International Journal of Electrical and Computer Engineering (IJECE)
Vol. 7, No. 5, October 2017, pp. 2514~2519
ISSN: 2088-8708, DOI: 10.11591/ijece.v7i5.pp2514-2519

The Effect of Bandwidth on Speech Intelligibility in Albanian Language by Using Multimedia Applications like Skype and Viber

Sabrije Osmanaj[1], Altin Shala[2], Blerta Prevalla[3]
[1.2]University of Pristina, Department of Electronics, Pristina, Kosovo
[3]Faculty of Computer Science, AAB University, Pristina, Kosovo

Article Info

Article history:

Received Jul 15, 2016
Revised May 17, 2017
Accepted Jul 11, 2017

Keywords:

Intelligibility
Measurment noise
Multimedia applications
Word error rate

ABSTRACT

This paper intends to analyze subjective measurements of intelligibility of speech on Albanian language during the conversation between two people using applications which today are very used for communication such as Skype and Viber. The measurement is done as follows: on the entry part of the transmission system sentences or words are spoken or just syllables while on receiving part is recorded what is heard; the percentage of words, sentences or syllables correctly received, on proportion to those imposed on the entry of the system, providing the percentage of intelligibility (the words, sentences or syllables). Methods of measurements are made at different speed of the Internet, in an environment without noise and with noise, in order to see the impact on understanding of the speech with different target parameters.

Corresponding Author:

1. INTRODUCTION

Historically, speech and its processing is handled in different ways in computer science, electrical engineering, linguistics, and psychology. The first steps of the development of these models started after Second World War when it began the discovery of computer, so, in the period from the 1940s until the late 1950s intensively was worked on the development of these models and speech intelligibility. Speech recognition, as an idea. firstly appeared several decades ago at scientific movies, where computers recognized speech and identify the person, no matter how fast and what language he spoke [1]. But even today, in reality it is not managed to design a program for speech recognition as was described in scientific movies.

Skype and Viber are computer and mobile applications used as a testing software, which provide communication through speech and writing. Skype and Viber use VoIP standard that means voice communication through the Internet [2]. These applications are included in the so-called web applications that function only by having access to Internet. Studies about the intelligibility of the word can be performed in different ways.

Since the invention of Alexander Bell, engineers and scientists have studied the phenomenon of speech communication as a communication between people through telecommunications equipment or communication between man and a machine. Starting from the 60's digital signal processing (DSP) and their presentation in the form of visualization and mathematical form have been somewhat abstract problem for mankind [3].

Today, modern technology has not yet been designed such that any software on the database have placed the words in Albanian language, except for some programs that work locally.

2. RESEARCH METHOD

2.1. The effect of Internet speed on speech intelligibility by using applications that work with the VoIP platform.

Some internet service providers offer Internet connection with wireless routers, so, in this case the conversations can be done in very practical, convenient and comfortable way using Skype and Viber. It should be taken into consideration the fact that internet connection with cable connection may be more efficient for the stability of the conversation between people because the speed of the internet is constant and at wireless connections it may vary [3], [4]. Also, the wireless connection may have an impact on instability of the communication since wireless connections are offered for access by many users and it affects the speed of the internet, then at connections without wires an impact can have effects such as interference [5], obstacles that arise due to the frequency bands and reflection of waves .

Another effect on using VoIP is the loss of packages and as a result is lost a part of the conversation, moreover an intelligibility of the speech is lost [2], [3]. Loss of packets due to load shedding and as a consequence the packs with audio data may remain on the network longer than is assigned to a frame, and if the time appointed is passed then the package is lost, which means that there is no destination until that moment. Internet services are divided into classes according to quality offered, ie the best of the class starts from A, B, C, up at the lowest quality under D [3].

