

# THE INBUILT RECORDER OF MOBILE PHONES – POSSIBILITIES OF FORENSIC SPEAKER IDENTIFICATION

Agata TRAWIŃSKA, Mateusz KAJSTURA

**ABSTRACT:** The study presents an experiment concerning recordings made by means of the inbuilt recorder of mobile phones. The recordings were investigated mainly from the point of view of forensic speaker identification. Thus, the recorded speech was examined with a view to finding a way of carrying out auditory speech analysis (especially description of a speaker's learned articulatory patterns) and also assessing the value of acoustic speech parameters (amongst other things, extraction of the structure of formants  $F_1$ – $F_3$ , which reflects both the anatomical characteristics of the vocal tract, such as size and shape and the speaker's articulatory habits).

Some procedures for securing recordings made on mobile phone inbuilt recorders were proposed. Characteristic signal-markers were observed in incoming calls recorded by means of some mobile phone inbuilt recorders. These signals-markers may constitute an additional source of information about the continuity of a recording and also may indicate the probable make of the telephone used as a recorder.

**KEY WORDS:** Forensic speaker identification; Recording function in mobile phone; Acoustic and phonetic analysis; Reference material; Securing procedures.

*Problems of Forensic Sciences, vol. LVII, 2004, 51–80*  
*Received 10 December 2003; accepted 12 May 2004*

## INTRODUCTION

Among recordings submitted by courts, prosecution or the police for speaker identification analysis, many are recordings of telephone calls concerning, for example, extortion of a ransom for a stolen car [3, 8]. Speech and Audio Analyses laboratories are frequently also asked to secure audio data which may be stored in the memory of mobile phones, such as the contents of the voice mailbox, voice sms or memos and then carry out relevant examinations [9, 14]. A well-known, and at the same time well-researched phenomenon linked to telephone transmission of speech is the occurrence of characteristic distortions of speech signals such as bandwidth reduction, low dynamic range and additive noise picked up and transmitted by the handset microphone [11]. In the case of GSM standard mobile telephony, the whole process of speech coding, the sort of speech coders used, the system of coding protection and further steps of signal processing have been strictly de-

efined [6]. The impact of cellular coding on the performance of speaker verification systems working both under commercial and forensic conditions has been widely studied in published research [1, 2, 4, 5]. A less well-known issue is the processing of the speech signal that occurs at the receiver circle i.e. by the base station and the particular mobile phone. Signal processing in the receiver circle is also not defined directly in the GSM standard and depends to a large extent on the parameters of the used equipment [6].

The recording function available on many mobile phones seems a very useful and simple way to make recordings of incoming calls or direct speech. In this way, an evidence event could be recorded too. The possibility of using the cellular phone as a recorder of evidence statements prompted us to analyse the usefulness of recordings made in this way – using speech and audio examinations, and especially examinations of the speaker’s identification. Our goal was to investigate:

- the technical capabilities of mobile phone recorders, including the best way to transfer the audio signal in order to secure it as material for identification purposes;
- usefulness of reference material recorded in a direct way in relation to measurable acoustic parameters of speech of a person whose speech has been recorded by means of the inbuilt recorder of a mobile phone (both direct speech and incoming calls).

## EXPERIMENT

### Equipment and reference material

The reference material for the speaker identification experiment was recorded on a portable digital recorder HHB MiniDisc (MDP 500) by means of a Schoeps cardioid microphone (CCM-40). This set enables high quality recordings in the transmission range 10–20 000 Hz. The evidence recordings were made on inbuilt recorders on the following mobile phones: Nokia 6310, 6510, 8310, Siemens ME45, SL45, S55, Motorola Accompli A008 and Ericsson T68. Additionally, Nokia 8210, Panasonic EB-GD67, PSTN telephone (Public Switched Telephone Network), and an ISDN telephone equipped with an answering machine were used.

Figure 1, presents an articulatory description of six Polish vowels, for which an analysis of measurable speech parameters was carried out (structure of formants  $F_1$ – $F_3$ ) [7, 13].

The possibility of carrying out an articulatory description of speech for recordings stored in the memory of a cellular phone was assessed on the basis of analysis of the following statements:

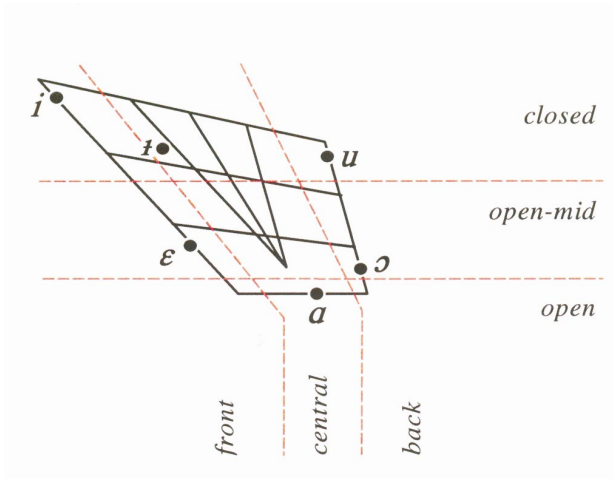


Fig. 1. System of six standard Polish vowels.

/na.jvɛ · ʃɛ.t̪ɨ.nʃɛ.sɔ.ɔ̃.pɔ.zna.ru/

Najwyższe czynsze są w Poznaniu (The highest rents are in Poznań).

/j.a.k.a.p.j.e.n.k.n.a.d͡z.i.ɔ.pɔ.gɔ.d.a/

Jaka piękna dziś pogoda (What a beautiful weather it is today).

/m.u.g.w̃.miu.zɛ.n.d͡n.i.k.pɔ.mu.t̪.s/

Mógł mi urzędnik pomóc (The clerk could have helped me).

/s.t̪.ʃɛ.m.p.j.a.s.tɛ.li.ɔ.t̪.ɛ.s.a.w.a.t̪.ɛ/

Strzępiaste liście sałaty (Lacinated leaves of lettuce).

/t̪.ʃɛ.d͡nɔ.ɛ.s.wɔ.n̪.t̪.sɛ.p.a.z̪ɛ.wɔ.pɔ.m.a.r.a.n̪.t̪.ʃɛ/

Trzydniowe słońce sparzyło pomarańcze (The sun lasting for three days burnt the oranges).

/a.bɛ.ɔ.vɔ.pɔ.jɔ.n̪.t̪.ɔ.ʃɛ.b.a.ɔ.v.l.u.b.i.t̪.ɔ/

Aby owo pojąć, trzeba owo lubić (To comprehend it, one should like it).

/ʃɛ.d.w.z.j.a.p.w.cɛ.m.v.rɛ.n̪.t̪.sɛ/

Szedł z jabłkiem w ręce (He was walking with an apple in his hand).

/t̪.ʃɛ.ṽ.jɛ.ɔ̃.w̃.zɛ.r.u.mɔ.z.n.a.mu.vi.d͡z̪.ɔ.rɛ.sɔ.t̪.s.j.a.li.za.t̪.s.ji/

Czy w więzieniu można mówić o resocjalizacji? (Is it justifiable to talk about a successful resocialisation in prison?)

In the presented part of the project, both isolated vowels and continuous statements were uttered by the same person, i.e. a woman aged 26. The fundamental frequency distribution ( $F_0$ ) of the speaker voice, registered directly on the MiniDisc, is presented below:  $F_{0\text{ mean}} = 205$  Hz;  $F_{0\text{ min}} = 180$  Hz;  $F_{0\text{ max}} = 272$  Hz.

The mean frequencies of the first three formants gained for steady-state parts of six Polish vowels and their relative standard deviations are presented in Tables I and II.

TABLE I. MEAN FREQUENCIES ESTIMATED FOR FORMANTS STRUCTURE ( $F_1$ - $F_3$ ) OF SIX POLISH VOWELS RECORDED ON MINIDISC HHB MDP 500 (REFERENCE MATERIAL)

Formant	Mean frequency [Hz]					
	i	ɨ	ɛ	α	ɔ	u
$F_1$	358.3	454.7	630.9	742.8	636.8	450.6
$F_2$	2640.6	1914.9	1775	1320.3	990.3	825.5
$F_3$	3346.2	2737.9	2737.9	2624.2	2455.5	2566.4

TABLE II. RELATIVE STANDARD DEVIATIONS ESTIMATED FOR FORMANTS STRUCTURE ( $F_1$ - $F_3$ ) OF SIX POLISH VOWELS RECORDED ON MINIDISC HHB MDP 500 (REFERENCE MATERIAL)

Formant	Relative standard deviation [%]					
	i	ɨ	ɛ	α	ɔ	u
$F_1$	8.81	7.42	5.63	6.46	7.69	8.33
$F_2$	2.66	5.57	5.85	4.76	8.69	5.18
$F_3$	4.73	4.03	3.17	5.82	5.49	4.18

Each analysed statement was simultaneously recorded on a cellular phone recorder as evidence material and directly on a MiniDisc digital recorder as reference material. The material collected in this way allowed us to analyse the influence of the examined algorithms coding the speech signal, recorded on the inbuilt recorder of the chosen cellular phone and the parameters of the telephone itself (including the characteristics of the band and the microphone) on the comprehensibility and acoustic parameters of speech.

## EXAMINATION PROCEDURES

First of all, before the main investigation, frequency responses of tested mobile phone recorders were examined. To analyse bandwidth and band characteristics for each of the investigated inbuilt recorders, pink noise was used. Pink noise was induced through active loudspeakers HHB Circle 5A and recorded on the inbuilt recorders of the tested mobile phones.

The broadest bandwidth for which it was possible to record pink noise was observed for Siemens mobile phone recorders (200–4000 Hz for Siemens SL45 in the case of audio signals transferred by digital connections, i.e. IrDA or com). The narrowest bandwidth was obtained for Ericsson (i.e. 300–3600 Hz in the case of audio signals transferred by acoustic coupling with FM radio transmission). The outcomes are graphically presented in Figure 2.

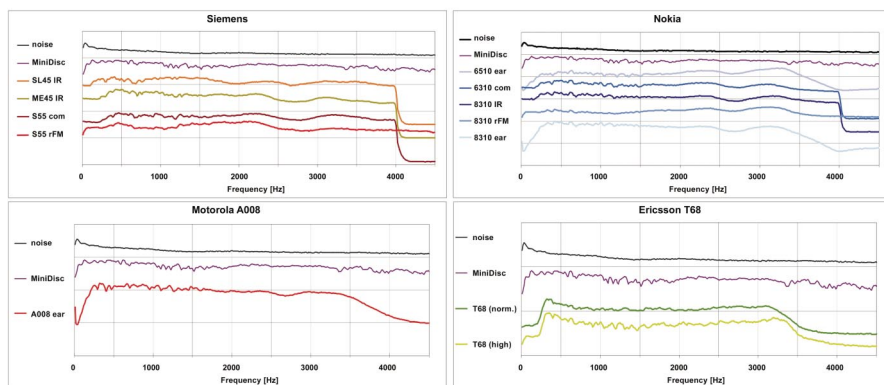


Fig. 2. Band characteristics depending on manner of acquisition of audio signal for inbuilt recorders of investigated mobile phones. For the Ericsson mobile phone, band characteristics are presented depending on the defined quality of recording: normal or high.

