Quantifying the Value of Subjective and Objective Speech Intelligibility Assessment in Forensic Applications

GIOVANNI COSTANTINI[1,2], ANDREA PAOLONI[3], MASSIMILIANO TODISCO[1,3]

[1]Department of Electronic Engineering
University of Rome "Tor Vergata"
Via del Politecnico, 1 - 00133 Rome
ITALY

[2]Institute of Acoustics and Sensors "Orso Mario Corbino"
Via del Fosso del Cavaliere, 100 - 00133 Rome
ITALY

[3]Ugo Bordoni Foundation
Viale del Policlinico, 147 - Rome
ITALY

massimiliano.todisco@uniroma2.it

Abstract: - Transcription from lawful interception is an important branch of forensic phonetics. Signals in that application context are often degraded, thus the transcript may not reflect what was really pronounced. In order to decide whether a given transcript generated from a lawful interception exercise reflects the views of the speakers instead of the transcriber's, an objective speech intelligibility measurement method is required. Usually, the intercepted signal can be affected by both speech intrinsic distortion and background/environmental noise distortion. Unfortunately, the original clean speech is never accessible to the forensic expert, who therefore must draw his assessment from the only available, distorted, signal. Consequently, the only way to assess the level of accuracy that can be obtained in the transcription of poor recordings is to develop an objective methodology for intelligibility measurements.

This paper addresses the issue by using three different objective approaches - namely the Signal-to-Noise ratio weighted with the "A" curves (S/NA), the Articulation Index (AI) and the Speech Transmission Index (STI) - to evaluate the intelligibility of a given signal. All of the three approaches were exercised with different types of noise, yielding results to be compared with speech intelligibility scores from subjective tests. The outcome gives high correlation evidence between objective measurements and subjective evaluations. Therefore, the proposed methodology is deemed rather useful to establish whether a given intercepted signal can be transcribed with sufficient reliability.

Key-Words: - Objective intelligibility, forensic phonetics, speech transmission index, transcript reliability.

1 Introduction

Intelligibility of speech refers to the amount of speech items that a normal listener can understand. More specifically the standard ISO 9921 [1] defines intelligibility as "the measurement of effectiveness in understanding speech." Intelligibility can be assessed at sentence level, at word level, and for each phoneme. Intelligibility plays a key role in communications; indeed, ensuring full intelligibility is the main purpose of any communication channel or any recording system.

In forensic applications it is crucial that the meaning of sentences and mentioned names reflect those actually uttered by the speakers rather than the views of the transcribers.

Covert recordings have become the most frequent sources of evidence in criminal trials, but in order to use the intercepted speech as a evidence it is mandatory to transpose it into written text. On numerous occasions the speech was almost unintelligible, however, some experts felt that they could draw from that signal a correct interpretation. Unfortunately, when the signal is almost understandable happen that the transcript does not correspond to anything that was said.

Difficulties of making a useful transposition of speech into written text are mainly due to words spoken in a low voice and/or covered by environmental noise. Probably the biggest threat to speech comprehension is competing noise, voices or other sounds reaching the listener.

In addition, as linguists know well, it is almost impossible to transform speech into written text without losing information.

In real applications inaccurate or misleading transcriptions are frequent due to the presence of both additive noise (background noise) and multiplicative noise (reverb). In many cases therefore there are harsh contrasts between the prosecutor and the defender about the transcription of poor recordings [2].

To assess the reliability of a transcript, it would be useful to have an intelligibility measure of the signal to be transcribed. Unfortunately, no subjective measurement can be used in forensic applications, because the content of the message is not known in advance, and therefore it is impossible to determine the percentage of words that have been accurately transcribed.

The only way to assess the intelligibility in forensic applications is to set up a system based on acoustic parameters which is able to predict the intelligibility of the measured signal.

Such a system would be also very useful in the forensic field to evaluate the performances of speech enhancement systems, and, more generally, in many other fields, to avoid the high cost of the subjective evaluation of signal intelligibility.

Objective measurements do not really measure the intelligibility but determine physical parameters to predict intelligibility according to a certain model.

Many objective speech intelligibility measurements have been proposed in the past [3-6]. Most of the literature in this field comes from information technology, where the problem is to study the impact of the transmission channel and the encoders on intelligibility of speech [7-9].

Three frequently used objective measurement methods are: the signal-to-noise ratio, with the noise filtered by an A-weighting curve (S/NA) [10], the Articulation Index (AI) [11,12], and the Speech Transmission Index (STI) [13].

Unfortunately, all these objective measurements need the clean signal to be available for comparison with the noisy signal.

All of them can be referred to as double-sided methods and are not suitable for predicting the intelligibility in forensic applications.

2 Assessment of Intelligibility

The "quality" of an audio signal is evaluated by three characteristics: the intelligibility, or the ability to accurately understand what is being said, the naturalness, or as the signal corresponds to that obtainable in direct listening and the quality, how the signal is pleasant.

These definitions have been formulated considering the analysis of the performance of a transmission system; in other words, we are interested to assess the sound quality that a transmission system with certain characteristics (bandwidth, signal to noise ratio, type of encoding) is able to guarantee. The measurement of the difference between the intelligibility of the output signal and the intelligibility of the input signal.

However, there are some applications, particularly forensic, where what we want to measure is the intelligibility of a signal starting from hard to hear audio.

The problem of evaluating the intelligibility of a single-side signal or having only the corrupted audio file that you intend to evaluate the intelligibility is very complex because the residual intelligibility depends on many parameters: the bandwidth, the signal to noise ratio, the type of noise, the signal type, the distortion, the encoding. In addition, the parameters that we have listed are not easy to estimate on the same signal we want to know intelligibility.

To assess the reliability and effectiveness of a transcription of speech signals, we must define an objective measure of intelligibility index closely correlated with the subjective performance of a group of listeners.

Traditionally, the intelligibility of speech refers to the accuracy with which a normal listener can understand a spoken utterance.

The known signal may consist of phrases, words or simple sounds without meaning (logatoms).

Algorithms for approximate intelligibility measures use a double-sided approach based on a comparison between the clean speech signal and the transmitted signal.

