. imitator BG	120.1	160.8	-40.2
Miller vs. anon. imitator	124.0	163.3	-39.3
Rewiński vs. imitator WO	221.7	204.9	16.8
Rokita vs. imitator AŁ	155.7	162.4	-6.7
Wałęsa vs. imitator BG	160.1	190.0	-29.9
Wałęsa vs. imitator JK	160.1	212.9	-52.8
Wałęsa vs. imitator AZ	160.1	201.6	-41.5

Table 4. Mean values of SFF in Hz for original speakers and their imitators for examined words.

Speech samples	Original	Imitator	Difference
Bartoszewski “dlaczego”	172.5	211.0	-38.5
Bartoszewski “ja”	179.1	236.2	-57.1
Gomułka “demokracji”	198.1	183.3	14.8
Gomułka “dlatego”	206.9	175.7	31.2
Gomułka “dzień”	220.4	202.1	-18.3
Gomułka “kobiet”	193.6	163.1	30.5
Gomułka “zamętu”	162.3	173.1	-10.8
Gomułka “ułatwić”	223.5	163.4	60.1
Miller “rządu”	104.0	129.7	-25.7
Miller “rokity”	81.5	154.4	-72.9
Miller “premier”	64.7	126.4	-61.7
Miller “pola”	123.4	146.4	23.0
Miller “pani”	110.6	138.4	-27.8

4. Discussion

The experiments performed resulted in SFF values of substantial dispersion. For the vowels extracted from continuous speech the difference in F_0 between the original speakers and their imitators reached in some cases more than 100 Hz, but in some other cases it was only a few Hz (see Table 2). This difference, averaged over all the examined vowels (see Table 3), is generally smaller (only few Hz for Bartoszewski, Gomułka and Rokita), but for the worst case (Mazowiecki vs imitator JK) it reaches 74 Hz. A similar spread of F_0 values may be observed for the same test material (see Table 4), thus the phonemic content of the examined utterances is not the main reason of the differences in SFF between the original speakers and their imitators.

It is necessary to stress that in the experiments the speech samples of the cabaret entertainers have been used. Thus, it was not a professional imitation, but some kind of entertainment in which the imitator wanted to obtain a proper artistic impression, reflecting the peculiarities of the original speaker such as his speaking tempo or articulation habits, paying less attention to the pitch of the imitated person. Some kind of a proof that such an attitude may be successful in the imitation of a voice of some other person is given in Table 5, taken out from some other studies [8] carried out by the author with the same group of original speakers and their imitators as in the present study. In this table the results of aural-perceptual voice recognition are presented. Looking at the data presented in Table 5 it may be seen that the imitators were quite successful in convincing the listeners that they are listening to the original speakers – for the speech samples of 4 seconds duration 38.3% of the listeners evaluated the imitations as the originals. It is also interesting to note that the listeners never recognized the original voices with 100% accuracy.

Table 5. Listeners' answers in percentage averaged over speakers.

Speech samples	Original	Imitation	Imitator	Other	Nothing
Original – 4 sec. duration	91.3	2.3	0	2.9	3.5
Original – 30 sec. duration	95.5	1.1	0	0.9	2.5
Imitation – 4 sec. duration	38.3	35.8	3.1	11.3	11.5
Imitation – 30 sec. duration	32.5	45.5	7.1	5.6	9.3

At this point it is interesting to quote the results of the experiments carried out by ERIKSSON and WRETILING [7]. They examined SFF of three Swedish politicians (Bildt, Loket and Parnevik) and the corresponding imitations. In this case, however, professional imitators have been employed. The imitators were working exclusively for the purposes of the experiment and they were doing their best to repeat faithfully the voices of the politicians, including their pitch. The results of that experiment are presented in Table 6. The presented results indicate that a professional imitator is able to repeat accurately the SFF values of the original speaker.

Table 6. Mean F_0 in Hz of Swedish politicians and their imitators [7].

Speech samples	Bildt		Loket		Parnevik	
	Mean F_0	Deviation	Mean F_0	Deviation	Mean F_0	Deviation
Original	134	4.11	146	3.41	85	3.76
Imitator	134	3.49	142	3.23	89	3.63
Difference	0	–	4	–	4	–

5. Conclusions

The result of the present study and the studies carried out by other investigators indicate that it is possible to repeat fundamental frequency characteristics of a given speaker by another speaker and that professional imitators are the best in realization of this task. On the other hand, it is interesting to note that even if an imitation of SFF characteristics of particular speakers by imitators is not perfect, they are able to successfully fool listeners, i.e. to convince them that they are listening to original speakers.

References

- [1] HOLLIEN H., *Forensic Voice Identification*, Academic Press, London 2002.
- [2] 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).
- [3] HOLLIEN H., MAJEWSKI W., DOHERTY E. T., *Perceptual identification of voices under normal, stress and disguise speaking conditions*, *Journal of Phonetics*, **10**, 139–148 (1982).

-
- [4] KÜNZEL, H. J., *Sprecher-Erkennung: Grundzüge forensischer Sprach-verarbeitung*, Kriminalistik-Verlag, Heidelberg 1987.
- [5] MASTHOFF H., *A report on a voice disguise experiment*, Forensic Linguistics – Int. J. Speech, Language and the Law, **3**, 1, 50–64 (1996).
- [6] MAJEWSKI W., *Aural-perceptual speaker verification under voice disguise conditions*, [in:] Proceedings of Subjective and Objective Assessment of Sound, CD Rom, Poznań 2004.
- [7] ERIKSSON A., WRETTLING P., *How flexible is the human voice? A case study of mimicry*, [in:] Proceedings of ESCA – Eurospeech 97, Rhodes, **2**, 1043–1046 (1997).
- [8] MAJEWSKI W., *Aural-perceptual voice recognition of original speakers and their imitators*, Archives of Acoustics, **30**, 4 (Supplement), 183–186 (2005).
- [9] 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).
- [10] GIGOŁŁA M., *Acoustical parameters of voices of original speakers and their imitators* [in Polish], Master's Thesis, Wrocław University of Technology, 2003.