A significant similarity was noticed between the characteristics of the Motorola A008 inbuilt recorder and Nokia mobile phone inbuilt recorders (making up a cohesive group). An important attenuation within the range 2500–3100 Hz was observed for all investigated mobile phones except for the Ericsson T68.

Essential differences were observed in the band characteristics of various Siemens models. For particular Siemens inbuilt recorders the characteristics varied more than for inbuilt recorders of other telephone makes. Obviously, the way of transferring the recordings of the pink noise from the inbuilt recorder influenced the bandwidth and band characteristics. Acoustic coupling with FM radio transmission, which is used to secure recordings from answering machines, telephones and faxes, seems to be inadequate for forensic speaker identification purposes. Substantial distortions caused by acoustic coupling mean that one has to be very careful when interpreting the outcomes (low correlation with direct recordings on MiniDisc).

It should be mentioned that the examined inbuilt recorders also differed in terms of maximum duration of recordings, for instance, from 30 seconds for Ericsson through 2–3 minutes for the majority of mobile phones to about 5 hours for the SL45. The examined mobile phone inbuilt recorders also differ in terms of technical possibilities of audio signal securing: Ericsson allows transfer of audio signal by acoustic coupling only, Nokia 6510 and Motorola A008 – by earphone, Nokia 8310 and all Siemens models – by IrDA and com connections (IrDA and com connections require appropriate software to exchange data between the mobile phone device and the PC), Nokia 8310 – by both earphones and IrDA connections.

The Ericsson inbuilt recorder was the only one offering a choice of quality of recording – normal (lasting up to 107 seconds) and high (up to 41 seconds; see Figure 2). However, no significant differences between so-called normal recordings and high quality ones were observed.

In the presented part of the undertaken research, the main focus was on so-called static parameters of speech. Hence, instead of describing the dynamics of articulatory activities (i.e. instead of the formants tracking method) we concentrated on average values extracted from the steady-state part of vowels. Mean values of  $F_1$ – $F_3$  formants for the manually segmented steady-state part of vowels were extracted by means of Praat software, version 4.0.51, both for reference material and material recorded on inbuilt recorders of mobile phones. Moreover, values of all formants obtained for vowels recorded on inbuilt recorders were each time verified against the characteristics of the band of the given mobile phone inbuilt recorder. In the experiment we tested different types of recordings made on inbuilt recorders of mobile phones, namely:

- direct speech recordings;
- incoming calls from mobile phones;
- incoming calls from PSTN telephone;
- incoming call (from PSTN) recordings made on answering machine of ISDN phone.

## RESULTS AND DISCUSSION

### Acoustic analysis

#### *Direct speech recordings*

Direct recordings made on a diverse group (in terms of, among other things, band characteristics) of inbuilt recorders of mobile phones, i.e. Siemens ME45, SL45, S55 and Nokia N6310, N8310 were transferred by IrDA/com connections. Linear correlation coefficients values relating to

these recordings and reference materials directly recorded on a MiniDisc HHB were high and ranged from 0.978 for Siemens S55 to 0.930 for Nokia 8310.

Detailed  $F_1$ – $F_3$  formants measurements obtained for recordings made on dictating machines with respect to relevant reference values extracted for particular vowels are presented in Table III. The ranges of intra-speaker variability for formants of six vowels with mean values for all reference recordings are also presented in this table.

TABLE III. CHANGE IN FORMANTS FREQUENCIES MEASURED FOR RECORDINGS MADE ON DIFFERENT DEVICES, OBTAINED BY IrDA/com CONNECTIONS

Formant	Vowel	ME45 [%]	SL45 [%]	S55 [%]	N6310 [%]	N8310 [%]	Intra-speaker variability [%]
$F_1$	i	38.7	6.5	34.7	40.9	16.5	8.81
	ɪ	29.2	2.1	−0.6	−5.6	−1.4	7.42
	ε	4.5	4.5	7.9	−5.2	10.8	5.63
	α	2.1	7.9	11.0	−0.4	6.7	6.46
	o	2.5	21.9	2.6	−7.0	7.2	7.69
	u	9.6	6.8	−1.9	−9.5	6.0	8.33
$F_2$	i	−1.4	−3.4	1.2	−1.0	1.8	2.66
	ɪ	4.2	−3.3	1.9	7.7	−3.2	5.57
	ε	17.1	−1.9	8.3	10.0	1.7	5.85
	α	−8.5	5.5	6.3	−9.0	3.1	4.76
	o	6.5	8.2	−22.6	7.8	−7.2	8.69
	u	−11.4	20.7	−	9.7	11.6	15.18
$F_3$	i	−4.1	−5.2	1.0	−4.5	0.8	4.73
	ɪ	1.5	0.1	6.2	4.2	−3.1	4.03
	ε	3.3	−4.2	6.4	3.6	−5.9	3.17
	α	4.3	−5.7	−6.8	−19.3	−3.5	5.82
	o	−11.7	−3.5	5.9	−9.4	−8.3	5.49
	u	−0.7	−0.8	−2.9	−9.7	−2.1	4.18

Taking into consideration intra-speaker variability, one may notice tendencies common to all direct speech recorded on the inbuilt recorder of various mobile phones. Namely, the average value of the first formant,  $F_1$ , deter-

mined for the test recordings was higher than the value gained for the reference material. Furthermore, the value of the third formant,  $F_3$ , for vowels / $\alpha$ / and / $\text{ɔ}$ / was lowered relative to the value determined for comparative vowels.

For direct recordings made on the inbuilt recorder of Nokia 6510 and Motorola A008 (the inbuilt recorders have similar band characteristics) and secured via the earphone connection, the linear correlation coefficients values relative to the reference material were high and amounted to 0.993 for Nokia 6510 and to 0.994 for Motorola A008. Beside high linear correlations of both materials, an important tendency of an increase in the  $F_1$  formant value for the high vowels /i/ and /ɪ/ was noticed. Further analysis showed opposite tendencies in  $F_2$  especially for the / $\varepsilon$ /, / $\alpha$ / and / $\text{ɔ}$ / vowels (see Table IV). Thus it may be concluded that the obtained results for  $F_1$ – $F_3$  formants recordings transferred via earphones connection do not reveal a considerable common tendency, even if the inbuilt recorders show similar band characteristics.

TABLE IV. CHANGE IN FORMANTS FREQUENCIES MEASURED FOR RECORDINGS MADE ON DIFFERENT DEVICES, OBTAINED BY EARPHONES CONNECTION

Formant	Vowel	N6510 [%]	A008 [%]	Intra-speaker variability [%]
$F_1$	i	12.7	33.8	8.81
	ɪ	10.5	12.7	7.42
	ε	3.4	-2.2	5.63
	α	14.9	2.1	6.46
	ɔ	11.8	1.2	7.69
	u	9.8	3.8	8.33
$F_2$	i	-0.3	0.2	2.66
	ɪ	6.2	-0.7	5.57
	ε	11.5	-6.6	5.85
	α	8.6	-3.9	4.76
	ɔ	10.9	-6.2	8.69
	u	-2.1	-4.1	15.18



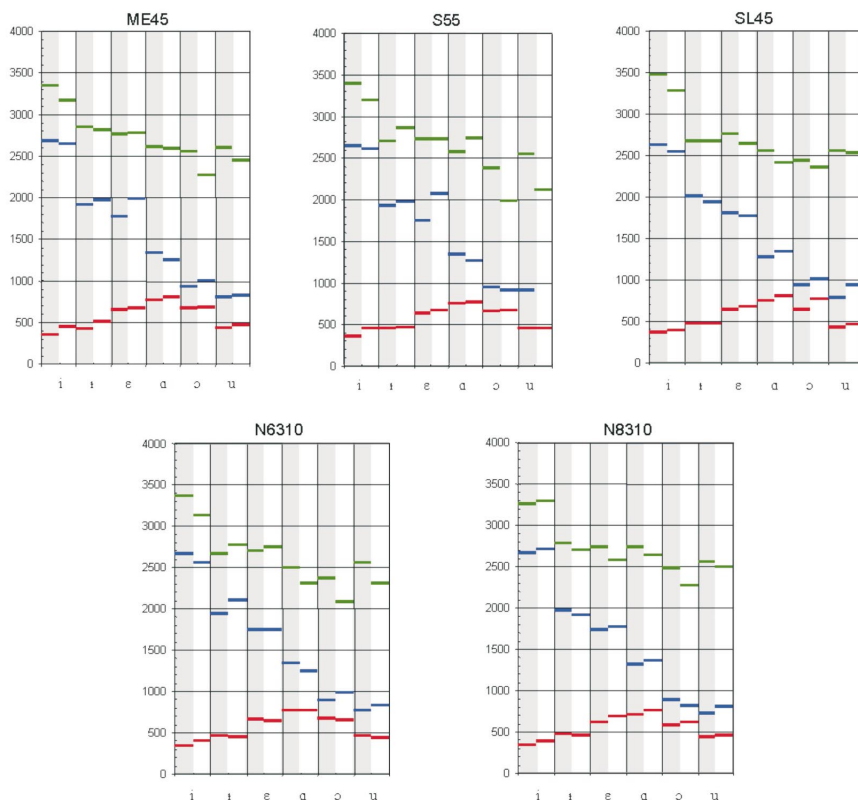


Fig. 3. Graphical presentation of changes in formants frequencies measured for recordings made by means of different devices, transferred by IrDA/com connections (grey background indicates reference material recorded on MiniDisc).

