Logatome intelligibility analysis of the FM demodulated signal

Svjetlana Kovačević, Jovan Galić and Tatjana Pešić-Brđanin

Abstract— **In this paper the analysis of logatome intelligibility of the FM demodulated signal is presented. The subjective tests include 50 listeners and 125 logatomes of CVC (consonant-vocal-consonant) type. The experiments are conducted for two values of deviation frequency: 20 kHz and 5 kHz. The confusion matrices for vowels and consonants in the initial and the final position are given as well. The average recognition rate for logatomes is 93.8% (for deviation 20 kHz) and 93.2% (for deviation 5 kHz).**

Index Terms— **FM modulation/demodulation; Logatomes; Statistical analysis; Subjective tests.**

I. Introduction

Despite the fact that digitalization of audio broadcast (DAB) signal has brought many benefits compared to traditional analog, the transition (so called "switchover") is expected to be progressive. Frequency Modulation (FM) is dominant technology for analog audio signal broadcast. FM broadcast band spans from 97.5 to 108 MHz (in Europe). Normally, each channel is 200 kHz wide, with deviation limited to 150 kHz (±75 kHz).

Speech signal quality can be evaluated in objective and subjective manner. The most used objective method for objective evaluation is Perceptive Evaluation of Speech Quality (PESQ) [1-2]. On the other hand, Mean Opinion Score (MOS) is the most used method for subjective evaluation [3]. Subjective tests are time-consuming but provide exact methodology. Disadvantages of subjective evaluations are difficult experiment preparation and need for great number of listeners [1].

For speech signal intelligibility, words without meaning (pseudowords or logatomes) can be used as well. Logatomes are used in experiments which examine speech recognition, decoding and perception of normally phonated and whispered speech [4-6]. They can be in different forms: CVC, VC, VCV, CVCV, where C denotes consonant and V vowel.

Because of the fact that logatomes of three-level structure CVC are easily pronounceable, that form of logatomes is used in this study, in which logatome recognition rate for the FM demodulated signal is analyzed.

Svjetlana Kovačević is with Faculty of Electrical Engineering, University of Banja Luka, Republic of Srpska, Bosnia and Herzegovina (e-mail: svjetlana.kovacevic@etf.unibl.org).

Jovan Galić is with Faculty of Electrical Engineering, University of Banja Luka, Republic of Srpska, Bosnia and Herzegovina (e-mail: jovan.galic@etf.unibl.org).

Tatjana Pešić-Brđanin is with Faculty of Electrical Engineering, University of Banja Luka, Republic of Srpska, Bosnia and Herzegovina (e-mail: tatjana.pesic-brdjanin@etf.unibl.org).

The remainder of this paper is organized in four Sections as follows. In Section II the basic characteristics of FM demodulation along with the description of exploited demodulator are presented. Section III describes the experimental preparation while Section IV gives the results and their discussion. Finally, in Section V concluding remarks and directions for future works are given.

II. FM Demodulator

All circuits that perform the function of differentiation in time domain can be used as converters of a frequency modulation signal into an amplitude modulation (AM) signal. It's possible to extract a modulating signal from an AM signal using an envelope detector. For the analysis presented in this paper, the slope detector is designed and realized (shown in Fig. 1.a), with the frequency response of the input selective circuit given in Fig. 1.b [7].

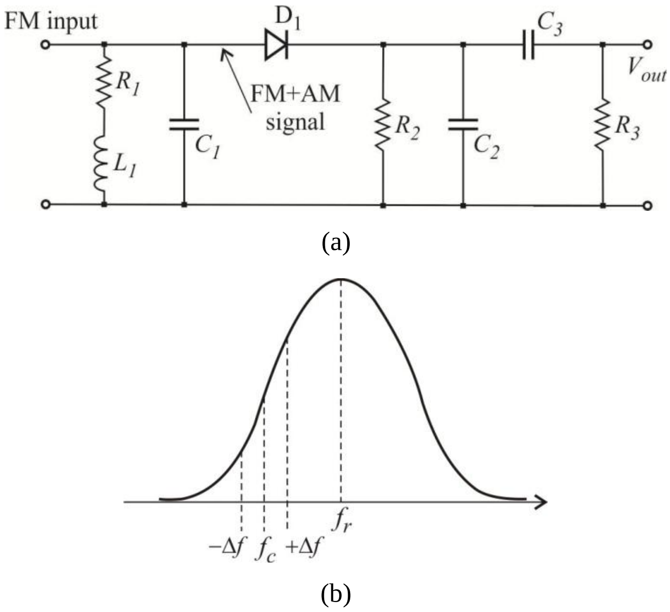

Fig. 1. Slope detector (a) and frequency response of the input selective circuit ($R_1L_1C_1$) (b) [7]

The measured value of resonant frequency of the input selective circuit is 1.454 MHz for used coil with inductivity $L_1 = 1\mu\mathrm{H}$ and quality factor $Q = 24$ and the capacitor with $C_1 = 12\mathrm{nF}$. Also, the part of the characteristics which is considered to be the most linear is determined, and includes the frequency range from 1.4105 MHz to 1.4501 MHz. The carrier frequency f_c is set in the middle of the interval and amounts 1.4303 MHz (the frequency deviation is at the most 20 kHz).

