# *Intelligibility of spoken Albanian language by using multimedia applications like skype and viber*

*Blerta Prevalla*[*]

*Altin Shala*[**]

**Abstract**

This paper intends to analyze subjective measurements of intelligibility of speech in Albanian language during the conversation between two people using applications which today is very used for communication such as Skype and Viber

Intelligibility of speech has to do with the clarity of speech that is heard, meaning of the entry of the system we have words without meaning while to the exit we mark what is heard. Speech intelligibility represents the percentage of correctly recognized words to the number of words uttered at the entrance. 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, provides the percentage of intelligibility (the words, sentences or syllables).

A concrete example of why is needed digitalization of the Albanian language and why should we design models and algorithms for this is the case of Google Search by Voice, where even if you speak a word that is not clear in English language, the system finds the closest potential word and generates it.

Methods of measurements are made at different speed of the Internet, in an environment without noise and with noise, in

[*] *Blerta Prevalla, Phd.Cand., Universiteti AAB, Fakulteti i Shkencës Kompjuterike, blerta.prevalla@aab-edu.net*

[**] *Altin Shala, MSc, Universiteti AAB, Fakulteti i Shkencës Kompjuterike,, altin.shala@aab-edu.net*

order to see the impact on understanding of the speech with different target parameters.
**Key words**: *the word error rate WER, intelligibility, noise, Skype, Viber, VoIP*.

## *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 the Second World War, when it began the discovery of computers, so, in the period from the 1940s until the late 1950s intensively was working on the development of these models and speech intelligibility[1]. Speech recognition, as an idea occurred several decades ago at scientific movies, where computers recognized speech and identify the person, no matter how fast and what language he spoke. But even today, in reality it is not managed to design a program for speech recognition as described in scientific movies.

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

[1] Jurafsky, Daniel and James H. Martin , *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*, 2006, Draft of June 25, 2007.

[2] *VoIP –Do të thotë komunikimi me zë përmes Internet Protokolit siç janë aplikacionet MSN në facebok, Viber edhe shume forume bisedash që funksionojnë me lidhje në internet

Since the invention of Alexander Bell, engineers and scientists have studied the phenomenon of speech communication as communication between people through telecommunications equipment or communication between man and machine. Starting from the 60's to 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 of Albanian language, except for some programs that work locally.

## *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 router and in this case is very practical, convenient and comfortable conversations 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. The reason is the speed of the internet which is constant and at wireless connections it may vary. Also, at wireless communication may have an impact on instability since wireless connections are offered for access by many users and it affects the speed of the internet, then at connections without wires even greater impact may have other effects such as interference, obstacles that arise due to the frequency bands and reflection of waves[4].

Another effect on using VoIP is the loss of packages and as a result is lost a part of the conversation, moreover an intelligibility

[3]Maxhuni, Adnan. *Gjetja e modelit për sintezën e të folurit nga tekstet e shkruara në gjuhën shqipe*, Prishtinë: Universiteti i Prishtinës 'Hasan Prishtina', 2014.faqe 10-11

[4] Shërbim për testim online të shfletuesit. http://voiptest.8x8.com/. Hapur më: 15.04.2016

of the speech is lost. 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 the quality offered, ie the best of the class starts from A, B, C, up at the lowest quality under D.

**Jitter (vibration)** -represents standard deviations between packets or frames of data.

The size of the jitter represents the ripple data evaluation. Any delay caused by jitter, are major network damage affecting the quality of the voice[5]. Increased delays depend on the physical distance between communicators and media which is used. When transmission is implemented with optical fiber or other media packs have a tolerance for a delay in the amount 5μs / km.

ITU[6] 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.

The total delay in the system consists 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 improvement measures to delay variations
5. The delay in the process of decoding

[5] Sun, L. *Spech quality prediction for voice over internet protocol networks* , University of Plymouth, Ph.D. January 2004

[6] ***ITU**-International Telecommunication Union (Unioni ndërkombëtar për telekomunikacion)

## *Testing and extracting results*

**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 of the known words sequences can contain three types of errors[7]. Similar to the error for the 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).