$F_3$	i	1.4	-1.8	4.73
	ɪ	1.3	1.5	4.03
	e	-0.4	-2.9	3.17
	a	1.7	0.9	5.82
	ɔ	2.6	-9.2	5.49
	u	-3.6	-1.4	4.18

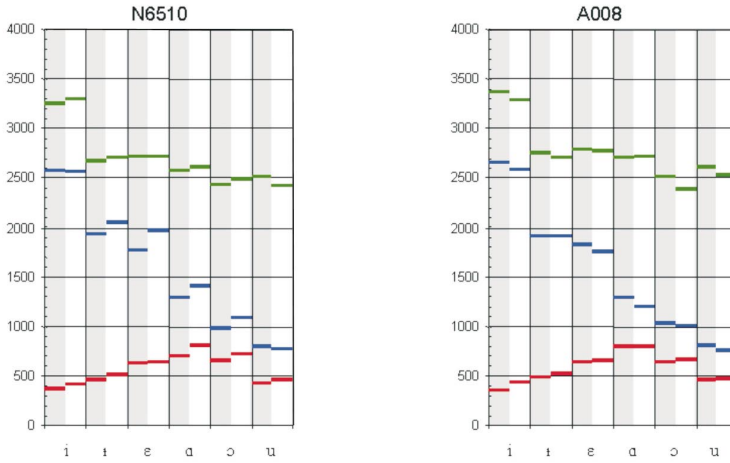


Fig. 4. Graphical presentation of changes in formants frequencies measured for recordings made by means of different devices, transferred by earphones connection (grey background indicates reference material recorded on MiniDisc).

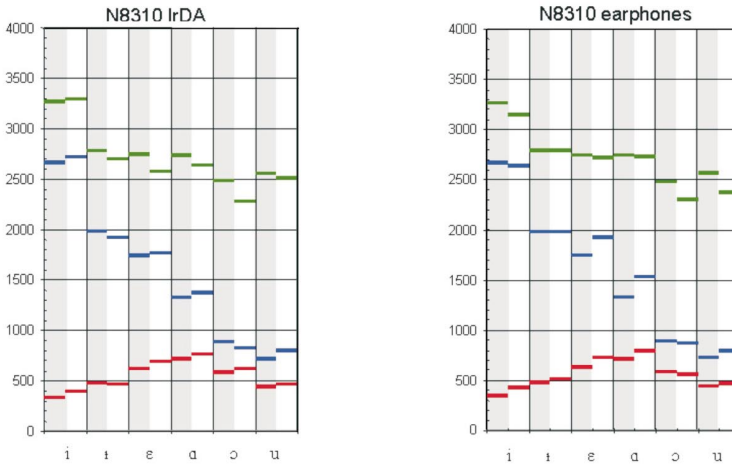


Fig. 5. Graphical presentation of change in formants frequencies measured for recordings made by means of N8310, transferred by IrDA/com and earphones connections (grey background indicates reference material recorded on MiniDisc).

The results for direct speech recorded on the inbuilt recorder of mobile phones depending on the means of securing the audio signal are presented in Figure 3 and Figure 4.

In order to verify results for direct speech recordings made on the inbuilt recorder of various mobile phones and different ways of transferring data,

i.e. IrDA/com and earphone connections, the results gained for the Nokia 8310 recordings transferred both by IrDA/com were analysed. The  $F_1$ – $F_3$  measurements for direct speech recorded on the inbuilt recorder of the Nokia 8310 and transferred in both ways are presented in Figure 5.

The accuracy of the formants structure ( $F_1$ – $F_3$ ) determined for direct recordings made on the inbuilt recorder of the Nokia 8310 relative to the structure determined for the reference material recorded on the HHB MiniDisc reveals better correlation for the recordings transferred by IrDA/com than by the earphone connection. It is also worth mentioning that compared with the reference material, the changes in the mean values of the formants  $F_1$  and  $F_3$  for the majority of vowels extracted from the recordings secured with IrDA/com are less than intra-speaker variability; in the case of  $F_3$  this tendency relates to all vowels.

To sum up, the results obtained for direct speech recordings made on the inbuilt recorder of mobile phones and transferred by IrDA/com connections were adequate relative to the reference material recorded directly on the HHB MiniDisc. Moreover, a transfer by IrDA/com connections makes the transmitted data less dependent on the make and model of the mobile phone than in the case of transfer of data by earphone connection.

### ***Incoming calls from mobile phones***

In order to test acoustic parameters of speech for incoming calls recorded on inbuilt recorders, calls were made from the Nokia 8210 to three phones, i.e.: Siemens ME45, Siemens S55 and Nokia 6310 and from the Panasonic EB-GD 67 to the Siemens S55. To obtain reference material, the speaker's speech was simultaneously recorded directly on an HHB MiniDisc recorder. All incoming calls registered by means of the inbuilt recorder of Siemens ME45, Siemens S55 and Nokia 6310 were transferred by IrDA/com connections.

Good correlation between the reference material and incoming calls recorded on the inbuilt recorders was observed (the linear correlation coefficients were high) and, furthermore, the make of the mobile phone from which the call was made did not influence the extracted parameters. The main factor affecting the measured formants frequencies was the make of the inbuilt recorder. For incoming calls from mobile phones, the observed tendencies were similar, i.e. the value of the  $F_1$  formant was shifted up for most vowels, the value of the  $F_2$  formant was shifted down for / $\alpha$ / and the value of the  $F_3$  formant was shifted down for / $\text{o}$ / and / $\text{i}$ /. The results for incoming calls from mobile phones recorded on the inbuilt recorder of three selected mobile phones and transferred by IrDA/com connection are presented in Figure 6 and Figure 7.

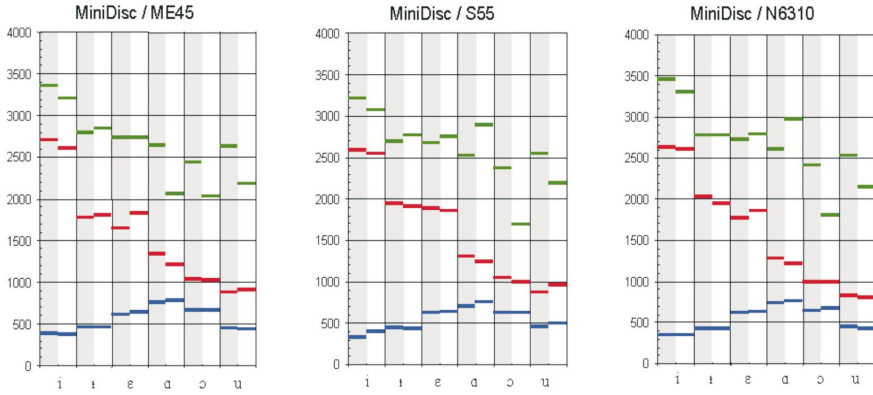


Fig. 6. Graphical presentation of change in formants frequencies measured for recording of incoming GSM calls (from N8210 to different telephones) in comparison with reference material (on grey background).

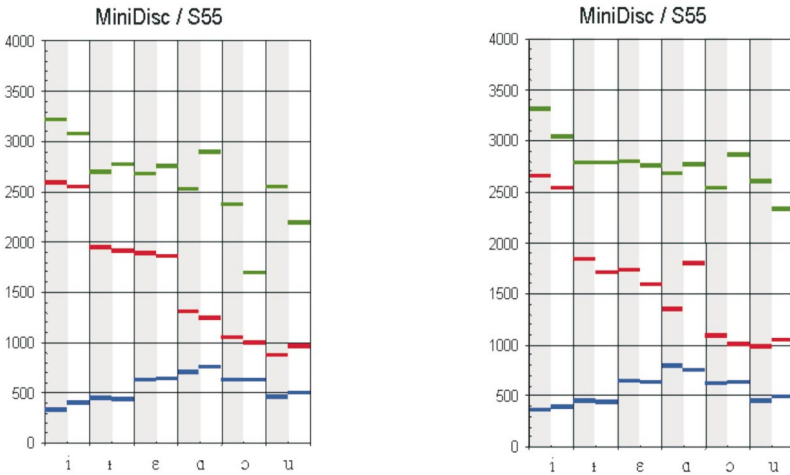


Fig. 7. Graphical presentation of change in formants frequencies measured for recording of incoming GSM calls (on left from N8210, on right from EB-GD 67) in comparison with reference material (on grey background).

### ***Incoming calls from PSTN telephone***

Incoming calls from the PSTN were recorded on the inbuilt recorder of four mobile phones (Siemens S55, Motorola A008, Nokia 6310, 6510). The same PSTN telephone was used for all tests.

Poor correlation between the reference material and the recordings obtained from the inbuilt recorders was observed. If the IrDA/com connection was used to transfer recorded audio signal, the secured data depended on

particular inbuilt recorders to such an extent that it was impossible to point to any common or crucial tendencies for changes in the values of the first three formants.

For the recordings secured by an earphone connection, the formants structure measurements showed only slight deviations from the reference material, being less than intra-speaker variability. Moreover, for PSTN incoming calls recorded on two similar inbuilt recorders (i.e. Nokia 6510 and Motorola A008) and secured with an earphone connection, similar tendencies were observed: for  $F_1$  the value of the vowel /ɔ/ formant rises, for  $F_2$  the value of the vowel /α/ formant falls and for  $F_3$ , the value of the vowel /i/ formant falls too (see Table V).

TABLE V. CHANGES IN FORMANTS FREQUENCIES MEASURED FOR RECORDINGS OF INCOMING PSTN CALLS MADE ON DIFFERENT DEVICES, OBTAINED BY DIFFERENT MEANS

Incoming calls from PSTN		By IrDA/com		By earphones connection		Intra-speaker variability [%]
		N6310 [%]	S55 [%]	N6510 [%]	A008 [%]	
$F_1$	i	-9.4	1.2	2.0	15.6	8.81
	ɛ	-3.7	-5.5	5.9	5.2	7.42
	ε	1.4	-2.4	2.7	5.2	5.63
	α	-0.6	4.5	14.0	-0.7	6.46
	ɔ	-3.1	4.7	20.8	9.5	7.69
	u	-9.1	12.6	6.3	0.8	8.33
$F_2$	i	-4.1	-2.9	-1.9	-6.2	2.66
	ɛ	8.3	2.4	2.0	0.2	5.57
	ε	10.1	6.8	-2.7	-0.8	5.85
	α	-19.1	-7.5	-4.2	-10.4	4.76
	ɔ	-1.0	10.3	8.9	3.2	8.69
	u	-34.8	24.8	0.9	-8.1	15.18
$F_3$	i	-9.9	-5.5	-7.8	-3.3	4.73
	ɛ	6.2	1.2	0.4	-5.1	4.03
	ε	3.3	3.2	0.7	2.3	3.17
	α	7.5	8.4	4.4	-0.5	5.82
	ɔ	22.2	-16.9	-1.2	-1.4	5.49
	u	-17.5	-11.6	1.5	-7.4	4.18

A worse linear correlation was achieved between recordings of incoming calls from the PSTN made on inbuilt recorders of varying band characteristics (compared with the connection set described above).

The obtained and discussed results above for the first three formants show that incoming calls from the PSTN recorded on the inbuilt recorder of mobile phones should be secured with earphone connections since such recordings manifest similar tendencies for particular investigated parameters and tend to have a better correlation of results relative to the reference material, i.e. the direct speech recorded on the MiniDisc.