Proceedings of 4th International Conference on Electrical, Electronics and Computing Engineering, IcETRAN 2017, Kladovo, Serbia, June 05-08, ISBN 978-86-7466-693-7

Assuming that the modulating signal is a telephone voice signal which occupies the frequency range from $f_{\min} = 300\text{Hz}$ to $f_{\max} = 3400\text{Hz}$ and that the index of modulation is $m = 0.6$, the following elements are selected for the envelope detector: $R_2 = 1\text{k}\Omega$ and $C_2 = 47\text{nF}$.

For the elimination of DC components at the output of the envelope detector, a high-pass RC filter is used. The elements of RC filter are calculated taking into account $R_3C_3 \geq (3 \div 5) / \omega_{\min}$, which means that $R_3C_3 > 0.001592$. This requirement is chosen because of the fact that all significant spectral components of the speech signal should not be distorted in the high-pass filter. Since the demodulated signal was fed to a headphone (8 Ω) or the high-resistance loudspeaker (2 kΩ), it was necessary to select a large capacitor. In this realized FM demodulator, the capacitor with $C_3 = 470\mu\text{F}$ is used [7].

III. EXPERIMENTAL SETUP

In order to determine the logatome intelligibility, as subjective evaluation of the quality of demodulated signals, the listening tests are performed. First, the speech database consisting of utterances of logatomes is created. Logatomes are arranged in two tables, each with 125 CVC logatomes. They are balanced logatomes, formed in a way that each of the 25 consonants occurs 5 times in the initial and 5 times in the final position, and each of the 5 vowels occurs 25 times in the medial position [6].

We used a software package PRAAT [8] to record the speech database, one recording session for each table. Logatomes are uttered by male person, with pause between two spoken logatomes 4-5 seconds. Recording is performed in a quiet room, with measured ambient noise level of 29 dB(A). Recordings are in the mono-channel technique, with the sampling frequency of 44.1 kHz and 16 bits per sample in wav format.

In addition, listening tests are performed under the same conditions, i.e. in a quiet room. The 50 students aged between 20 and 25 years from the Faculty of Electrical Engineering in Banja Luka participated in experiments. They were volunteers, without reported voice and hearing disorders. Their task was to write all perceived logatomes in the prepared forms, simultaneously while listening.

The block diagram used to perform experiment is given in Fig. 2. FM signal demodulator, laptop, the FM modulator (FM stereo / FM-AM signal generator 1100 [9]) and a headset are used in preparation. The laptop is used for generating a voice message signal, which is pre-recorded. FM modulator is a device that performs modulation using an output signal from the sound card of the laptop. The output of FM modulator generates an FM signal, and is directly connected to the realized demodulator. FM modulator is tuned to carrier frequency 1.37 MHz, while the level of the FM signal is set to 13 dBm (20 mW). The experiments are carried out for two values of the deviation frequency: 20 kHz and 5 kHz with a pause of several days for each individual listener. One table is used for the frequency deviation of 20 kHz, and the other for the deviation of 5 kHz.

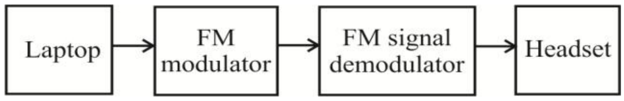

Fig. 2. Block diagram used to perform the experiment

IV. RESULTS AND DISCUSSION

In Serbian language, 25 consonants classified in 7 phonetic classes according to manner of articulation are presented in Table I [10].

TABLE I
GROUPS OF CONSONANTS ACCORDING TO MANNER OF ARTICULATION

Group	**Consonant(s)**
Plosives	/p, t, k/ and /b, d, g/
Fricatives	/f, s, š, h/ and /z, ž/
Affricates	/c, č, ć/ and /dž, đ/
Nasals	/m, n, nj/
Vibrant	/r/
Laterals	/l, lj/
Approximants	/v, j/

For each deviation frequency 250 responses for all listeners and each individual consonant in the final and initial position are obtained in total (5 times per table x 50 listeners). Also, there are 1250 responses for each vowel (25 times per table x 50 listeners).

The obtained results are classified in groups of vowels and consonants (according to manner of articulation) and shown in three tables, which represent the confusion matrices.