This approach is not usable in forensic applications because the expert witness has only the noisy version of the signal.

The ISO/TR 4870:1991 [14] outlines how the subjective intelligibility of speech changes according to the signal to noise ratio, where masking noise is defined as speech-shaped filtered white noise [14], to provide a noise spectrum that is somewhat representative of the everyday real-life noises, including the babble of many voices, that often interfere with speech communications.

Fig. 1 shows that speech to background noise ratio greater than 7.5 dB is required for adequate intelligibility (> 80%).

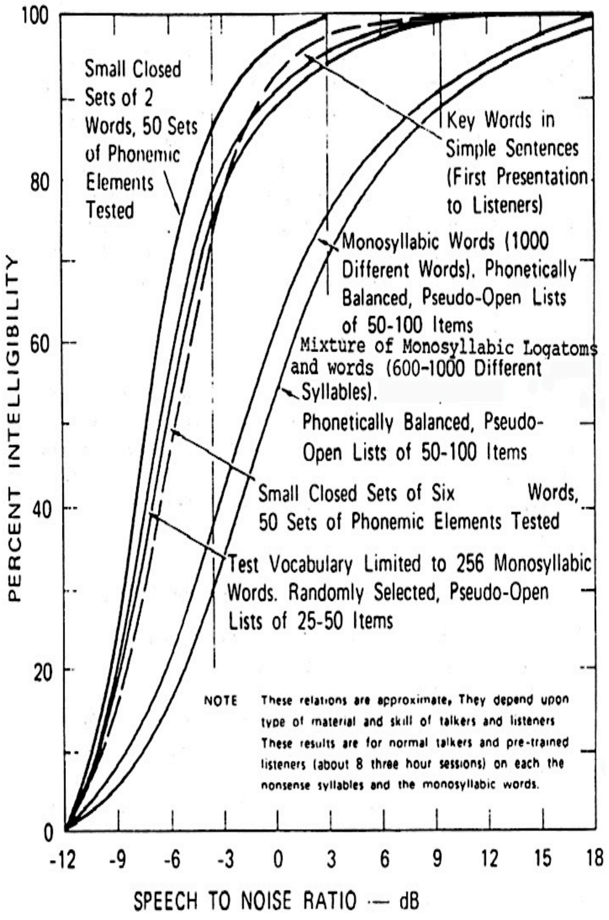


Figure 1: Intelligibility versus speech to noise ratio (data from ISO/TR 4870:1991)

3 Speech Corpus

Both subjective and objective tests are conducted using the corpus collected during the European project SAM EUROM 1 [15] and the Italian project CLIPS [16].

We extract three different corpora:

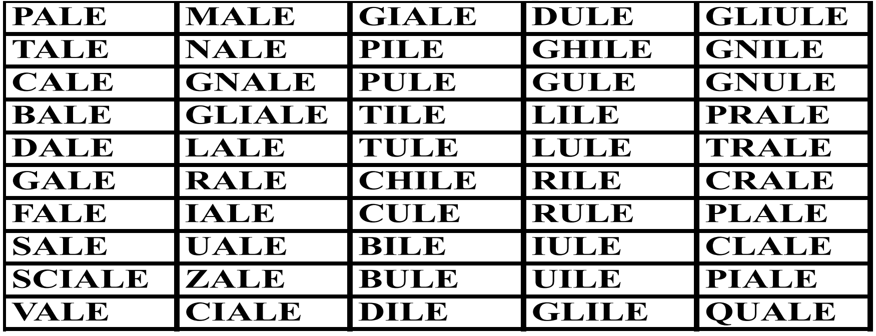
a) 50 rimed words
b) 24 Italian, meaningful or meaningless, sentences
c) 10 Italian simple sentences, 15 Italian simple words and 19 phonemes

In the corpora *a* degradations considered include additive noise. In particular, the corpus have been properly made noisy by adding Pink, Hammer and Babble noise. Each word appeared in five different degrees of signal to noise ratio (S/N = 2, 0, -2, -4, -6 dB) and read by 4 different voices, two men and two women. At the end of operations, therefore, can be found to have 60 different corpora each formed by 50 different words. Table I shows the complete speech corpus.

In the corpora *b* degradations considered include additive Babble noise and multiplicative noise [17,18]. The noisy speech appeared in three different grades of signal to noise ratio (S/N = +4, 0, -4 dB) each with two types of reverb (T60 = 0.95s and 2.03s), used to simulate Office and Lobby environment [19], so we obtain six differently degraded signals. Each sentence is read by 4 different voices: two men and two women. At the end of operations, therefore, can be found to have 24 different signals, each formed by different sentences. Table II shows the complete speech corpus.

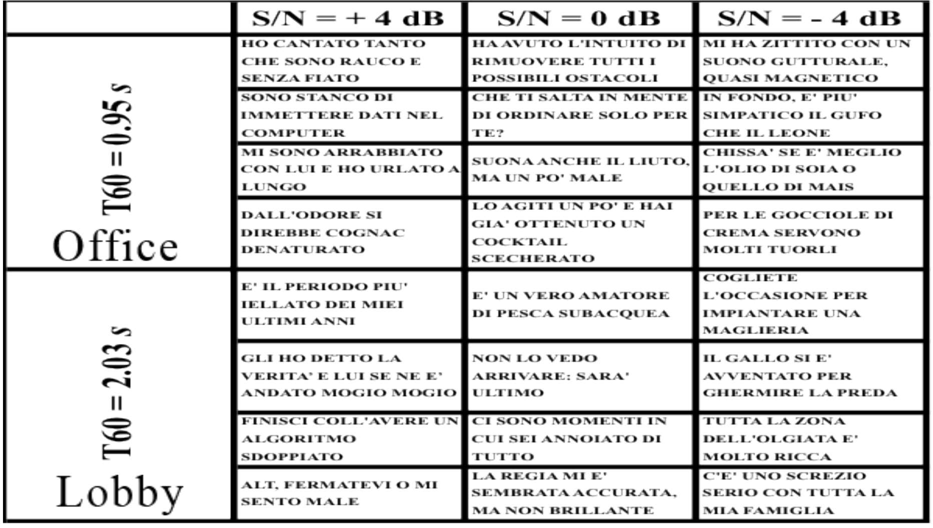
Finally, in the corpora *c* degradations considered include additive Babble noise. The noisy speech appeared in five different grades of signal to noise ratio (S/N = 6, 3, 0, -3, -6 dB) and read by a men. At the end of operations, therefore, can be found to have 50 different signals, each formed by different sentences, 75 different signals, each formed by different words and 95 different signals, each formed by different phonemes. Table III shows the complete speech corpus.