And finally we have 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 at 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.

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

- Substitutions - Replacement of words
- Deletions - Termination of words

[7] Limani, Myzafere. *Elektroakustika*, Ligjerata të autorizuara, Universiteti i Prishtinës, 2005

- Insertions – Insertion of words
- W – Total number of words
-

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

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

Complete word uttered is: In the next week I will travel to the US at Electronics VLSI training.

| Incoming Speaker | Në | Javën e | ardhshme | udhëtoj | Për në | SHBA | në trajnimet | për | Elektronikë | VLSI |
|---|---|---|---|---|---|---|---|---|---|---|
| Output Receiver | Për | Javën e | Tashme | ***** | Për në | SHBA | Në trajtimet | Në | Elektrikë | SI |
| Evaluation | S | | S | D | | | S | I | S | I |

By applying the above expression

Word error rate (WER) =$100\frac{4+2+1}{10}$=70%

WER=70%

**While understanding counts:**

Scale of understanding the words is defined as the percentage of correctly recognized words with the number of words uttered at the entrance.

$$\text{Speech understandability} = \frac{\boldsymbol{Number\ of\ words\ correctly\ recognized}}{\boldsymbol{Total\ number\ of\ words}} * \mathbf{100}\% \text{.....(4.3)}$$

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

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

Table 1. Scale of understanding the performance of speech

The derived results are made in the dialogue between two male persons. So in transducer system at the entrance to the reading of the text is Dardan Mehmeti age 25 with a 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 understanding the speech in Albanian language on depending texts that are read for testing based in other languages worldwide and different models. Another achievement is that the results of our language consulted in terms of understanding and adapting for communication model, are similar to Serbian and Croatian.[8]

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

| | |
|---|---|
| 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 for quality of the internet by measuring with Skype application

- Download speed-100 Mbps

[8] N. Caka, A.Caka:, "***Lista e 100 fjalëve të para të radhitura sipas dendurisë së paraqitjes në korpusin njëmilionfjalësh të gjuhës shqipe***", Universiteti i Prishtinës, Prishtinë, 2006.

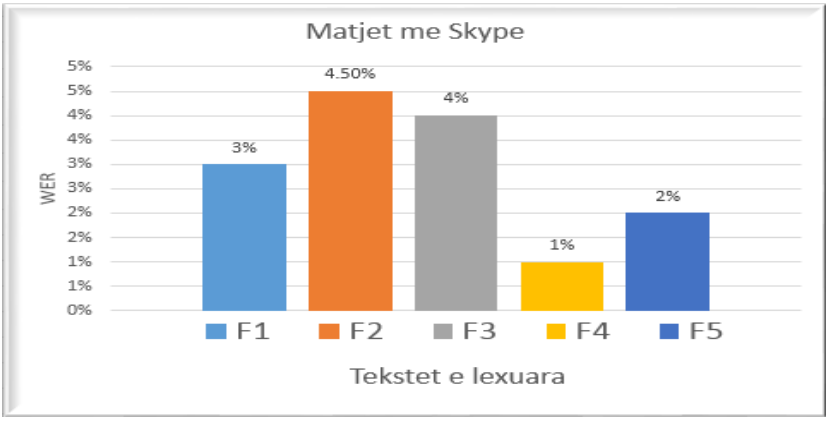
- upload speed - 50 Mbps
- Ping - 30 ms
- Ping between the Skype server and PC's that we have made the measurements is Ping - 66ms
- Jitter - 21ms
- Line of site quality is B
- MOS[9] - 4:29
- Distance measurement from the local host Internet providers Star Link server in the Art Motion Pristina.

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

| 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 below we present the dependence of the error of words in relation to the text read.