Table II shows the confusion matrix for vowels for two deviation frequencies. Numbers of perceived phonemes correctly recognized are highlighted. Other values in the table represent the number of incorrectly recognized phonemes. The empty cells in the table are with zero misrecognition.

TABLE II
CONFUSION MATRIX OF VOCALS

Vocals			**Stimulant**				
			a	e	i	o	u
Response	a	20 kHz	1248				
		5 kHz	1249			1	1
	e	20 kHz		1242	2		
		5 kHz		1247			
	i	20 kHz			1248		3
		5 kHz	1	1	1250		
	o	20 kHz	1			1214	1
		5 kHz		1		1247	
	u	20 kHz				35	1246
		5 kHz				2	1248
	*	20 kHz	1	8		1	
		5 kHz		1			1
ARR [%]		20 kHz	99.84	99.36	99.84	97.12	99.68
		5 kHz	99.92	99.76	100	99.76	99.84

* - recognized consonant instead vowel; ARR - average recognition rate

One can see that the recognition accuracy for all vocals is very good; vocals are correctly perceived with success higher than 97%. It may also be noticed that better results are obtained for the frequency deviation of 5 kHz. The most prominent substitution for vowels is substitution of vowel /o/ as /u/ for the frequency deviation of 20 kHz, and amounts to 2.8%.

TABLE III
CONFUSION MATRIX FOR CONSONANTS IN THE INITIAL POSITION

Initial position consonants		**Stimulant**													
		Plosives		Fricatives		Affricates		Nasals		Vibrant		Laterals		Approximants	
		20 kHz	5 kHz	20 kHz	5 kHz	20 kHz	5 kHz	20 kHz	5 kHz	20 kHz	5 kHz	20 kHz	5 kHz	20 kHz	5 kHz
Response	Plosives	1464	1406	1	14	4	4				2		2	4	5
	Fricatives	4	26	1493	1437	42	41						1	39	1
	Affricates	1		3	6	1186	1194								
	Nasals		1					746	735			1	5		
	Vibrant	1	4		32					250	245				
	Laterals	1	1								1	492	477	10	17
	Approximants		7		2	2	1				1	4	15	444	477
	*	16	2	1	1	1					1			3	
	**	13	53	2	8	15	10	4	15			3			
ARR [%]		97.60	93.73	99.53	95.80	94.88	95.52	99.47	98.00	100.00	98.00	98.40	95.40	88.80	95.40

* - recognized vowel instead of consonant ; ** - same phonetic group but incorrect phoneme; ARR - average recognition rate

TABLE IV
CONFUSION MATRIX FOR CONSONANTS IN THE FINAL POSITION

Final position consonants		**Stimulant**													
		Plosives		Fricatives		Affricates		Nasals		Vibrant		Laterals		Approximants	
		20 kHz	5 kHz	20 kHz	5 kHz	20 kHz	5 kHz	20 kHz	5 kHz	20 kHz	5 kHz	20 kHz	5 kHz	20 kHz	5 kHz
Response	Plosives	1478	1477		4		1	2					1	39	3
	Fricatives		2	1493	1464		1	1						7	3
	Affricates			1		1243	1244	3							
	Nasals	6						710	699						
	Vibrant									250	250			1	
	Laterals							2				496	499	62	15
	Approximants		1	2	2			1				2		390	478
	*	3	1		1	1		4	1						
	**	13	19	4	29	6	4	27	50			2		1	1
ARR [%]		98.53	98.47	99.53	97.60	99.44	99.52	94.67	93.20	100.00	100.00	99.20	99.80	78.00	95.60

* - recognized vowel instead of consonant; ** - same phonetic group but incorrect phoneme; ARR - average recognition rate

Table III shows the confusion matrix of consonants in the initial position. For the deviation of 20 kHz, the lowest recognition rate is obtained for approximants 88.80%. There is pronounced replacement of approximants with fricatives (7.80%), and affricates with fricatives (3.36%).

For the deviation of 5 kHz, the lowest recognition rate is obtained for plosives and amounts to 93.73% (with the most pronounced replacement with phonemes of the same group, 3.53%). The recognition of laterals and approximants for the frequency deviation of 5 kHz is 95.40%, and their mutual substitution is often: phonemes from the group of approximants are with 3.40% recognized as laterals and vice versa, laterals are with 3.00% recognized as approximants. It may also be noticed that plosives, fricatives, nasals, laterals and vibrant have more successful recognition for the frequency deviation of 20 kHz.

Table IV shows a confusion matrix of consonants in the final position. The worst recognition for the frequency deviation of 20 kHz is obtained for approximants and amounts to 78.00%, wherein the replacement is expressed with laterals 12.40%. Nasals are correctly recognized with rate 94.67%, with the highest confusion within the same group (3.60%). Other groups of phonemes are correctly recognized over 98.00%.

