ACÚSTICA 98

ACCURATE PREDICTION OF SPEECH INTELLIGIBILITY IN LARGE REVERBERANT AND NOISY ENCLOSURES

FILLOL C. and LEGROS C.

Laboratoire d'Acoustique de Toulouse-Le Mirail
5, allées Antonio Machado 31058 Toulouse, France.
05 61 50 36 73
fillol@cict.fr

SUMMARY

The well-known methods such as the speech transmission index STI or RASTI show important deviations between subjective intelligibility tests and results calculated for large reverberant and noisy enclosures. This study presents a new objective method as an accurate predictor of speech intelligibility in such spaces. This model is based on the evaluation of sound energy ratio which separately reflect the influences of room, background noise level and intrinsic characteristics (frequency response, directivity) of multiple loudspeaker system. The validity of this approach is illustrated by considering several combinations of room (church, underground stations), signal-to-noise ratio (-15dB(A) to +10dB(A)) and loudspeaker system (central system, distributed loudspeakers system). A total of 150 configurations has been measured. A correlation coefficient of r=0.92 and a standard deviation of σ=6% between results calculated from the prediction model and subjective intelligibility tests have been found.

INTRODUCTION

In some large enclosures, such as underground stations or factory halls, which involve long reverberation time, low signal-to-noise ratio and represent more difficult environments for public address systems, the prediction of speech intelligibility is not satisfactory. Nevertheless, it is necessary to characterize the quality of sound systems in such spaces in order to deliver audible announcements (emergency calls, commercial calls) to group of listeners. The official standard method RASTI shows important deviations between subjective intelligibility tests and results calculated for large reverberant and noisy enclosures. Therefore, it was necessary to develop objective method allowing the estimation of speech intelligibility in such spaces. The present work describes this procedure and compares the accuracy of various predictors of speech intelligibility from a set of measurements of both physical quantities and speech intelligibility scores in real rooms.

PROCEDURE

Both subjective tests and RASTI measurements were performed in five rooms with volumes from 600-40000 m^3 and 1-kHz reverberation times from 2-10s. Several combinations of signal-to-noise ratios (-15dB(A) to +10dB(A)) and loudspeaker systems (central system, distributed loudspeakers system) were used.

Subjective testing

Speech intelligibility scores were obtained using 10 lists of 34 triphonemic phonetically balanced French words. The word lists were similar to those used by Lafon [1] but they were transformed in order to be homogenized and perfectly balanced with French language. When they were presented to the listener, the words were embedded in a carrier sentence. This carrier sentence is necessary to take into account the interfering effects of reverberation. Words are scored according to phonetics instead of spelling. For each receiver position, the scores of four to nine subjects were averaged to obtain one score representative of the receiver location. For each source-receiver combination, background noise was emitted by a loudspeaker positioned in the same plan or mixed with speech message in order to avoid the localization effect (Cocktail party). Subjects varied in age from approximately 20 years old high school students to working age adults. According to configurations, the speech message was emitted by central loudspeaker or by multiple loudspeaker system. The central loudspeaker was particularly used for measurements in churches. This loudspeaker is a sphere of 30 cm diameter having directional properties similar to a human speaker. Several positions source-receiver were chosen in the expectation of yielding a wide range of intelligibility scores.

Objective measurements

For each room, source and listener position combination, impulse responses were recorded using MLS method. From these recorded pulses, the Definition D_{50} and the reverberation times were calculated. Background noise level and speech level recordings were made at each receiver location. For the different type of loudspeakers, polar responses and impulse responses were taken in anechoic chamber using multiple microphone locations in order to measure vertical and horizontal radiation.

PREDICTIVE METHOD

The new predictor is based on the concept of the ratio of useful/detrimental sound energy [2]. Moreover, it take in account the Definition, speech and background noise levels and two criteria based on sound energy ratio which separately reflect the influence of frequency response and directivity of sources. This model is based on an estimation of phoneme speech intelligibility following "S-shape" regression law:

$$I(\%)=100^{*}\left(1-10^{-\frac{(S/B)_{eq}+40}{60*q}}\right)^{n}, \qquad [1]$$

where $(S/B)_{eq}$ is the equivalent signal-to-noise ratio for a dynamic of 60dB which include level variations from whispered voice to shouted voice. n and q are parameters of regression.

The equivalent signal-to-noise ratio is defined by

$$(S/B)_{eq}=10\log\left(\frac{D_{50}^{s}+R_{dir}+R_{rf}}{(1-D_{50}^{s})+10^{-\frac{S/B}{10}}}\right). \qquad [2]$$

• The sound energy ratio D_{50}^{s} is a criteria which performs effects of room. This ratio is calculated in the same way as the Definition D_{50} except that the source-enclosure impulse response undergo a deconvolution operation of loudspeaker impulse response. $(1 - D_{50}^{s})$ symbolize detrimental effect of reverberation on speech intelligibility. The influence of room is assimilated to Definition omitting direct sound energy.