Table I – Speech Corpus *a*



PALE	MALE	GIALE	DULE	GLIULE
TALE	NALE	PILE	GHILE	GNILE
CALE	GNALE	PULE	GULE	GNULE
BALE	GLIALE	TILE	LILE	PRALE
DALE	LALE	TULE	LULE	TRALE
GALE	RALE	CHILE	RILE	CRALE
FALE	IALE	CULE	RULE	PLALE
SALE	UALE	BILE	IULE	CLALE
SCIALE	ZALE	BULE	UILE	PIALE
VALE	CIALE	DILE	GLILE	QUALE

Table II – Speech Corpus *b*



	S/N = + 4 dB	S/N = 0 dB	S/N = - 4 dB
T60 = 0.95 *s* Office	HO CANTATO TANTO CHE SONO RAUCO E SENZA FIATO	HA AVUTO L'INTUITO DI RIMUOVERE TUTTI I POSSIBILI OSTACOLI	MI HA ZITTITO CON UN SUONO GUTTURALE, QUASI MAGNETICO
	SONO STANCO DI IMMETTERE DATI NEL COMPUTER	CHE TI SALTA IN MENTE DI ORDINARE SOLO PER TE?	IN FONDO, E' PIU' SIMPATICO IL GUFO CHE IL LEONE
	MI SONO ARRABBIATO CON LUI E HO URLATO A LUNGO	SUONA ANCHE IL LIUTO, MA UN PO' MALE	CHISSA' SE E' MEGLIO L'OLIO DI SOIA O QUELLO DI MAIS
	DALL'ODORE SI DIREBBE COGNAC DENATURATO	LO AGITI UN PO' E HAI GIA' OTTENUTO UN COCKTAIL SCECHERATO	PER LE GOCCIOLE DI CREMA SERVONO MOLTI TUORLI
T60 = 2.03 *s* Lobby	E' IL PERIODO PIU' IELLATO DEI MIEI ULTIMI ANNI	E' UN VERO AMATORE DI PESCA SUBACQUEA	COGLIETE L'OCCASIONE PER IMPIANTARE UNA MAGLIERIA
	GLI HO DETTO LA VERITA' E LUI SE NE E' ANDATO MOGIO MOGIO	NON LO VEDO ARRIVARE: SARA' ULTIMO	IL GALLO SI E' AVVENTATO PER GHERMIRE LA PREDA
	FINISCI COLL'AVERE UN ALGORITMO SDOPPIATO	CI SONO MOMENTI IN CUI SEI ANNOIATO DI TUTTO	TUTTA LA ZONA DELL'OLGIATA E' MOLTO RICCA
	ALT, FERMATEVI O MI SENTO MALE	LA REGIA MI E' SEMBRATA ACCURATA, MA NON BRILLANTE	C'E' UNO SCREZIO SERIO CON TUTTA LA MIA FAMIGLIA

Table III – Speech Corpus *c*

SENTENCES	WORDS	PHONEMES
IL FULMINE HA COLPITO L'ALBERO	AEROPLANO	PALE
QUEL CANTANTE HA UNA BELLA VOCE	BIGLIETTO	TALE
ABBIAMO PREPARATO UNA TORTA MOLTO DOLCE	COLAZIONE	CALE
IL TRENO PARTIRÀ IN RITARDO	ELEGANTE	BALE
UN MESE DI VACANZA PASSA IN FRETTA	FATICA	DALE
QUEI SIGNORI NON SANNO MAI COSA FARE NÉ DOVE ANDARE	SETTIMANA	GALE
NEL GRANDE PARCO UN BAMBINO GIOCAVA CON SUO PADRE	GINOCCHIO	FALE
LA RAGAZZA CHE È APPENA ENTRATA, NON LA CONOSCO	GOVERNO	SALE
CHIAMAI IL MEDICO PERCHÉ AVEVO MALE AGLI OCCHI	INDUSTRIA	SCIALE
QUEL RAGAZZO NON DICE MAI LA VERITÀ	MACCHINA	VALE
	MODELLO	MALE
	OROLOGIO	NALE
	PADRONE	GNALE
	PRINCIPE	GLIALE
	RAGAZZO	LALE
		RALE
		ZALE
		CIALE
		GIALE

4 Intelligibility Evaluations

A first experiment was conducted in order to obtain the subjective intelligibility score.

The speech corpora *a* have been subjected to a group of 12 normal-hearing listeners, 4 for every degradation condition, using software developed for this purpose under the Max/MSP [20] environment, that deliver each item at chance many times as listener agreed. One test set consists of 50 different test signals. The listener fill in the proper space the word heard. Fig. 2 shows the application interface. The result of the subjective tests is shown in Fig. 3. We note that, for the same S/N, bubble noise leads to significantly higher values of the intelligibility than the other two types of disturbance.

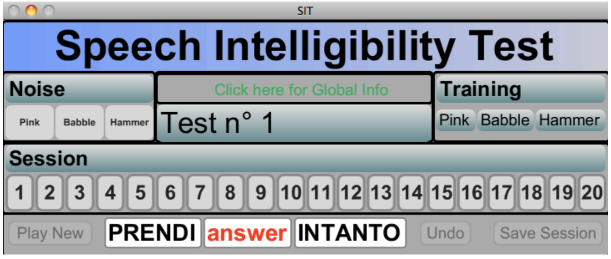


Figure 2: Interface used for the subjective listening tests on Corpus *a*

A second experiment was conducted to obtain intelligibility scores using the speech corpora *b*. The speech corpus was submitted to a group of 24 normal-hearing listeners. One test set consists of 24 different test signals. The listener fills in the proper space the sentence he/she has heard. Fig. 4 shows the application interface used for test.