### ***Incoming calls (from PSTN) recorded on the digital answering machine of an ISDN phone***

A call made from a PSTN phone was recorded as a message on the digital answering machine of an ISDN telephone. This message was simultaneously recorded on a MiniDisc as reference material. Next, the message recorded on the answering machine of the ISDN telephone was played back and recorded on the MiniDisc using the earphones connection of the ISDN telephone. Furthermore, the message stored on the answering machine of the ISDN device was remotely accessed and recorded on the Siemens ME45 inbuilt recorder.

The linear correlation coefficients values for both ways of securing the material from the ISDN answering machine were high and slightly better for the recording transferred by the earphones connection of the ISDN device. It was also observed that the differences in the mean values of the formants for both investigated recordings were less than intra-speaker variability. The few examples of substantial differences showed parallel tendencies: for  $F_1$  the value of the vowel /i/ formant rises significantly, for  $F_2$  the value of the vowel /α/ formant falls significantly and for  $F_3$  the value of the vowel /ε/ formant increases significantly. The only inconsistency in tendencies was noted for the  $F_3$  formant value of /α/; the  $F_3$  value of /α/ extracted from the recording transferred by earphones connection increases and decreases for the recording remotely accessed and recorded on the Siemens ME45 inbuilt recorder (see Table VI).

### **Phonetic analysis**

Auditory evaluation (comprehensibility and clarity) of collected recordings lead us to the conclusion that the recordings made on the inbuilt recorder of the Siemens S55 were of the worst quality. For instance, for the Siemens S55, with incoming calls recorded from a PSTN telephone: the vowel /u/ and some realizations of /i/ and /ɔ/ were interpreted as noise (see Figure 8).

Besides, for some realisations, a lack of opposition between vowels / $\alpha$ / and / $\text{ɔ}$ / was noticed. A high level of attenuation of speech sounds clarity – up to a complete lack of intelligibility – especially of hard voiceless consonants (such as / $f$ /, / $p$ /, / $t$ /, / $s$ /) and nasal ones (/n/, /m/), both in initial and middle parts of words, was observed (see Figure 9 and 10).

Distortion of speech was even worse in the case of background noise or speaker movement. Similar attenuation of consonantal clarity, particularly in initial sounds was also perceived in recordings on the Nokia 6310i and, to a lower degree, on the Siemens ME45. It should be added that time parameters, e.g. duration of vowels, should be carefully evaluated. In a few cases it was observed that vowels spoken in isolation recorded on inbuilt recorders were shorter than their counterparts extracted from reference material. Moreover, some interruptions appeared during vowels, as presented in the above spectrograms (see Figure 8).

TABLE VI. CHANGES IN FORMANTS FREQUENCIES MEASURED FOR RECORDINGS MADE ON ANSWERING MACHINE OF ISDN PHONE AND OBTAINED BY DIFFERENT MEANS

Incoming calls from PSTN	By earphones connection of ISDN [%]	By IrDA/com of ME45 [%]	Intra-speaker variability [%]	
Recorded on ISDN Answering machine				
$F_1$	i	15.6	6.2	8.81
	ɪ	1.3	0.2	7.42
	ɛ	1.0	-3.7	5.63
	α	3.4	0.8	6.46
	ɔ	3.1	1.1	7.69
	u	0.5	11.6	8.33
$F_2$	i	-2.4	0.8	2.66
	ɪ	-0.6	-2.2	5.57
	ɛ	-1.3	-3.5	5.85
	α	-5.3	-7.1	4.76
	ɔ	-1.0	-1.6	8.69
	u	0.5	-3.3	15.18

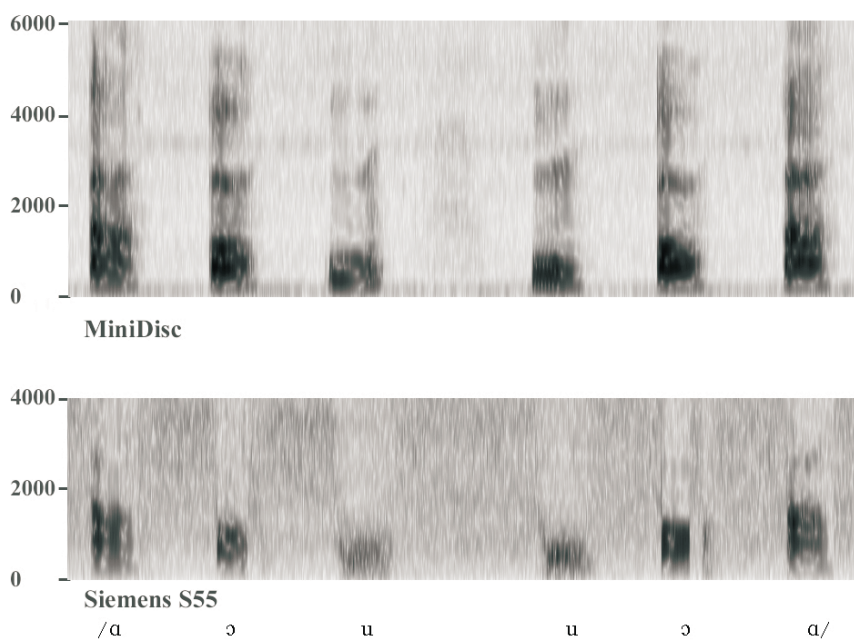


Fig. 8. Spectrogram of back vowels repeated two times.

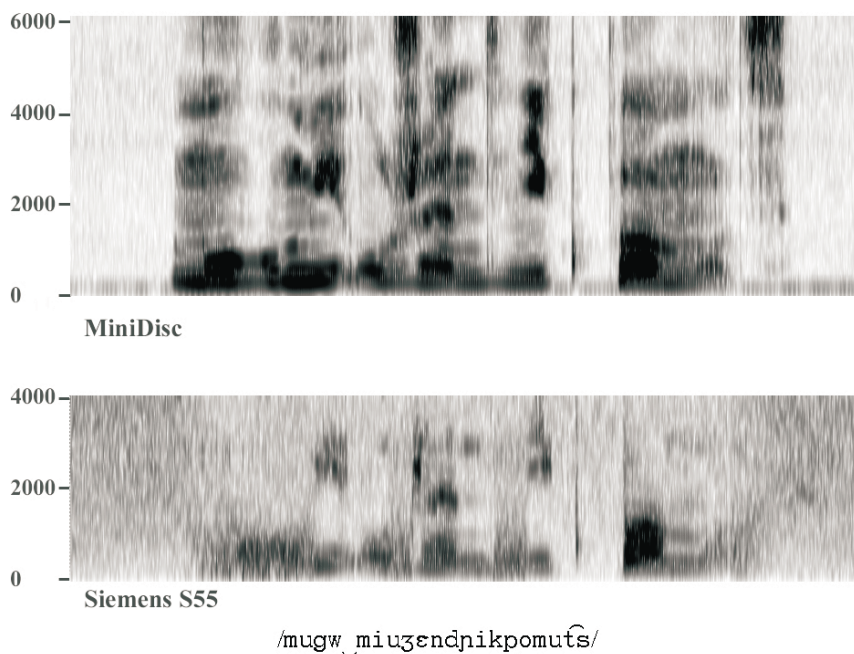


Fig. 9. Spectrogram of utterances: The clerk could have helped me.



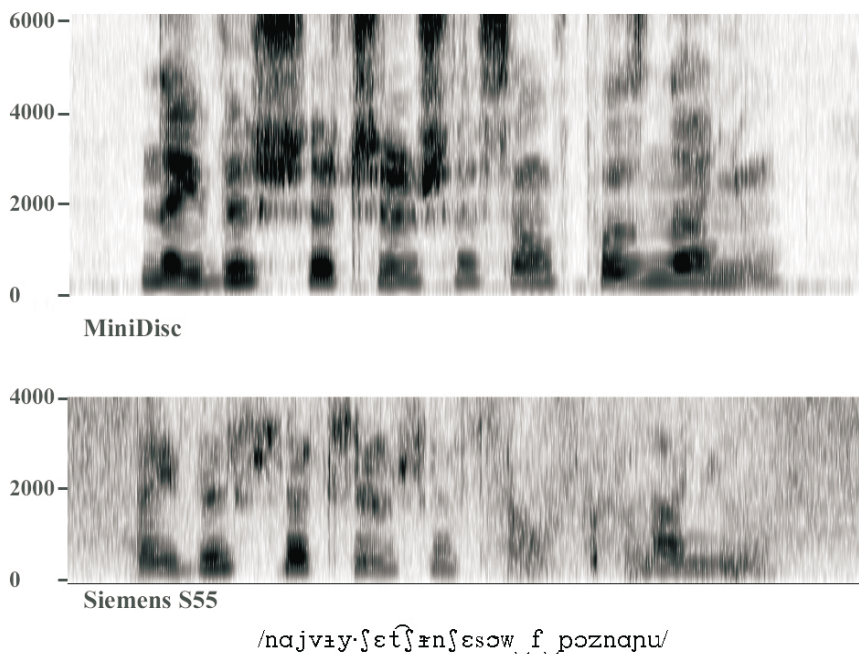


Fig. 10. Spectrogram of utterances: The highest rents are in Poznań.

Incoming calls from PSTN		By earphones	By IrDA/com	Intra-speaker
Recorded on ISDN		connection	of ME45	variability
Answering machine		of ISDN	[%]	[%]
		[%]		
$F_1$	i	15.6	6.2	8.81
	ɪ	1.3	0.2	7.42
	ɛ	1.0	-3.7	5.63
	α	3.4	0.8	6.46
	ɔ	3.1	1.1	7.69
	u	0.5	11.6	8.33
$F_2$	i	-2.4	0.8	2.66
	ɪ	-0.6	-2.2	5.57
	ɛ	-1.3	-3.5	5.85
	α	-5.3	-7.1	4.76
	ɔ	-1.0	-1.6	8.69
	u	0.5	-3.3	15.18

$F_3$	i	-3.3	0.6	4.73
	ɪ	3.5	9.5	4.03
	ɛ	5.8	-4.2	3.17
	α	2.8	-	5.82
	ɔ	0.2	-	5.49
	u	-	-	4.18

In the presented research, the analysed recordings were conducted under restricted conditions, i.e. in a soundproof room and using high quality equipment. The authors decide to exclude factors that are common in real forensic cases such as background noise or speaker movement; however the impact on speech quality of such factors in real-life conditions should be kept in mind.