Jitter (vibration) - represents standard deviations between packets or frames of data. The size of the jitter represents the evaluation of the ripple data.When transmission is implemented with optical fiber or other media packs it has a tolerance for a delay in the amount 5μs / km [6]. ITU recommends that delays of the jitter should not be greater than 150 ms for most applications and a limitation for applications with voice communication about 400mS [2], [6].

The total delay in the system is consisted of the following components:

1. The delay in the process of coding
2. The delay due to waiting
3. The delay of the transmission
4. The delay due to the delay variation and improvements of the delay variations
5. The delay in the process of decoding

3. RESULTS AND ANALYSIS

Calculation of Word Error Rate - measurement parameter for measuring intelligibility of speech is the word error rate. For some simple recognition systems (such as for example the isolated words), the performance is simply the percentage of lost words to total words. However, this measurement parameter is not effective because the known words sequences can contain up to three types of errors [9]. Similarly with the error of recognition of digits, the first error known as replacement of words, occurs when an incorrect word is accepted as a correct word. The second error, known as suppression of words, occurs when a spoken word is not known (ie, sentences have not recognized the spoken word at the entrance) [7]. And the third error, known as the introduction of words while processing this case happens when words involved are accepted by their knowledge (ie, the sentence is recognized and accepted with more words than is provided in entrance such as noise). One such example would be:

Spoken sentence at the entrance: Good evening, is there anything new from you?

Sentence understood and accepted at the exit: Good evening as much as you are, there is something from you!

The error rate is defined as the percentage of the words incorrectly accepted to the number of words uttered at the entrance [10].

$$WER = 100\% * \left(\frac{S+D+I}{|W|}\right) \quad (1)$$

- Substitutions - Replacement of words
- Deletions - Termination of words
- Insertions – Insertion of words
- W – Total number of words

$$Word\ error\ rate = 100\% * \left(\frac{\boldsymbol{Number\ of\ error\ words}}{\boldsymbol{Total\ number\ of\ words}}\right) \quad (2)$$

To understand measurements better, we will give an example of a conversation between a transmitter and a recipient using Viber.

- An example of how calculated WER, between the speakers and receiver using Viber.

REF: i *** ** UM the PHONE IS i LEFT THE portable **** PHONE UPSTAIRS last night
HYP: i GOT IT TO the ***** FULLEST i LOVE TO portable FORM OF STORES last night
Eval: I I S D S S S I S S

By applying the above expression

Word error rate (WER)=$100\frac{6+3+1}{13}$=76.9%

WER=76.9%[7],[8]

While intelligibility is counted as:
The scale of intelligibility of the words is defined as the percentage of correctly recognized words to the number of words uttered at the entrance.

$$\text{Speech understandability} = \frac{\textit{Number of words correctly recognized}}{\textit{Total number of words}} * \mathbf{100}\% \quad (3)$$

The level of satisfaction is expressed in percentage and understanding has several divisions according to the results issued [1].

Table 1. The Scale of Intelligibility of Speech Performance

Performance level	Excellent	Good	Enough	Weak
Intelligibility	76 - 100	66 - 75	61 - 65	30 – 60

The derived results are from a dialogue between two male persons. So at the transmitter, at the entrance of the system the text is read by Dardan Mehmeti age 25 with readable and clear tone, while the receiver or at the output is 30 years old Altin Shala who also assessed the results of these measurements. Regarding the methods and manner of measurements, is selected the way to test different texts to view and compare the results of intelligibility of speech in Albanian language depending on the texts that are going to be read for testing and in other languages worldwide and different models. Another achievement if we see the results is that Albanian language in terms of intelligibility and adaptivity for communication, ais similar to Serbian and Croatian [4].

In the table 2 is presented a WER statistic for converstation done by different age. All this is tested in English language by recording audio files and then tested by the speakers.