Using the same corpora *b*, we also investigate the role of the noise suppression algorithms [21-24] on the intelligibility.

In [25-28] has been shown that noise suppression algorithms do not improve the intelligibility, but in some cases it is worsened.

The signals were given to 4 different experts in the field of speech enhancement in forensic applications asking them to operate a restoration of the signal to improving the intelligibility through the methods usually adopted by them.

The results of subjective measures of intelligibility are reported in Fig. 6. The values of intelligibility were obtained by averaging the values of the measures on the sentences belonging to the same class of degradation. The experts are identified by labels *Expert1* … *Expert4*, while the original signal is labeled *Original*.

We note that the enhancement methods used by the experts not significantly improve the intelligibility of the signal; indeed in some conditions the operation of enhancement leads to a significant deterioration of intelligibility. For example, the intelligibility of the condition of low degradation (+4 dB, office), equal to about 90%, is reduced to 50% by the enhancement system used by the Expert 4.

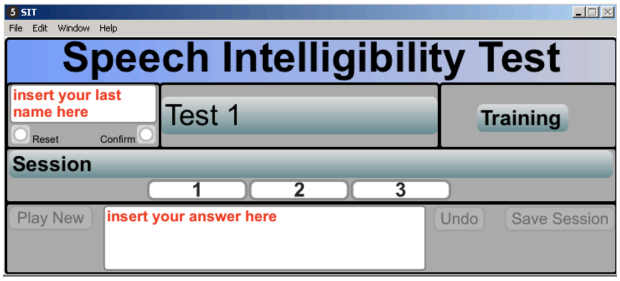


Figure 3: Interface used for the subjective listening tests on Corpus *b*

A third experiment was conducted to obtain intelligibility scores using the speech corpora *c*. The speech corpus was submitted to a group of 10 normal-hearing listeners. One test set consists of 44=10+15+19 different test signals regarding phonemes, words and sentences.

The listener fills in the proper space what he/she has heard. Fig. 4 shows the application interface used. The averaged results of the subjective tests, regarding sentences, words and phonemes are shown in Fig. 7. There is also the possibility of a training that allows to better understand the test, and to adjust the audio signal level.