## CONCLUSIONS

### **Proposed procedures for transferring audio data recorded by means of inbuilt recorder of mobile phones**

On the basis of results gained during this experiment it should be stated that the best possible correlation between material recorded on an inbuilt recorder of a mobile phone and reference speech recorded directly on a digital recorder of similar technical parameters to the HHB MiniDisc may be achieved when:

- direct utterances and also incoming calls from mobile phones recorded on the inbuilt recorder of mobile phones are secured using IrDA/com connections;
- incoming PSTN calls (and as a special case PSTN messages recorded on an ISDN answering machine) are secured by the audio method of acquisition, using the analogue earphones connection.

### **Identification process**

During identification examinations of a speaker, in accordance with the method of speaker identification used by the authors, considerable attention is paid to linguistic analysis of the speech of the person being identified, especially their articulatory habits. Only in the case of significant correlation of articulatory features, is it possible to carry out a reliable identification.

However, as has been noted in the "Experiment" section, even in the highly controllable conditions that were maintained in the experiment, substantial attenuation of the quality of speech recorded on the inbuilt record-

ers of mobile phones could be observed. The attenuation process encompasses both direct speech and incoming calls, affecting consonants and vowels to a varying extent. In real-life cases, higher degradation of the speech signal should be expected due to interference from the acoustic background or speaker movement. Due to such conditions, the speech is characterised by low clarity, leading to probable lack of understanding of some words. Then, linguistic – especially articulatory – analysis is particularly impeded. Good correlation of acoustic speech features with the reference material obtained in the experiment shows that measurable features may constitute additional, though insufficient, information in the process of speaker identification [2, 12]. In order to interpret investigation results properly, it is advisable to take into consideration the band characteristics of the evidence mobile phone.

It may be concluded that in real forensic cases there are insufficient reasons to record reference material on the same or similar inbuilt recorder that was used to record the speech of the speaker being identified, especially since information concerning the evidence inbuilt recorder is often unavailable.

There are no clear and stable acoustic speech parameters which are identically or at least similarly processed by all inbuilt recorders. What is more, especially for incoming calls recorded on the inbuilt recorders of mobile phones, the number of variables (such as the make and type of used devices, e.g. GSM, PSTN or ISDN) makes it impossible to re-create circumstances similar to those occurring when the evidence material was recorded. Besides, such reconstruction might be based on case details, such as, for example, the statement of the suspect and not directly on analysis of the recording parameters.

### **Additional observations for incoming call recordings**

In the course of the experiment it was noticed that some of the studied mobile phones (all Nokia models: 6510, 6310, 8310 and Motorola A008) periodically (usually every 5 or 10 seconds) registered a tone of definite frequency and given duration during the recording of incoming calls. The frequency of this signal-marker for Nokia phones was 1175 Hz and for Motorola phones 2500 Hz. The presence and frequency-time characteristics of such signals, in the authors' opinion, may be helpful in the investigation of continuity of such recordings and may be used as a marker for a particular make of mobile phones. It should be stressed, however, that the signal is audible to the speaker being recorded, so it excludes the possibility of secret recording of phone calls. No such signal was noticed in the incoming calls recorded on the inbuilt recorder of Siemens mobile phones.

The findings presented here are only preliminary due to the fact that the project is still ongoing and one has to be very careful drawing conclusions.

It is necessary to examine the same connection sets presented in the section “Examination procedures” for inbuilt recorders of other mobile phones, and also to broaden the group of speakers whose utterances are analysed. It should be also added that the authors observed the need for taking into consideration more than just so-called static parameters (i.e. mean values of formants extracted for the steady-state part of vowels): dynamic measurements of articulatory activity should also be undertaken.

#### References:

1. Boves I., Kuitert M., Speaker Verification with GSM coded telephone Speech, Proceedings of EUROSPEECH-97 Conference, Rhodes 1997, pp. 975–978.
2. Deutsch W. A., Moosmüller S., Formant frequency measurement on spectrally reduced speech signals, Abstracts IAFP Annual Conference, Vienna 2003, p. 18.
3. Dolecki J., Rzeszotarski J., Nagrania bankowych zleceń telefonicznych jako przedmiot badań fonoskopijnych, *Problemy Kryminalistyki* 2003, nr 240, s. 44–48.
4. Eskelinen-Rönkä P., Niemi-Laitinen T., The unknown effect of speech compression, Abstracts IAFP Annual Conference, Vienna 2003, p. 17.
5. Gazit R, Metzger Y., Toledo O., Speaker verification over cellular networks, Proceedings of a Speaker Odyssey – The Speaker Recognition Workshop, Crete 2001, pp. 125–128.
6. Hołubowicz W., Szwabe M., GSM – ależ to proste!, Wydawnictwo Holkom, Poznań 1999.
7. Jassem W., Podstawy fonetyki akustycznej, PWN, Warszawa 1973.
8. Kredens K., Góralewska-Łach G., Language as sole incriminating evidence: the Augustynek case, *Forensic Linguistics* 1998, vol. 5, pp. 1350–1371.
9. Künzel H. J., Beware of the “telephone effect”. The influence of telephone transmission on the measurement of formant frequencies, *Forensic Linguistics* 2001, vol. 8, pp. 80–99.
10. Künzel H. J., Rejoinder to Francis Nolan’s: The “telephone effect” on formants: a response, *Forensic Linguistics* 2002, vol. 9, pp. 83–86.
11. Meuwly D., Voice analysis, [in:], Encyclopedia of forensic science, Siegel J., Saukko P., Knupfer G. [ed.], Academic Press, London 2000.
12. Nolan F., The “telephone effect” on formants: a response, *Forensic Linguistics* 2002, vol. 9, pp. 74–82.
13. Wierchowska B., Fonetyka i fonologia języka polskiego, PWN, Warszawa 1980.
14. Vermeulen J., The need for speed, Abstracts IAFP Annual Conference, Vienna 2003, p. 16.

# FUNKCJA DYKTAFONU W TELEFONACH KOMÓRKOWYCH – MOŻLIWOŚCI IDENTYFIKACJI MÓWCY

Agata TRAWIŃSKA, Mateusz KAJSTURA

## WSTĘP

Wśród nagrań, odnośnie do których sąd, prokurator lub policja zlecają przeprowadzenie badań identyfikacyjnych mówiącego, wiele z nich to nagrania rozmów telefonicznych, w tym także prowadzonych za pośrednictwem telefonu komórkowego, np. przy wymuszeniach okupu za skradziony samochód [3, 8]. Laboratoria fonoskopijne proszone są także o procesowe zabezpieczenie nagrań, które mogą ewentualnie znajdować się w pamięci telefonu komórkowego, na przykład w postaci zgłoszeń przechowywanych w skrzynce poczty głosowej czy MMS-ów, a następnie o przeprowadzenie dalszych badań szczegółowych [9, 14].

Dobrze znanym, a zarazem szczegółowo opracowanym zagadnieniem związanym z transmisją telefoniczną, jest występowanie charakterystycznych zniekształceń sygnału mowy, które polegają m.in. na ograniczeniu pasma mowy czy zmniejszeniu dynamiki sygnału; niektóre ze zniekształceń sygnału użytecznego, jakie charakteryzują nagrania rozmów telefonicznych, wynikają z parametrów samego aparatu telefonicznego, zwłaszcza charakterystyki mikrofonu [11]. Dla telefonii komórkowej w standardzie GSM w sposób jednoznaczny został zdefiniowany cały proces obróbki i transmisji sygnału mowy [6]. Licznych opracowań doczekał się także wpływ kodowania mowy w telefonii komórkowej na rezultaty uzyskiwane przez automatyczne systemy weryfikacji tożsamości mówiącego na podstawie mowy, testowane zarówno w zastosowaniach komercyjnych, jak i w warunkach analizy kryminalistycznej [1, 2, 4, 5]. Jak dotychczas niewielką jednakże uwagę poświęcono procesom przetwarzania sygnału mowy, jakie następują na etapie odbioru sygnału m.in. przez stację bazową. Obróbka sygnału mowy na etapie odbioru nie jest także zdefiniowana w standardzie GSM i w dużej mierze może zależeć od rozwiązań zastosowanych w urządzeniach odbiorczych [6].

Funkcja nagrywania w aparatach telefonów komórkowych wydaje się być użytecznym a zarazem łatwym sposobem utrwalenia rozmów przychodzących, jak również wypowiedzi bezpośrednich. Za pomocą tej funkcji może zostać utrwalona także rozmowa, która następnie stanowi materiał dowodowy. Możliwość wykorzystania telefonu komórkowego jako rejestratora wypowiedzi dowodowych skłania do przeanalizowania przydatności utrwalonych w ten sposób nagrań z punktu widzenia badań fonoskopijnych, a zwłaszcza z badań identyfikacyjnych mówiącego. Ponadto w prezentowanym eksperymencie:

- zbadano możliwości techniczne aparatów telefonów komórkowych, a zwłaszcza zaproponowano najlepszy ze względu na dalsze badania fonoskopijne sposób zabezpieczania nagrań, utrwalonych w pamięci telefonu;
- określono przydatność materiału porównawczego zarejestrowanego w sposób bezpośredni pod względem parametrów mierzalnych mowy osoby, której wy-

powiedzi dowodowe zostały utrwalone za pośrednictwem dyktafonu telefonu komórkowego (zarówno wypowiedzi bezpośrednich, jak i rozmów przycho-  
dzących).

## EKSPERYMENT

### Urządzenia i materiał referencyjny

Materiał referencyjny w prezentowanych eksperymentalnych badaniach identyfikacyjnych mówcy stanowiły nagrania utrwalone za pomocą przenośnego rejestratora cyfrowego MiniDisc marki HHB (MDP-500) oraz mikrofonu o kardoidalnej charakterystyce marki Schoeps (CCM-40). Zestaw ten pozwala na utrwalenie nagrań o wysokiej jakości w zakresie częstotliwości 10–20 000 Hz. Nagrania dowodowe stanowiły zaś wypowiedzi utrwalone za pomocą dyktafonów następujących telefonów komórkowych: Nokia 6310, 6510, 8310, Siemens ME45, SL45, S55, Motorola Accompli A008 oraz Ericsson T68. Ponadto do realizacji połączeń telefonicznych, które następnie rejestrowane były na dyktafonach telefonów komórkowych, użyto aparaty: Nokia 8210, Panasonic EB-GD67 oraz stacjonarny aparat pracujący w sieci PSTN a także telefon ISDN wyposażony w cyfrowy aparat zgłoszeniowy.

Na rycinie 1 zamieszczono artykulacyjny opis sześciu polskich samogłosek, dla których przeprowadzono analizę parametrów mierzalnych mowy (struktura formantowa  $F_1$ – $F_3$ ) [7, 13].