Table 2.WER for Communication Only Men and Only Women Stood Aside and Older

Word Error Rate (WER) %			
	Young	**Old**	**Difference**
Overall	30.4	40.4	10.0
Males	30.1	38.8	8.7
Females	32.4	46.1	13.7

While we are at the texts used for reading and interpetation it's much easier to use equivalence of titles with the following abbreviations:

Table 3.Types of Text Read

F1	*The text recognized by the receiver*
F2	*Unrecognized text for the receiver*
F3	*Pronunciation of 50 consonants with two letter words*
F4	*Pronunciation of the 150 most frequently used words in Albanian language*
F5	*Pronunciation of 50 longer words that are less used in Albanian language*

Measurements done with Skype – Specifications of quality of the internet for measurments done with the Skype application

- Download speed-100 Mbps
- Upload speed - 50 Mbps
- Ping - 30 ms
- The Ping between the Skype server and the PC we have made the measurements is Ping - 66ms
- Jitter - 21ms
- Line of site quality is B
- MOS - 4:29
- The distance measured from the local host of the Internet provider Star Link server in the Art Motion Pristina.

As we indicated above for the use of text abbreviations, at the beginning is measured the text for F1 with the title: amplifier circuits with many stages, a total of 477 words, and then so on, for the other texts.

Table 4. Measurement Results with Skype

Read texts	Speech understandability %	WER (Word Error Rate) %
F1	97%	3%
F2	95.5%	4.5%
F3	96%	4%
F4	99%	1%
F5	98%	2%

In the diagram presented at the Figure 1 we see the dependence of the error of words in relation to the text read.

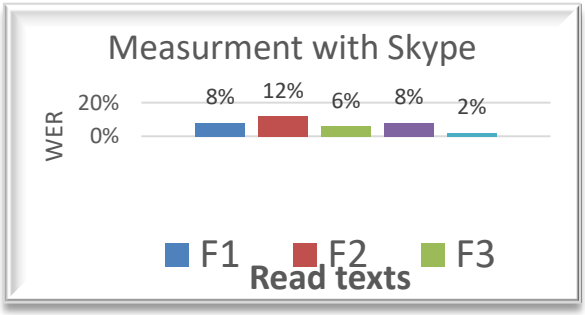

Figure 1. Graphical Representation of the Measurements with Skype

Measurements with Viber application

The data about quality of internet for Viber app measurements in a clean environment without noise.

- Download speed - 4Mbps
- Upload speed - 1.8 Mbps
- Ping - 48 ms
- Jitter - 1 ms
- Line of Internet quality is B
- MOS 4.1
- The distance of measurement from the local host server Ipko in Frankfurt, Germany

Just like in Skype we have used the same method for measurement in Viber also, but here we have used Viber in Smart Phones with specifications outlined above.

Table 5. Measurement Results with Viber

Read texts	Speech understandability %	WER (Word Error Rate) %
F1	92%	8%
F2	88%	12%
F3	94%	6%
F4	92%	8%
F5	98%	2%

From these data we present the following diagram presented in the Figure 2.

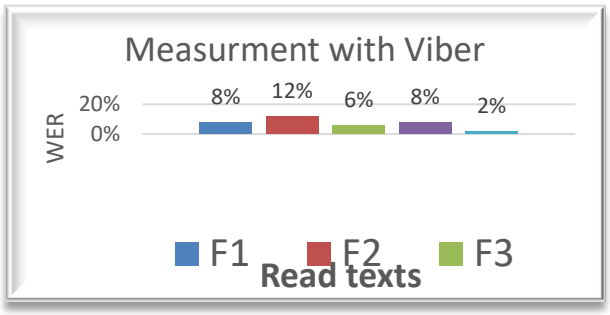

Figure 2. Graphical Representation of the Measurements with Viber

Below we present the difference between applications Skype, Viber and Zoiper in terms of intelligibility of speech in environments without noise, to see it as a summary of all texts in total meaning the average value of all texts.

Table 6. Final Measurement Results between Skype and Viber

Used Apps	Speech understandability % (Average value)
Skype	97.1%
Viber	91.6%

Whereas the presentation has a chart like the one presented in the Figure 3.