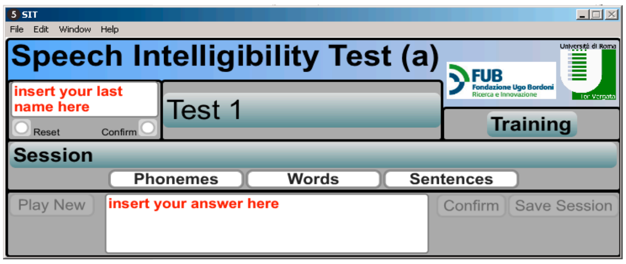


Figure 4: Interface used for the subjective listening tests on Corpus *c*

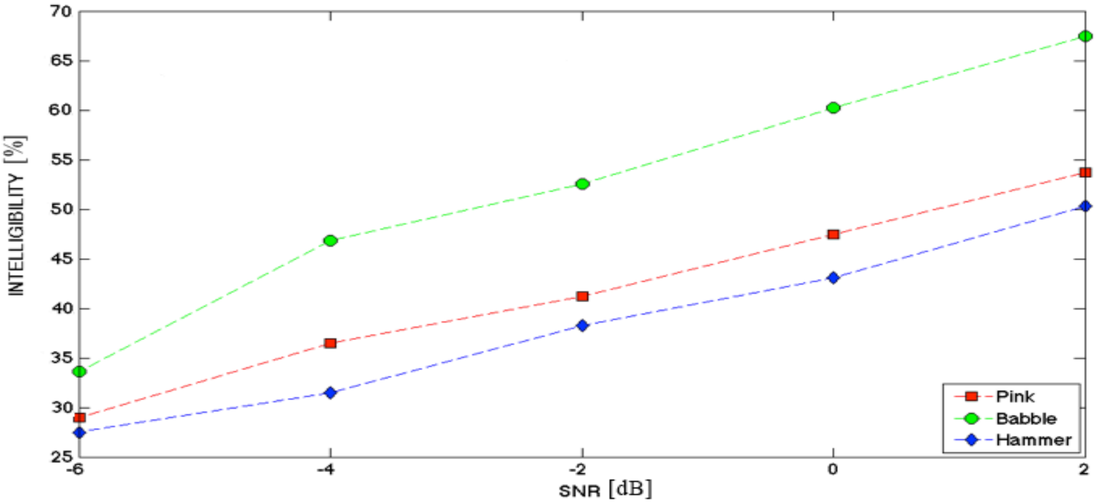


Figure 5: Subjective tests on Corpus *a*

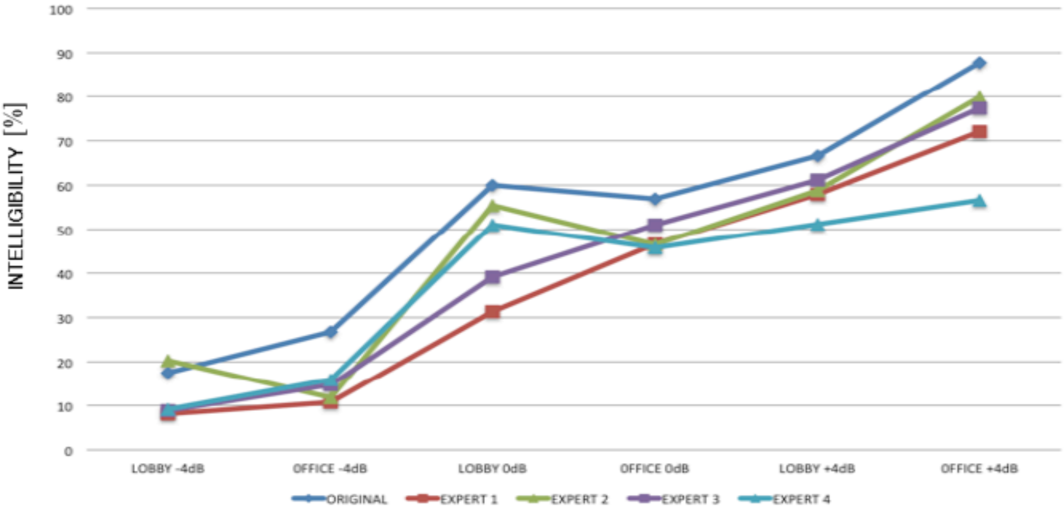


Figure 6: Subjective tests on Corpus *b*

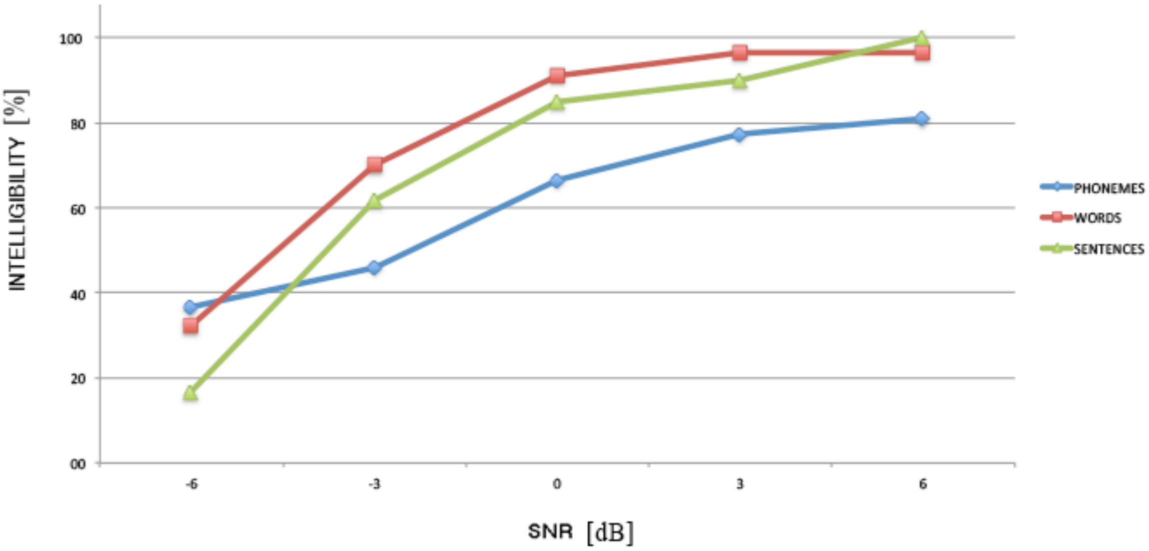


Figure 7: Subjective tests on Corpus *c*

5 Objective Measures

Three frequently used objective measurement methods were evaluated for use, based on: the signal-to-noise ratio, with the noise filtered by an A-weighting curve (S/NA), the Articulation Index (AI) and the Speech Transmission Index (STI).

The Signal to Noise ratio Weighted with the curves "A" (S/NA) is the simplest and easiest method proposed. It can be formally presented as

$$SN_A = S_A - N_A \tag{1}$$

where S_A is the A-weighted long-term average speech level and N_A the A-weighted long-term average level of background noise, measured over any particular time.

The Articulation Index (AI) estimates the intelligibility of speech from the spectral properties of the speech and the masking noise.

It has been shown to accurately predict performance in a variety of phonetically balanced intelligibility tests across a wide range of different listening environments [12]. The AI was calculated using the 20-band method described by [11].

In the Speech Transmission Index (STI) theory the intelligibility of speech is related to the preservation of the spectral differences between successive speech elements, the phonemes. This can be described by the envelope function. The envelope function is determined by the specific sequence of phones of a specific utterance. Unfortunately, all those objective measurements need the clean signal to be available for comparison with the noisy signal. All of them can be referred to as double-sided methods and are not suitable for predicting the intelligibility in forensic applications.

To this end, we propose a single-sided intelligibility measurement based on STI.

6 Single-sided STI-based measures

The intelligibility of speech is related to the preservations of the spectral differences between successive speech elements, the phonemes. This can be described by the envelope function. The envelope function is determined by the specific sequence of phones of a specific utterance.

The STI-based measure is computed as follows. The noisy signal were first bandpass filtered into seven octave bands starting from 125 Hz to 8000 Hz. The envelope of each band was computed using the power of the signal. In particular, let us consider a discrete time-domain signal x(n) filtered in the kth octave band, we define the envelop function as:

$$Env_k(m) = \frac{1}{N_e - 1} \sum_{n=mh}^{mh+N_e-1} h(n-mh)\left[x(n)\right]^2 \tag{2}$$

where Ne is the window size, h is the hop size, m ∈ {0, 1, 2,…, M} the hop number, h(n) is a finite-length sliding Hanning window and n is the summation variable. After that, we compute the normalized envelope spectrum as follows:

$$s_{k,f_i} = \frac{\left|\sum_{p=0}^{N_s-1} w(p) Env_k(p) \cdot e^{-\frac{i2\pi p f_i}{F_s}}\right|}{\sum_{p=0}^{N_s-1} Env_k(p)} \tag{3}$$

where Ns is the window size, Fs is the sampling rate, fi is the 14 frequencies in the range 0.63 Hz to 12.5 Hz at 1/3-octave step, w(p) is a finite-length rectangular window and p is the summation variable. The SNR in each band is computed as:

$$SNR_{k,f_i} = 10\log_{10}\left(\frac{s_{k,f_i}^2}{1 - s_{k,f_i}^2}\right) \tag{4}$$

and subsequently limited to the range of [-15, 15] dB The Transmission Index (TI) in each band is computed by linearly mapping the SNR values between 0 and 1 using the following equation:

$$TI_{k,f_i} = \frac{SNR_{k,f_i} + 15}{30} \tag{5}$$

For each octave band, the average TI over a specified frequency range gives the Modulation Transfer Index (MTI), as given by:

$$MTI_k = \frac{1}{n}\sum_{i=1}^{n} TI_{k,f_i} \tag{6}$$

Finally, the STI-based measure is obtained as a weighted mean of the MTI over seven octave bands, and is written:

$$STI = \sum_{k=1}^{7} W_k \cdot MTI_k \tag{7}$$

The sum of these weighting factors W_k is 1 [12].