Możliwość dokonania opisu artykulacyjnego mowy dla nagrań zarejestrowanych w pamięci telefonu komórkowego oceniono na podstawie analizy następujących zdań:

/nɔjvɛ · ɕɛtʃɛnʃɛsɔwʃpɔznɔɲu/  
Najwyższe czynsze są w Poznaniu.

/jakɔpʲɛnknɔdʒiɛpɔgɔda/  
Jaka piękna dziś pogoda.

/mugw miuzɛndɲikpɔmutʃ/  
Mógł mi urzędnik pomóc.

/stʃɛmpjastɛliɛtʃɛsawatɛ/  
Strzępiaste liście sałaty.

/tʃɛdɲɔvɛsɔntʃɛspɔzɛwɔpɔmarɔntʃɛ/  
Trzydniowe słońce sparzyło pomarańcze.

/ɔbɛɔvɔpɔjɔntʃɛtʃɛbaɔvɔlubitʃ/  
Aby owo pojąć, trzeba owo lubić.

/ʃɛdwzjɔpwcɛmvrɛntʃɛ/  
Szedł z jabłkiem w ręce.

/tʃɛ · vjɛwʒɛɲumɔznɔmuvidʒ · ɔrɛsɔtʃjalizɔtʃɛsji/  
Czy w więzieniu można mówić o resocjalizacji?

W prezentowanej części projektu zarówno izolowane samogłoski, jak i wypowiedzi ciągle, wypowiedane były przez tę samą osobę, tj. kobietę w wieku 26 lat. Poniżej przedstawiono rozkład częstotliwości podstawowej tonu krtaniowego uczestnika eks-

perymentu uzyskane dla wypowiedzi referencyjnych zarejestrowanych bezpośrednio na cyfrowym rejestratorze MiniDisc HHB:  $F_{o\text{sr}} = 205$  Hz;  $F_{o\text{min}} = 180$  Hz;  $F_{o\text{max}} = 22$  Hz.

Częstotliwości średnie trzech pierwszych formantów uzyskane dla ustabilizowanych fragmentów sześciu polskich samogłosek oraz odpowiednie względne odchylenia standardowe zamieszczono w tabelach I i II.

Każda analizowana wypowiedź była równocześnie rejestrowana na dyktafonie wybranego telefonu komórkowego jako materiał dowodowy oraz bezpośrednio na rejestratorze cyfrowym MiniDisc jako materiał referencyjny. Zgromadzony w ten sposób materiał pozwalał przeanalizować wpływ rozpatrywanych jednocześnie algorytmów kodujących sygnał mowy, zaimplementowanych w rejestratorze-dyktafonie wybranego telefonu komórkowego, jak również parametrów samego aparatu telefonicznego (m.in. charakterystyki pasma oraz mikrofonu) na zrozumiałość i parametry akustyczne mowy.

### Przebieg badań

Przed przystąpieniem do rejestracji nagrań określono pasmo przenoszenia dyktafonów każdego z badanych telefonów komórkowych. W tym celu na dyktafonach telefonów testowych zarejestrowano szum różowy (szum, w którym intensywność sygnałów jest odwrotnie proporcjonalna do częstotliwości; charakterystyka częstotliwościowa takiego szumu jest zbliżona od charakterystyki ludzkiego słuchu), mianowicie przy użyciu aktywnych monitorów marki HHB Circle 5A odtwarzano szum różowy i utrwalano go na dyktafonie badanego aparatu telefonicznego.

Najszerze pasmo, w jakim możliwe było zarejestrowanie częstotliwości z zakresu emitowanego szumu różowego, zaobserwowano dla dyktafonów Siemens (np. dla SL45 wynosiło 200–4000 Hz; są to wyniki dla sygnału pozyskiwanego poprzez złącza cyfrowe, IrDA lub com). Najwęższe zaś pasmo przenoszenia uzyskano dla telefonu Ericsson (300–3600 Hz dla sygnału pozyskanego sprzężeniem akustycznym z transmisją radiową w zakresie FM). Wyniki dotyczące pasma przenoszenia oraz charakterystyki pasma uzyskane dla poszczególnych telefonów z uwzględnieniem sposobu zabezpieczania nagrań z pamięci aparatu telefonicznego zaprezentowano w sposób graficzny na rycinie 2.

Na podstawie zgromadzonych wyników stwierdzono ponadto, iż dyktafony telefonów komórkowych różnych marek wykazują różne charakterystyki pasma przenoszenia. I tak, charakterystyka pasma przenoszenia dla dyktafonów Nokia była wyraźnie różna od charakterystyk pasma uzyskanych dla dyktafonów Siemens, a równocześnie odnotowano znaczne podobieństwo charakterystyki pasma dyktafonu Motorola A008 w porównaniu do spójnej, pod tym względem, grupy dyktafonów telefonów Nokia. Stwierdzono także znaczne zróżnicowanie w obrębie charakterystyki pasma dla różnych modeli dyktafonów telefonów tej samej marki, np. dla telefonów Siemens. Charakterystyka pasma dyktafonów poszczególnych modeli telefonów Siemens była bardziej zróżnicowana, aniżeli w grupie dyktafonów pozostałych marek telefonów komórkowych. Podkreślić jednakże należy, iż sposób pozyskiwania nagrania szumu różowego z dyktafonu istotnie wpływał na charakterystykę, a szczególnie na zakres pasma przenoszenia. Ponadto zaobserwowano znaczną ingerencję metody sprzężenia akustycznego z transmisją radiową FM na strukturę akustyczną głosu (m.in. wpływ zastosowanych w tym torze głośników i mikrofonów), co pozwoliło uznać ten sposób zabezpieczania materiału do badań identyfikacyjnych jako niedosta-



teczny (niska korelacja względem nagrań bezpośrednich utrwalonych na MiniDisc HHB). W związku z powyższym, sposób ten powinien być stosowany jedynie wówczas, gdy nie ma innych możliwości zabezpieczenia nagrań, a parametry uzyskane dla takiego materiału powinny być interpretowane z dużą ostrożnością (taki sposób pozyskiwania sygnału audio jest często wykorzystywany do zabezpieczenia nagrań z tzw. automatycznych sekretarek telefonów czy faksów).

Dodać również należy, iż dyktafony telefonów, które użyto w niniejszym eksperymencie, różniły się także maksymalną długością nagrania, jakie można było utrwalić w ich pamięci. Przykładowo najkrótsze, 30-sekundowe nagranie można było zarejestrować za pomocą dyktafonu Ericsson T68, najdłuższe zaś, około 5-godzinne na dyktafonie Siemens SL45; najczęściej jednak pojemność pamięci testowanych dyktafonów umożliwiały zapis 2–3 minutowego nagrania. Wśród cech różniących testowane aparaty telefonów komórkowych należy także wskazać techniczne możliwości zabezpieczenia nagrań: aparat Ericsson pozwalał transferować nagrania jedynie sprzężeniem akustycznym, telefony Nokia 6510 i Motorola A008 – wyjściem słuchawkowym, podczas gdy aparaty Nokia 8310 i wszystkich modeli marki Siemens umożliwiały zabezpieczenie nagrań cyfrowymi złączami, tj. poprzez port podczerwieni, IrDA lub szeregowy, com (transfer danych złączami cyfrowymi wymagał oczywiście odpowiedniego oprogramowania); wreszcie aparat Nokia 8310 pozwalał na dowolny wybór sposobu transferu nagrań.

Nadmienić także trzeba, iż spośród wszystkich przebadanych dyktafonów telefonów komórkowych tylko dla aparatu Ericsson możliwy był wybór jakości nagrania, tzn. normalnej i wówczas możliwe było zarejestrowanie nagrania trwającego 107 sekund lub też tzw. jakości wysokiej, dla której można było utrwalić tylko 41 sekund (por. rycina 2). Zarówno audytywne, jak i instrumentalne zestawienie nagrań zarejestrowanych z różną jakością nie wskazywało na występowanie między nimi istotnych różnic.

W prezentowanej części eksperymentu szczegółowej analizie poddano tzw. parametry statyczne mowy, czyli parametry opisujące najbardziej ustabilizowaną fazę artykulacji. W związku z tym zamiast opisu segmentów mowy w ujęciu dynamicznym, uwzględniającym m.in. zjawisko koartykulacji głosek, skoncentrowano się na ekstrakcji wartości średnich trzech pierwszych formantów z artykulacyjnej ustabilizowanej fazy sześciu samogłosek. Pomiary trzech pierwszych formantów  $F_1$ – $F_3$  dokonywane były w sposób automatyczny za pomocą oprogramowania Praat 4.0.51. W odniesieniu do nagrań zarejestrowanych za pomocą dyktafonów telefonów komórkowych, uzyskane w sposób automatyczny wartości poszczególnych formantów poddawane były weryfikacji poprzez analizę charakterystyki pasma uzyskanej dla testowanego dyktafonu. Nagrania, które zostały utrwalone za pomocą funkcji dyktafonu w pamięciach telefonów komórkowych, uwzględniały różne typy wypowiedzi, a mianowicie:

- wypowiedzi bezpośrednie;
- tzw. rozmowy przychodzące realizowane za pomocą telefonów komórkowych;
- tzw. rozmowy przychodzące realizowane za pomocą telefonu stacjonarnego PSTN;
- tzw. rozmowy przychodzące, realizowane za pomocą telefonu stacjonarnego PSTN, a następnie zarejestrowane przez cyfrowy aparat zgłoszeniowy telefonu ISDN.



## WYNIKI

**Analiza akustyczna*****Nagrania wypowiedzi bezpośrednich***

Nagrania wypowiedzi bezpośrednich, zarejestrowane za pomocą dyktafonów następujących telefonów komórkowych: Siemens ME45, SL45, S55 oraz Nokia N6310 i N8310 zabezpieczono poprzez złącza cyfrowe (port podczerwieni, IrDA lub port szeregowy, com). Współczynniki korelacji liniowej, jakie uzyskano pomiędzy pozyskanymi w ten sposób nagraniami testowymi a materiałem referencyjnym, rejestrowanym bezpośrednio i symultanicznie z testowymi wypowiedziami na cyfrowym urządzeniu MiniDisc, były wysokie i wynosiły od 0,978 dla aparatu Siemens S55 do 0,930 dla dyktafonu telefonu Nokia 8310.

Wartości trzech pierwszych formantów uzyskane dla wypowiedzi bezpośrednich zarejestrowanych na dyktafonach telefonów komórkowych wraz z odpowiednimi wartościami dla samogłosek wyekstrahowanych z materiału referencyjnego przedstawiono w tabeli III. W tabeli tej uwzględniono także zmienność naturalną mówiącego, wyznaczoną dla każdego z analizowanych formantów poszczególnych samogłosek referencyjnych.