For the deviation of 5 kHz the lowest recognition is obtained for the group of nasals, 93.20%, wherein the confusion is predominantly within the same group (6.67%). Then follows approximants with intelligibility of 95.60% and fricatives 97.60%. Approximants are in 3.00% of all cases identified as laterals and fricatives 1.93% as other phoneme in the same phonetic group. It may also be noted that plosives, fricatives and nasals are more successfully recognized for the deviation of 20 kHz. For both deviations, vibrant in the final position is successfully recognized by all listeners (100%).

If we compare the recognition of the corresponding group in the initial and the final position for the respective frequency deviation, one can see that the identification of consonants in the final positions is more successful for following groups: plosives, fricatives, affricates, vibrant, and laterals.

For the comparison of overall recognition rate between two values of the frequency deviations, statistical tests are needed.

Despite the fact that average logatome recognition rate (LRR) is higher for the deviation of 20 kHz (93.8% versus 93.2%), two tailed Wilcoxon statistical test showed that the increase in average LRR is not statistically significant (p=0.17).

In Fig. 3 the histogram for LRR for the frequency deviations of 20 kHz (a) and 5 kHz (b) is given. As can be seen from Fig. 3, for the deviation of 5 kHz, there are more listeners with very poor recognition rate and less listeners with excelent recognition. In addition, the median for the deviation of 5 kHz is lower for 0.4% and quartiles (both left and right) for 0.8%. These distribution characteristics could be important indicators that LRR for the deviation frequency of 5 kHz is moderately lower.

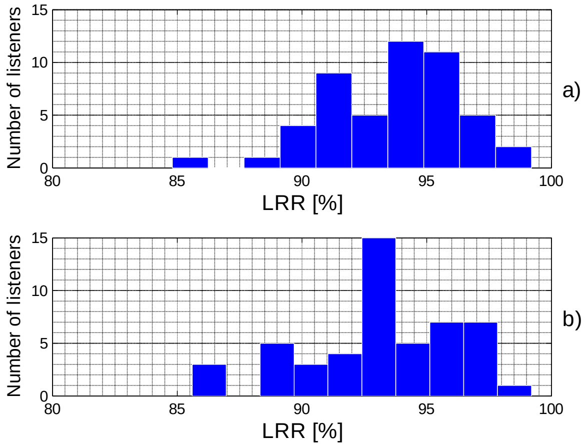

Fig. 3. The histogram of logatome recognition rate for deviation frequency of 20 kHz (a) and 5 kHz (b)

V. Conclusion

Subjective tests play important role in measuring of speech signal quality. In this research study, the recognition of logatomes for the FM demodulated signal has been analyzed, for 50 listeners and deviations of frequency 20 kHz and 5 kHz. The obtained results suggest that there is a moderate decrease (0.6%) in recognition rate for the deviation of 5 kHz. Prospective future works would include a comparison of recognition success score with demodulation of signal primarily intended for modern digital audio broadcasting.

Acknowledgment

The authors would like to thank the students and staff at the Faculty of Electrical Engineering, University of Banja Luka who participated in the experiment.

References

[1] Z. Becvar, J. Zelenka, M. Brada, T. Valenta, “Comparison of Subjective and Objective Speech Quality Testing Methods in the VoIP Networks“, International Conference on Systems, Signals and Image Processing (IWSSIP 2006), Budapest, pp. 427-430, 2006.

[2] ITU-T Recommendation P.862, “Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs“, 2001.

[3] ITU-T Recommendation P.800.1, “Mean Opinion Score (MOS) Terminology“, 2003.

[4] O. Scharenborg, “Reaching over the gap: A review of efforts to link human and automatic speech recognition research“, Speech Communication. 49 (5), pp. 336-347, 2007.

[5] K. Madsen: “The use of pseudowords for decoding instruction“, Ph.D. Thesis, University of Florida, 2014.

[6] S.T. Jovičić, M.M. Đorđević, “Percepcija fonema u šapatu: identifikacija i konfuzija“, chapter in the book S.T. Jovičić, M.Sovilj (ed.): “Govor i jezik: interdisciplinarna istraživanja, II“; Life Activities Advancement Center, 2008.

[7] Svjetlana Kovačević, Tatjana Pešić-Brđanin, Jovan Galić: “Logatomska razumljivost FM demodulisanog signala“, Symposium Infoteh, Jahorina, BiH, pp. 344-348, 2016.

[8] P. Boersma, D. Weenink, *Praat: Doing phonetics by computer*, http://www.fon.hum.uva.nl/praat/,accessed in december 2016.

[9] http://www.peaktech.de/productdetail/kategorie/signalgeneratoren/produkt/p-1100.html, accessed in december 2016.

[10] S. T. Jovičić, “Govorna komunikacija: fiziologija, psihoakustika i percepcija“, pp. 33, Nauka, Beograd, 1999.