Fig. 8 shows the block diagram of the STI-based measure.

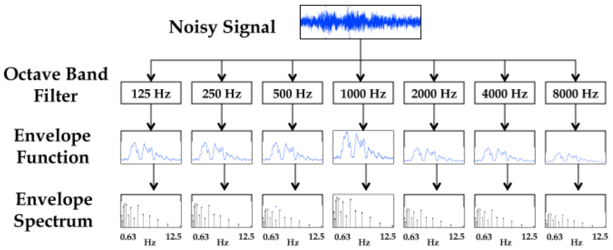


Figure 8: Block diagram of the STI-based measure

7 Experimental Results

Performances of the objective measures are presented in terms of the Pearson product-moment correlation coefficient *r* between the subjective intelligibility ratings and the objective measure, and is given by:

$$r = \frac{\sum_{i=1}^{n}\left(S_i - \overline{S}\right)\cdot\left(O_i - \overline{O}\right)}{\sigma_S \cdot \sigma_O} \tag{8}$$

where S and O are the subjective and objective scores, with means $\overline{S}$ and $\overline{O}$, and standard deviation σ_S and σ_O respectively, while n is the number the different degrees of signal to noise ratio considered. The coefficient ranges from -1 to 1 with 1 being the highest-correlated to subjective scores and vice versa.

A first experiment was conducted using for the intelligibility assessment the STI-based measure on Corpus *a*, *b* and *c*.

The experiment has highlighted the correlation between objective and subjective data in the particular conditions that are typical forensic applications. We note that all correlations are above 97%.

The results of these experiments are summarized in Fig. 9-11. Table IV shows the correlation between subjective and objective measures, for all degradations taken into account.

A second experiment was conducted using for the intelligibility assessment the STI-based measure on two real audio interceptions. We calculate a time-varying STI-based measure on a frame-by-frame basis. The short-time STI-based measure can be used to give a running measure of the speech intelligibility. In particular, we compute the STI-based measure using a sliding window of 500 milliseconds with 50% overlap. Finally, we link the STI-based measure to the Intelligibility by computing a linear fitting regarding the Pink noise curve shows in Fig. 9.

The second experiment shows the analysis of two speech samples of about 15 seconds concerning an actual case. These files are sampled at 8 KHz and quantized with 16 bits.

The first signal is very low quality in terms of S/N. The analysis shown in Fig. 12 allows to assess the intelligibility of individual segments (phrases).

In the figure you can see the trend of the signal amplitude, the sonogram (middle graph) and the intelligibility of different segments.

The estimated value of intelligibility is never more than 50%. The second speech segment has been recorded for a comparison. The result (Fig. 13) indicates that in this case the speech intelligibility is 100%.

8 Conclusion

The evaluation of the speech intelligibility is crucial to ensure the reliability of transcription. STI-based measures have proven to be reliable for predicting the intelligibility in forensic applications.

The present study demonstrates that the Speech Transmission Index is a good model in order to provide a tool for predicting speech intelligibility in additive and multiplicative noise conditions. The overall results show that the STI function provides a good estimate of speech intelligibility.

In particular, the experiments carried out have proven that our proposed STI measurement procedure is able to predict with sufficient accuracy speech intelligibility in conditions very close to those most frequently found in forensic applications, where both additive and multiplicative noise are involved.

Moreover, we developed a standalone application that operates a short-time STI-based measure; this application allows us to compute the objective intelligibility locally on a noisy signal, using window length of 500ms.

Interested readers are invited to download our system from the site indicated below and test it on their own signals.

http://voice.fub.it/SSIM/

Table IV – Correlation between subjective and objective measures

Corpus *a*		**Corpus *b***	**Corpus *c***
Pink	0.99	0.98	0.98
Babble	0.97		
Hammer	0.99		