Uwzględniając naturalną zmienność mowy, można wskazać występowanie wspólnych tendencji dla wszystkich bezpośrednich wypowiedzi zarejestrowanych na dyktafonach różnych telefonów komórkowych (Siemens ME45, SL45, S55 oraz Nokia N6310, N8310). Mianowicie, wartość średnia pierwszego formantu  $F_1$  wyznaczona dla nagrań testowych była wyższa aniżeli wartość, jaką uzyskano dla materiału referencyjnego. Ponadto, wartość trzeciego formantu  $F_3$  dla samogłosek /a/ i /o/ ulegała obniżeniu względem wartości wyznaczonej dla samogłosek porównawczych.

Te same nagrania zarejestrowane w pamięciach telefonów komórkowych o podobnej charakterystyce pasma, tj. Nokia 6510 i Motorola A008, zabezpieczono także w inny sposób, a mianowicie wyjściem słuchawkowym. Uzyskane współczynniki korelacji liniowej dla nagrań testowych dokonanych za pomocą dyktafonów dwóch wspomnianych telefonów i zabezpieczonych wyjściem słuchawkowym oraz nagrań referencyjnych były także wysokie i wynosiły odpowiednio 0,993 dla aparatu Nokia 6510 i 0,994 dla aparatu Motorola A008. Dla nagrań testowych zabezpieczonych poprzez wyjście słuchawkowe uzyskano wysokie współczynniki korelacji liniowej, a także zaobserwowano analogiczne przesunięcia formantów w porównaniu z wartościami uzyskanymi dla materiału referencyjnego, tj. wartość pierwszego formantu  $F_1$  wysokich samogłosek /i/ i /ɨ/ ulegała wyraźnemu podwyższeniu. Podczas szczegółowej analizy nagrań zabezpieczonych z dyktafonów aparatów N6510 i Motorola A008 poprzez wyjście słuchawkowe ujawniono jednakże występowanie odmiennych tendencji, zwłaszcza dla wartości formantu drugiego  $F_2$  wyekstrahowanego dla trzech samogłosek /ɛ/, /a/ i /o/ (por. tabela IV). Uwzględniając powyższe, można stwierdzić, iż nagrania zabezpieczone wyjściem słuchawkowym z dyktafonów o podobnej charakterystyce pasma nie wykazują analogicznych zniekształceń w obrębie parametrów mierzalnych mowy tak, jak miało to miejsce w przypadku nagrań wypowiedzi bezpośrednich transferowanych złączami cyfrowymi z pamięci dyktafonów nawet o różniących się charakterystykach pasma.

Na poniższych ilustracjach (ryciny 3, 4) przedstawiono zniekształcenia parametrów mierzalnych mowy, wyekstrahowanych z nagrań wypowiedzi bezpośrednich zarejestrowanych na dyktafonach telefonów komórkowych w zależności od sposobu zabezpieczania sygnału mowy. W celu zweryfikowania wyników uzyskanych dla nagrań bezpośrednich zarejestrowanych na dyktafonach różnych telefonów komórkowych i zabezpieczonych różnymi torami, tj. IrDA/com oraz wyjściem słuchawkowym, przeanalizowano wyniki uzyskane dla nagrań bezpośrednich zarejestrowanych na jednym dyktafonie – Nokia N8310 i równolegle zabezpieczanych obydwoma torami, cyfrowym i analogowym. Wyniki przedstawiono na ilustracji poniżej (rycina 2).

Struktura formantowa ( $F_1-F_3$ ) wyznaczona dla nagrań bezpośrednich wypowiedzi zarejestrowanych na dyktafonie telefonu Nokia 8310 względem struktury ustalonej dla nagrań referencyjnych wskazuje lepszą korelację dla nagrań zabezpieczanych złączami cyfrowymi IrDA/com, aniżeli wyjściem analogowym. Ponadto zauważyć należy, iż dla formantu pierwszego,  $F_1$  i trzeciego,  $F_3$  większości z analizowanych samogłosek nieadekwatność wartości uzyskanych dla nagrań zarejestrowanych na dyktafonie i zabezpieczonych przez port IrDA/com względem nagrania referencyjnego mieści się w zakresie naturalnej zmienności mówiącego; w przypadku  $F_3$  prawidłowość ta odnosi się do wszystkich samogłosek.

### ***Tzw. rozmowy przychodzące realizowane za pomocą telefonów komórkowych***

Aby przeanalizować wartość identyfikacyjną parametrów mierzalnych mowy wyekstrahowanych z nagrań tzw. rozmów przychodzących, zarejestrowanych w pamięci telefonów komórkowych, wykonano połączenia z aparatu telefonu komórkowego Nokia 8210 na następujące aparaty: Siemens ME45, Siemens S55 i Nokia 6310 oraz z aparatu komórkowego Panasonic EB-GD 67 na telefon Siemens S55. Zgodnie z procedurami niniejszego eksperymentu tzw. rozmowy przychodzące były jednocześnie rejestrowane na magnetofonie cyfrowym MiniDisc jako materiał referencyjny. Każda z tzw. rozmów przychodzących, która została utrwalona za pomocą dyktafonów telefonów komórkowych aparatów Siemens ME45, S55 i Nokia 6310, zabezpieczono poprzez złącza cyfrowe (IrDA/com).

Parametry mierzalne uzyskane dla materiału referencyjnego oraz dla nagrań tzw. rozmów przychodzących realizowanych z telefonów komórkowych, a następnie rejestrowanych na dyktafonach trzech aparatów telefonicznych (Siemens ME45, S55 i Nokia 6310), wykazywały dobrą korelację. Ponadto należy podkreślić, że parametry mierzalne mowy określone dla tzw. rozmów przychodzących realizowanych przy pomocy różnych telefonów komórkowych, tj. Nokia 8210 oraz Panasonic EB-GD 67, a następnie utrwalonych na tym samym dyktafonie telefonu Siemens S55, nie różniły się w sposób istotny. Najistotniejszym czynnikiem, który decydował o stopniu zniekształcenia struktury formantowej względem materiału referencyjnego, był dyktafon, na którym dokonywano rejestracji. Pomimo odnotowanych różnic ilościowych nagrania tzw. rozmów przychodzących realizowanych z telefonów komórkowych i zarejestrowanych na dyktafonach telefonów komórkowych, tj. Siemens ME45, S55 i Nokia 6310, a następnie zabezpieczonych złączami cyfrowymi wskazywały analogiczne zniekształcenia mowy. Podobieństwa te dotyczyły: wartości pierwszego formantu  $F_1$  – dla większości analizowanych samogłosek była ona wyższa aniżeli wartość referencyjna, ponadto dla samogłoski /a/ odnotowano tendencję do obniżania wartości formantu drugiego, podobnie obniżenie wartości zaobserwowano

dla samogłoski wysokiej i przedniej /i/ oraz tylnej, półotwartej /ɔ/, przy czym obniżenie dla tych dwóch samogłosek dotyczyło wartości formantu trzeciego,  $F_3$ . Omówione wyniki w sposób graficzny zaprezentowano na rycinach 6 i 7.

### ***Tzw. rozmowy przychodzące realizowane za pomocą telefonu stacjonarnego PSTN***

Wszystkie tzw. rozmowy przychodzące realizowane były z tego samego aparatu telefonu PSTN, a następnie rejestrowano je na dyktafonach czterech telefonów komórkowych (Siemens S55, Motorola A008 oraz Nokia 6310 i 6510). Dla tej konfiguracji połączeń przetestowano dwa sposoby zabezpieczenia nagrań z pamięci telefonów komórkowych, tj. transfer złączami cyfrowymi oraz wyjściem słuchawkowym.

W przypadku zabezpieczania nagrań poprzez złącza cyfrowe uzyskano niskie współczynniki korelacji, a ponadto zniekształcenia mowy tak dalece zależały od dyktafonu użytego do rejestracji, iż nie można było nawet wskazać analogicznych tendencji w odniesieniu do analizowanej struktury formantowej.

Drugim z testowanych sposobów zabezpieczania nagrań tzw. rozmów przychodzących realizowanych z telefonu PSTN, a następnie utrwalonych na dyktafonach czterech aparatów telefonów komórkowych, było wyjście słuchawkowe. Pomiar trzech pierwszych formantów przeprowadzone dla tak zabezpieczonych nagrań tzw. rozmów przychodzących wykazywały niewielkie odchylenie względem pomiarów uzyskanych dla materiału referencyjnego; dodać należy, iż odchylenie to mieściło się w zakresie naturalnej zmienności mowy uczestnika eksperymentu. Ponadto zestawienie nagrań tzw. rozmów przychodzących realizowanych z telefonu PSTN, a następnie zarejestrowanych na dyktafonach o podobnej charakterystyce pasma, a więc na dyktafonach telefonów Nokia 6510 i Motorola A008, wykazało podobne tendencje w obrębie analizowanych formantów. Mianowicie, w nagraniach tych zaobserwowano wyższe wartości pierwszego formantu  $F_1$  dla samogłoski /ɔ/ w porównaniu z ustalonymi dla odpowiednich nagrań referencyjnych, podczas gdy dla samogłosek /a/ oraz /i/ ujawniono niższe wartości – odpowiednio – drugiego  $F_2$  oraz trzeciego  $F_3$  formantu (por. tabela V).

Porównanie wartości trzech pierwszych formantów wyekstrahowanych z nagrań zarejestrowanych na dyktafonach o różniących się charakterystykach pasma wykazywało gorszą wzajemną korelację.

Podsumowując wyniki uzyskane dla nagrań tzw. rozmów przychodzących realizowanych z telefonu PSTN, a następnie rejestrowanych na dyktafonach różnych telefonów komórkowych, należy stwierdzić, iż takie nagrania powinny być zabezpieczane przez wyjście słuchawkowe. Transfer nagrań wyjściem analogowym pozwolił uzyskać lepszą korelację materiału testowego względem referencyjnego, a ponadto różnice wynikające z dyktafonu użytego do rejestracji były mniejsze niż w przypadku transferu danych złączami cyfrowymi.

### ***Tzw. rozmowy przychodzące, realizowane za pomocą telefonu stacjonarnego PSTN, a następnie zarejestrowane przez cyfrowy aparat zgłoszeniowy telefonu ISDN***

Podczas przedstawianego eksperymentu wykonano również połączenia z telefonem PSTN na aparat ISDN wyposażony w cyfrowy rejestrator zgłoszeń, za pomocą

którego dokonano utrwalenia tej rozmowy; równocześnie rozmowa została utrwalona jako materiał referencyjny na rejestratorze MiniDisc. Następnie, zgodnie z praktyką zabezpieczania materiałów ze stacjonarnych aparatów telefonicznych, nagranie przechowywane w pamięci telefonu ISDN zostało odtworzone i zabezpieczone poprzez wyjście słuchawkowe aparatu. Ponadto wiadomość utrwalona w pamięci tzw. automatycznej sekretarki telefonu ISDN została odsłuchana poprzez zdalny dostęp do skrzynki aparatu ISDN i jednocześnie zarejestrowana na dyktafonie telefonu komórkowego Siemens ME45; zabezpieczenia danych z dyktafonu Siemens dokonano poprzez złącza cyfrowe.