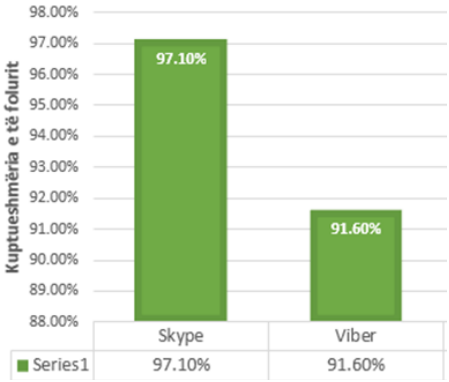

Figure 3. Graphical Representation of Measurements between with Viber and Skype.

4. CONCLUSION

Results achieved for the intelligibility of speech are perfect for all applications that we used, meaning the values are over 76% which is considered the highest degree. From this we conclude that the Albanian language is a language that is easily understood and the main reason why we get these positive results are the vowels which give the meaning of the words. Either so, we still remain reserved for these results because measurements made in this report are only between two persons who are familiar with each other, if the testing is made with more people who are unfamiliar between themselves, is expected to have lower values of intelligibility of speech. A reduction of intelligibility can be found at the measurements done between people who communicate and don't recognize each other which in our case measurements are carried out between persons who are known to each other and this is one of the reasons for this results with high values.

The future of this field is in that there is much to be done and required to work in groups from various fields engineering, programming, linguistics, because is equally challenging and also very necessary for this modern time. Truly, the forces of relevant national experts should be joint together to digitalize the Albanian language because this will be an advantage not only in speech intelligibility but also in many automatic systems such as robotics, medicine to the machines that work with signalization of the voice, which also are used in our country, then in the future the car manufacturing industry of information technology is meaningless to develop applications and machines and not integrate Albanian language on them.

REFERENCES

[1] Roger L. Freeman, *Fundamentals of Telecommunications*, IEEE Press, Willey Interscience, 2009.

[2] Thomas F. Quatieri. *Discrete-Time Speech Signal Processing*. Prentice-Hall, 3rd edition. 1996.

[3] J.Rodman. *The Effect of Bandwidth on Speech Intelligibility*. Commonwealth Telecommunications Organization, September 2006.

[4] Veton Këpuska. *VOn Wake-Up-Word Speech Recognition Task, Technology, and Evaluation Results against HTK and Microsoft SDK 5.1*.Invited Paper: World Congress on Nonlinear Analysts, Orlando 2008, *Journal of Nonlinear Analysis, Theory, Methods & Applications. & Klein*, T. 2008.

[5] H.K. Palo, Mihir Narayan Mohanty. *Classification of Emotional Speech of Children Using Probabilistic Neural Network*. International Journal of Electrical and Computer Engineering (IJECE) Vol. 5, No. 2, April 2015, pp. 311~317 ISSN: 2088-8708.

[6] D. Jurafsky and James H. Marti. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition Copyright 2006, draft of June 25, 2007.

[7] R. Vipperla..*Automatic Speech Recognition for ageing voices, Institute for Language*, Cognition and Computation School of Informatics University of Edinburgh 2011.

[8] Kayode Francis Akingbade, Okoko Mkpouto Umanna, Isiaka Ajewale Alimi. Voice-Based Door Access Control System Using the Mel Frequency Cepstrum Coefficients and Gaussian Mixture Model, *International Journal of Electrical and Computer Engineering (IJECE)* Vol. 4, No. 5, October 2014, pp. 643~647 ISSN: 2088-8708.

[9] Saeed V. Vaseghi. *Multimedia Signal Processing, Theory and Applications in Speech, Music and Communications, Department of Electronics*, School of Engineering and Design Brunel University, UK, 2007.

[10] L. Liu and L. Sun: *Performance Analysis of Voice Call using Skype*, Centre for Security, Communications and Network Research Plymouth University, United Kingdom, 2011.