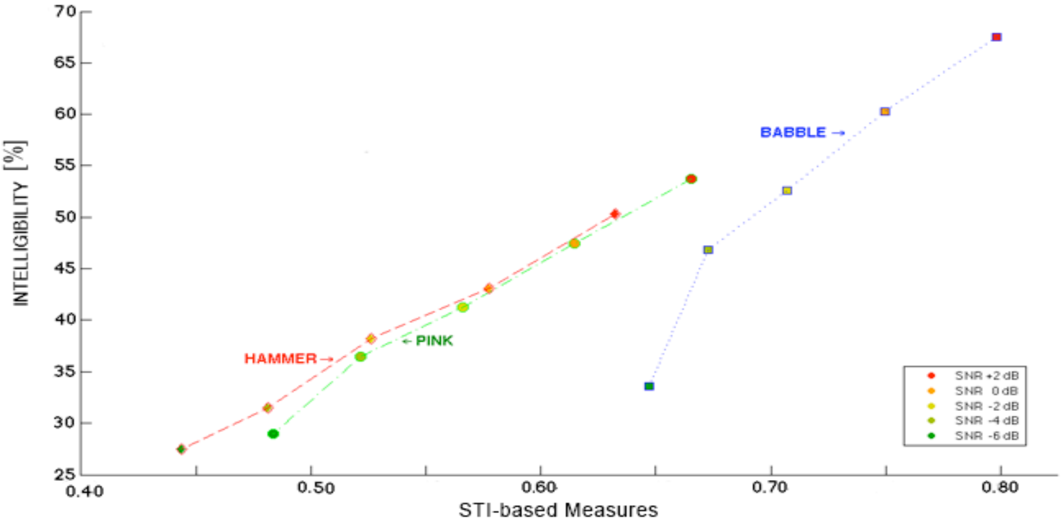


Figure 9: Corpus *a*: subjective intelligibility versus STI-based measures

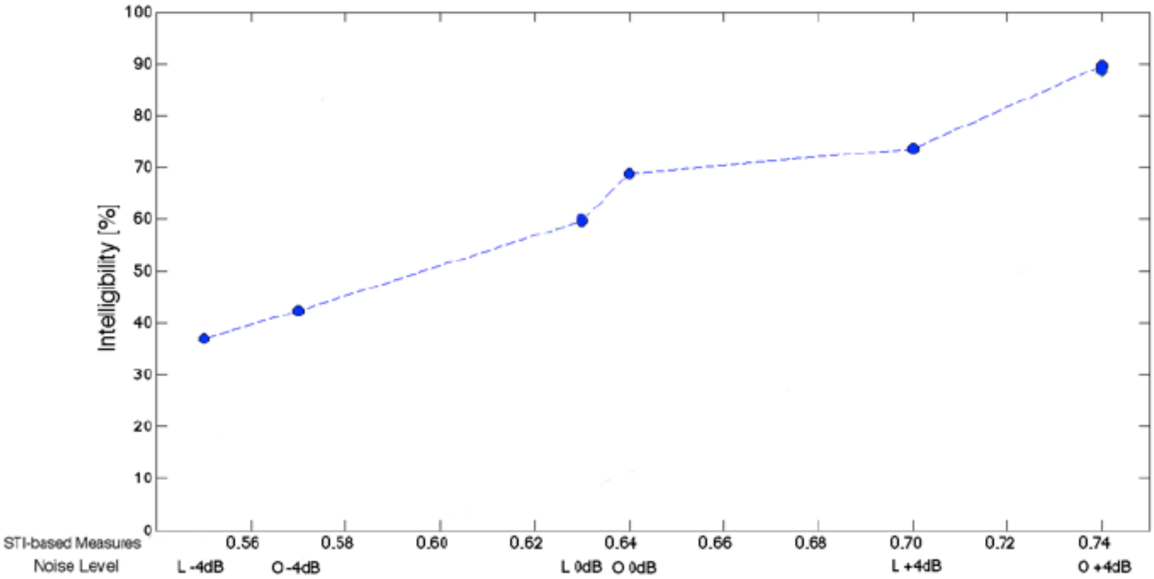


Figure 10: Corpus *b*: subjective intelligibility versus STI-based measures

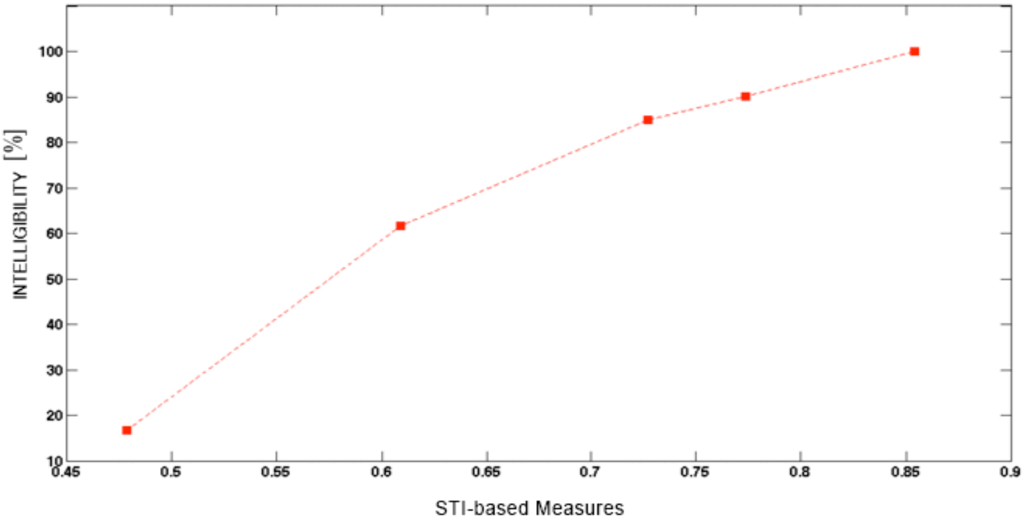


Figure 11: Corpus *c*: subjective intelligibility versus STI-based measures

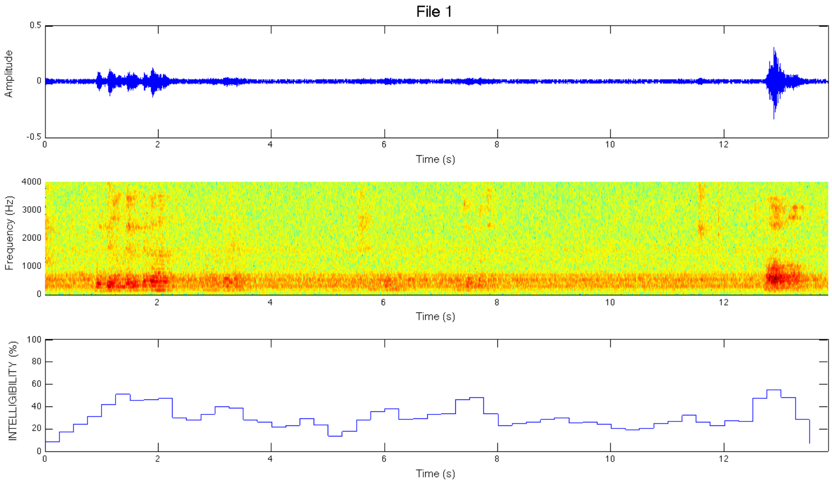


Figure 12: Intelligibility score versus time in a real case with poor quality signal. The intelligibility has been predicted using STI-based measurement.