Współczynniki korelacji liniowej wyznaczone dla obu sposobów zabezpieczenia nagrania z tzw. automatycznej sekretarki ISDN były wysokie (dla toru audio współczynnik wynosił 0,993, dla zdalnego odtworzenia i rejestracji na dyktafonie Siemens ME45 – 0,992). Zaobserwowano ponadto, iż różnice w wartościach formantów wyznaczonych dla wiadomości z sekretarki dla toru audio i dla dyktafonu ME45 względem wartości uzyskanych dla materiału referencyjnego mieszczą się w zakresie naturalnej zmienności mówiącego. W pozostałych zaś nielicznych przypadkach, gdy odstępstwa te są istotne statystycznie, wykazują analogiczne tendencje, mianowicie: dla  $F_1$  w sposób istotny podwyższeniu ulega wartość formantu dla samogłoski /i/, wartość formantu drugiego,  $F_2$ , ulega istotnemu obniżeniu dla niskiej samogłoski /a/, wreszcie wartość formantu  $F_3$  ulega istotnemu podwyższeniu dla półotwartej przedniej samogłoski /ɛ/. Jedyną istotną niezgodnością w tendencjach zaobserwowanych dla obu sposobów zabezpieczania nagrania z aparatu zgłoszeniowego telefonu ISDN dotyczyła wartości trzeciego formantu samogłoski /a/, który w nagraniu zabezpieczonym wyjściem słuchawkowym aparatu ISDN w sposób istotny ulegał podwyższeniu, zaś nagrania zabezpieczonego przy pomocy zdalnego odtworzenia i dyktafonu telefonu komórkowego jego wartość ulegała obniżeniu (por. tabela VI).

### Analiza fonetyczna

Zgromadzone podczas eksperymentu nagrania poddano także ocenie audytywnej. W oparciu o kryteria zrozumiałości i wyrazistości mowy, nagrania zarejestrowane za pomocą dyktafonu telefonu komórkowego Siemens S55 zostały uznane jako najgorsze. Dla przykładu w nagraniach rozmów telefonicznych realizowanych z telefonu stacjonarnego PSTN, a następnie rejestrowanych na dyktafonie Siemens S55, samogłoski /u/ oraz niektóre z izolowanych realizacji samogłosek /i/ i /o/ były interpretowane jako tło akustyczne lub zwykły szum (por. rycina 8).

Ponadto w niektórych realizacjach zacierała się fonetyczna opozycja pomiędzy samogłoską /a/ i /o/. Szczególnej degradacji, po zanik zrozumiałości, ulegały spółgłoski zarówno w nagłosie, jak i w pozycji interwokalicznej; dotyczyło to zwłaszcza twardych bezdźwięcznych spółgłosek (np. /f/, /p/, /t/ i /s/) oraz nosowych (por. ryciny 9, 10).

Zniekształcenia mowy utrwalonej w pamięci telefonów komórkowych pogłębiały się w tych nagraniach, w których rozmówca poruszał się w trakcie rejestrowanej rozmowy. Zaznaczyć również należy, iż największej degradacji ulegały głoski występujące w nagłosie, zwłaszcza po pauzie fonetycznej; proces ten dotyczył w podobnym stopniu różnych wypowiedzi, tj. zarówno bezpośrednich, jak i tzw. rozmów przy-

chodzących; odnotowywany był także dla większości aparatów telefonicznych (w większym stopniu dla telefonu Nokia 6310i, w mniejszym dotyczył aparatu Siemens ME45). Podobnie parametry czasowe uzyskane dla segmentów mowy wyekstrahowanych z nagrań zarejestrowanych na dyktafonach telefonów komórkowych powinny być poddane ostrożnej interpretacji. Zaobserwowano bowiem, zwłaszcza dla realizacji izolowanych samogłosek, iż samogłoski z nagrań utrwalonych w pamięci telefonów komórkowych były krótsze lub w obrębie trwania głoski występowały nie-naturalne przerwy (por. rycina 8).

Eksperyment prezentowany w niniejszym artykule był przeprowadzony przy znacznie ograniczonych warunkach, tzn. nagrań dokonywano w wytłumionym pomieszczeniu studyjnym, a do rejestracji użyto wysokiej jakości urządzenia. Autorzy zdecydowali się także nie badać szczegółowo takich warunków, jak przemieszczanie się rozmówców w trakcie rejestracji lub wpływ zakłóceń pochodzących z otoczenia, jakkolwiek należy oczekiwać, iż uwzględnienie wspomnianych czynników wpłynęłoby na pogorszenie zaprezentowanych dotychczas wyników.

## WNIOSKI

### **Proponowane procedury zabezpieczania nagrań zarejestrowanych w pamięciach telefonów komórkowych**

Na podstawie rezultatów uzyskanych podczas niniejszego eksperymentu należy stwierdzić, iż najlepszą korelację względem materiału porównawczego zarejestrowanego w sposób bezpośredni pozwalają uzyskać nagrania zarejestrowane w pamięci telefonu komórkowego zabezpieczane w następujący sposób:

- poprzez złącza cyfrowe IrDA/com – wypowiedzi bezpośrednie, jak i rozmowy przychodzące z telefonów komórkowych zarejestrowane na dyktafonach telefonów komórkowych;
- poprzez analogowe wyjście słuchawkowe – rozmowy przychodzące z PSTN (jako szczególny przypadek również wiadomość z PSTN zarejestrowana na automatycznej sekretarce ISDN).

### **Badania identyfikacyjne**

Podczas badań identyfikacyjnych mówiącego, zgodnie ze stosowaną przez autorów metodą językowo-pomiarową, duża uwaga poświęcana jest fonetycznemu opisu nawyków mownych osoby identyfikowanej. Możliwość dokonania takiej analizy decyduje często o możliwości przeprowadzenia wiarygodnych badań i sformułowaniu wniosków końcowych z dużym stopniem prawdopodobieństwa.

Tak więc, co zostało omówione szerzej w rozdziale „Eksperyment”, nagrania dokonane za pomocą dyktafonów telefonów komórkowych pomimo kontrolowanych warunków, w jakich przeprowadzany był niniejszy eksperyment, cechowały się ograniczoną wyrazistością mowy, a często i brakiem zrozumiałości. Zniekształcenia mowy w nagraniach zarejestrowanych w pamięciach telefonów komórkowych dotyczyły zarówno nagrań wypowiedzi bezpośrednich, jak i tzw. rozmów przychodzących i obejmowały tak spółgłoski, jak i samogłoski. Należy podkreślić, iż w przypadku realnych nagrań dowodowych zarejestrowanych za pomocą funkcji dyktafonu telefonu komórkowego zniekształcenia te będą prawdopodobnie większe, choćby ze wzglę-

du na zakłócenia wynikające z tła akustycznego towarzyszącego rejestrowanej rozmowie czy też z poruszania się rozmówców. Możliwość dokonania wiarygodnego opisu nawyków artykulacyjnych mówiącego jest zatem ograniczona. W tej sytuacji źródłem istotnych, acz niewystarczających informacji o mówcy, stają się parametry mierzalne mowy, w tym zwłaszcza analiza struktury formantowej samogłosek [2, 12]. Do właściwej interpretacji parametrów mierzalnych uzyskanych dla nagrań zarejestrowanych w pamięci telefonu komórkowego należy uwzględnić – o ile jest to możliwe – charakterystykę pasma danego dyktafonu.

Należy zatem podkreślić, iż w opinii autorów nie ma dostatecznych przesłanek, które skłaniałyby do rejestrowania materiału porównawczego za pomocą dyktafonu takiego samego telefonu komórkowego jak ten, na którym zostały utrwalone dowodowe wypowiedzi, szczególnie że informacje o urządzeniu dowodowym zwykle są trudne do ustalenia. Należy także zaznaczyć, iż nawet bardzo podobne, ze względu na charakterystykę pasma, rejestratory telefoniczne nie pozwalają przyjąć hipotetycznych założeń o rodzaju i stopniu zniekształceń w obrębie struktury formatowej. Wreszcie, w przypadku rozmów przychodzących, wielość możliwych zmiennych dotyczących m.in. aparatu, z którego dzwonił (ISDN czy PSTN, model a zatem i charakterystyka mikrofonu) wydaje się uniemożliwiać rekonstrukcję warunków właściwych dla dowodowego połączenia. Informacje o poszczególnych ogniwach połączenia pochodzą zresztą z zeznań świadków lub części stron, nie wynikają natomiast z bezpośredniej analizy nagrania dowodowego.

### **Uwagi dodatkowe**

Podczas prezentowanego eksperymentu zaobserwowano, iż niektóre z badanych telefonów (wszystkie modele telefonów Nokia: 6510, 6310i, 8310 oraz Motorola A008) podczas rejestrowania rozmowy przychodzącej periodycznie (zwykle co 5 lub 10 s) rejestrowały określony ton. Dla Nokii był to ton o częstotliwości 1175 Hz, zaś dla Motoroli o wartości 2500 Hz. Obecność takiego sygnału w nagraniu może być pomocna w badaniach ciągłości nagrania, jak również sygnał ten może być potraktowany jako marker właściwy dla marki telefonu, na którym dokonano rejestracji rozmowy. Zaznaczyć przy tym należy, iż sygnał ten jest słyszalny dla rejestrowanego rozmówcy, co z kolei demaskuje niejawną formę nagrania rozmowy telefonicznej. Sygnał taki nie występował w nagraniach bezpośrednich zarejestrowanych za pomocą dyktafonu telefonu komórkowego.

Ponieważ prezentowane wyniki są tylko wynikami wstępnymi prowadzonego projektu, wszelkie uogólnienia powinny być więc ostrożne. Autorzy zaznaczają jednocześnie, iż niezbędne jest przebadanie analogicznych zestawień, przedstawionych w rozdziale „Wyniki” dla innych dyktafonów telefonów komórkowych, a przede wszystkim konieczne jest poszerzenie grupy osób, których wypowiedzi są rejestrowane (oczywiście należy zweryfikować zaobserwowane tendencje dla głosu kobiecego względem wyników dla głosu męskiego). Ponadto słuszne wydaje się powtórne przeanalizowanie zebranych już danych z perspektywy opisu dynamicznego mowy, tzn. uwzględniającego zjawisko koartykulacji.