[9] ***MOS**-Mean Opinion Score (parametër matës objektiv për kualitetin e shërbimeve)




**Measurements with Viber application.**

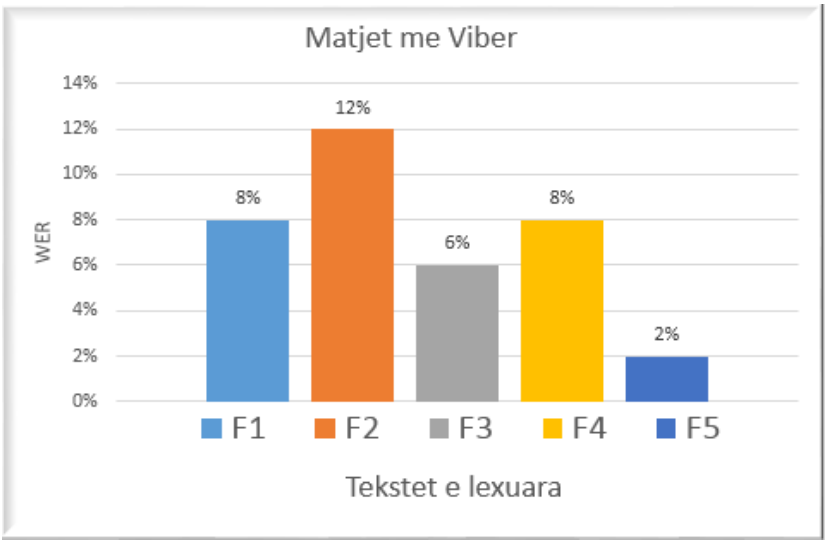
*Data about quality of internet for Viber app measurements in a clean environment without noise.*

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

Just like in Skype we have done the same measurement method for Viber but here we have used Viber on Smart Phone with specifications outlined above.

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

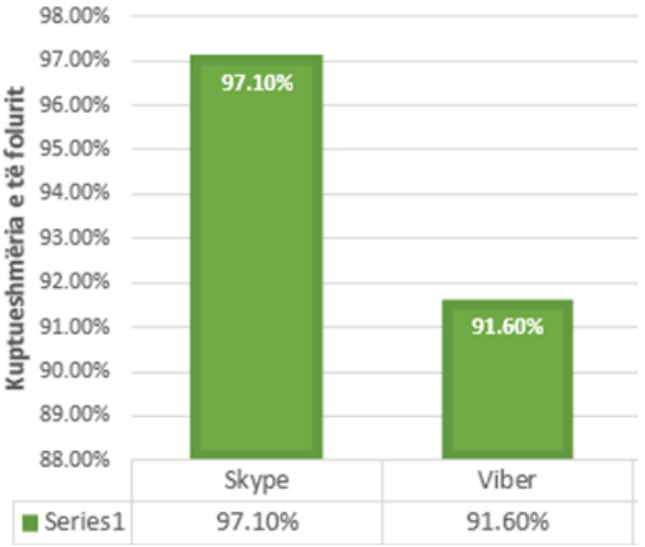
From these data we present the following diagram.




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

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

Whereas the presentation has a chart of this form.




## *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 words. We remain reserved for these results because measurements made in this report are made only between two persons who are familiar with each other, if any testing are made with more people who are unfamiliar between them, is expected to have lower values of intelligibility of speech, a reduction of understanding may be to the measurements when people who communicate do not 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 these high value results.

## *Recommendations*

For the future in this field, we can say 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 national relevant experts should be joint and digitalization of 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.

## *Biblography*

Këpuska, Veton. VOn Wake-Up-Word Speech Recogntion Task, Technology, and Evaluation Results against HTK and Microsoft SDK 5.1. Invited Paper: World Congress on Nonlinear Analysts, Orlando 2008, To appear in Journal of Nonlinear Anal ysis, Theory, Methods & Applications. & Klein, T. 2008.

Limani, Myzafere. Ligjerata të autorizuar, Prishtinë 2005.

Jurafsky. Daniel 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.

Beci, Bahri. Fonetika e gjuhës shqipe, Tiranë, 2004