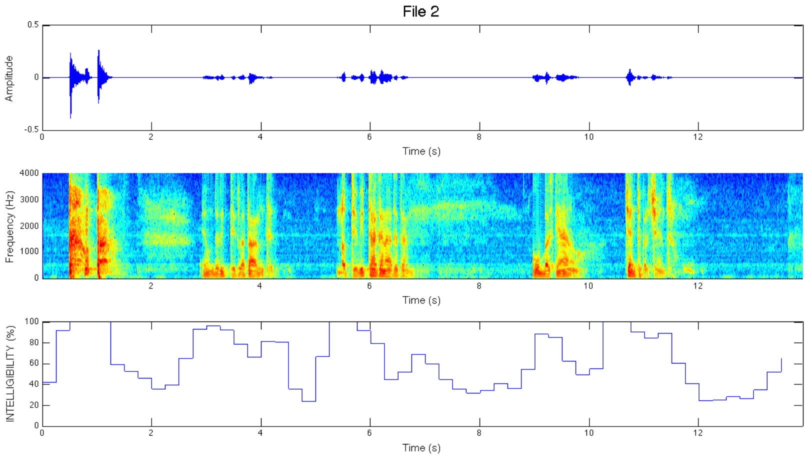


Figure 13: Intelligibility score versus time in a real case with good quality signal. The intelligibility has been predicted using STI-based measurement.

References:

[1] ISO 9921-1, Ergonomic assessment of speech communication – Part 1: Speech interference level and communication distances for persons with normal hearing capacity in direct communication (SIL method), International Standards Organization, 1996.

[2] Fraser, H., "Issue in transcription: factors affecting the reliability of transcripts as evidence in legal cases", Speech Language and the Law, 10 (2), pp. 203-226 2003.

[3] Steeneken, H. J. M., "The Measurement of Speech Intelligibility ", TNO Human Factors, Soesterberg, the Netherlands, 2002.

[4] Ma J., Hu Y., Loizou C.: "Objective measures for predicting speech intelligibility in moist conditions based on new band importance functions" JASA 125, May 2009.

[5] Costantini G., Todisco M., Perfetti R., Paoloni A., Saggio, G, "Single-Sided Objective Speech Intelligibility Assessment based on Sparse Signal Representation", IEEE International Workshop on Machine Learning for Signal Processing, Sept. 23–26, 2012, Santander, Spain.

[6] Costantini, G, Paoloni, A, Todisco, M, "Objective speech intelligibility measures based on Speech Transmission Index for forensic applications", 39th International AES Conference on Audio Forensics: Practices and Challenges. Hillerød, Denmark, pp. 182-188, 2010.

[7] Kitawaki, N., and Yamada, T., "Subjective and Objective Quality Assessment for Noise Reduced Speech", ETSI Workshop on Speech and Noise in Wideband Communication, May 2007, Sophia Antipolis, France

[8] Liu W. M., Jellyman K. A., Evans N. W. D., and Mason J. S. D., "Assessment of Objective Quality Measures for Speech Intelligibility", INTERSPEECH 2008, 9th Annual Conference of the International Speech Communication Association Brisbane, Australia September 22-26, 2008

[9] Voznak, M., "Non-intrusive speech quality assessment in simplified E-model", WSEAS Transactions on Systems, Volume 11, Issue 8, August 2012, Pages 315-325

[10] Hu Y., Loizou, P.C. "A Comparative Intelligibility Study of Speech Enhancement Algorithms", Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on Volume 4, Issue, 15-20 April 2007, Page(s): IV-561 - IV-564.

[11] Kryter K., "Methods for the calculation and use of the Articulation Index", JASA 34, 1689–1697 November 1962.

[12] Kryter, K., ANSI S3.5-1969, ‘‘American National Standards Methods for Calculation of the Articulation Index’’ American National Standards Institute, New York, 1969.

[13] Payton K. L. "A method to determine the speech transmission index from speech waveforms", JASA 106, 3637-3648, 1999.

[14] ISO/TR 4870, Acoustics – The construction and calibration of speech intelligibility tests, 1991.

[15] Chen D., Fourcin A., et alii, "EUROM A spoken language resource for the EU", ESCA EUROSPEECH '95 Madrid September 1995.

[16] Leoni F. A., "The CLIPS Corpus", available at http://www.clips.unina.it.

[17] Costantini G., Uncini A., "Real-time room acoustic response simulation by an IIR adaptive filter", Electronics Letters , vol. 39, Issue 3 , 6 Feb 2003, pp. 330-332.

[18] Costantini G., Casali D., and Uncini A., Suitable loss function intended for adaptive simulation of room acoustic response, ECCTD '03, European Conference on Circuit Theory and Design, September 1- 4, 2003, Kraków, Poland, pp. I-129 - I-132.

[19] Costantini G., Casali D., "Adaptive Room Acoustic Response Simulation: a Virtual 3D Application", WSEAS Transactions on Acoustics and Music, Issue 1, Volume 1, January 2004, pp. 25-28.

[20] Cycling74 Max/MSP, documentation available on the web at: http://cycling74.com/

[21] Costantini G., Casali D., "Speech Noise Reduction Using Adaptive Spline Neural Networks", WSEAS Transactions on Circuits and Systems, Issue 1, Volume 3, January 2004, pp. 155-158.

[22] Dutta, M., Vig, R., "Speech compression with masked modulated lapped transform and SPIHT algorithm", WSEAS Transactions on Systems, Volume 4, Issue 11, November 2005, Pages 2157-2162

[23] Costantini G.,. Carota M, "Multilayer Spline Neural Networks for Speech Denoising in Frequency Domain", WSEAS Transactions on Systems, Issue 2, Volume 3, April 2004, pp. 569-572.

[24] Costantini G., Casali D., “Spline neural networks for speech noise reduction operating in time and frequency domain”, ECCTD '03, European Conference on Circuit Theory and Design, September 1- 4, 2003, Kraków, Poland, pp. III-173 - III-176.

[25] Boll S. F., “Suppression of acoustic noise in speech using spectral subtraction”, IEEE Transactions on Acoustics, Speech, Signal Processing, vol. ASSP-27, pp. 112-120, 1979.

[26] Boll S. F., “Speech enhancement in the 1980’s: noise suppression with pattern matching,” in Advances in Speech Signal Processing, S. Furui and M. M. Sonhdi, Eds., New York: Marcel Dekker, 1992.

[27] Hu Y., Loizou, C., A Comparative Intelligibility Study of Speech Enhancement Algorithms, Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on Volume 4, Issue, 15-20 April, Page(s): IV-561 - IV-564, 2007.

[28] Jianfen M., “Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions”, J. Acoust. Soc. Am. 125 (5), May 2009.