• R_{rf} and R_{dir} characterize the influence of frequency response and radiation of a single loudspeaker or a multiple loudspeaker system on the intelligibility score. The criteria R_{rf} compare a reference filter of flat frequency response between 100-4000Hz with frequency response of source in the same range. The gain of reference filter corresponds at mean level of source frequency response in the range of interest. Energy of loudspeaker frequency response included in $\pm 1.5dB$ of mean level, for the range 100-4000Hz, is counted useful. Peaks and dips are considered detrimental for speech intelligibility. R_{rf} is therefore compared to a useful/detrimental ratio of spectral irregularities.

• The sound energy ratio R_{dir} is the ratio between the Definition D_{50} calculated with global impulse response and the "Definition" D_{50}^{s} calculated with room impulse response. This criteria expresses the proportion of energy coming from reference axis to off axis radiation.

• The model takes into account the background noise by signal-to-noise ratio in dB(A). Nevertheless, it can't be used for impulse noise such as clap but just steady-state noise.

RESULTS

Correlation of predictive scores with subjective score

Figure 1 plots the results of mean speech intelligibility test scores versus $(S/B)_{eq}$. The line shown on the figure is the result of fitting the "S-shape" regression law to the data.

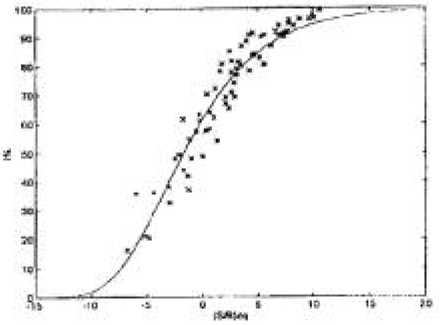

Figure 1: Measured speech intelligibility scores versus $(S/B)_{eq}$.

This curve only shows the intelligibility score of three large reverberant and noisy enclosures (churches, factory hall) using a central system. A correlation coefficient of r=0.95 and a standard deviation of $\sigma = 6.4\%$ between results calculated from prediction model and subjective intelligibility tests is found. Speech intelligibility scores were significantly related to $(S/B)_{eq}$. The accuracy of predictive model could be improve, principally for scores less than 25%, if one pay attention to the subject behaviour in difficult hearing conditions.

S/B(A)	d en m	RASTI	I_c(%)	I_m(%)	I_s(%)	I_p(%)
+10	2	0.61	87	93	96.6	95
	8	0.29	32	71	90	92
	16	0.23	17	43	77.5	79
+5	2	0.53	79	90	96.1	94
	8	0.27	30	56	83.6	86
	16	0.15	0	32	66.7	77
0	2	0.4	54	85	93.9	91
	8	0.19	8	43	70.6	79
	16	0.12	0	19	49.4	66

Table 1: RASTI measurements, sujective and predictive intelligibility scores.

Comparison

Table 1 summarizes the results calculated from RASTI measurements and predictive method for a church with volume 40000 m^3 and a 1-kHz reverberation time of 10s. The results are given for three signal-to-noise ratio from 0dB(A) to +10dB(A) and for three receiver positions from 2 to 16 m. The RASTI values cannot be directly compared to predicted scores by the model. To establish a relationship between these different values, the RASTI measurements must be convert to word scores (I_c(%)) [3] as also intelligibility scores predicted (I_m(%)) [4]. The model only calculates phoneme intelligibility scores. On the other hand, to show the validity of predictive method, the subjective scores I_s(%) and the predictive scores I_p(%) are given without modification (Table1). The RASTI values greatly under estimate speech intelligibility. The average standard deviations are 18% for RASTI prediction and 4% for predictive model. The results show that intelligibility calculated by RASTI measurements can lead to serious errors in the large reverberant enclosures. It is important to note that under the same speech conditions, the predictive model estimates intelligibility accurately.

CONCLUSION

Nowadays, it is important to characterize the acoustic quality of sound systems in factory halls and public spaces in order to deliver intelligible announcements. The actual predictive method don't correctly estimate the speech intelligibility in the large reverberant and noisy enclosures. The present study proposes a predictive model based on the evaluation of sound energy ratio which separately reflect the influences of room, background noise level and intrinsic characteristics of sources. This model predicts accurately the intelligibility scores for central systems in different reverberant spaces. A correlation coefficient of r=0.9 and a standard deviation of $\sigma = 6\%$ for a multiple loudspeaker system formed of four sources located on a plane network have been also found. To estimate the validity of predictive model, the experimental database must be increased with more complicated acoustic configurations for distributed systems. The construction of this model can lead to the optimization of sound systems steering recommendations of positioning and radiation characteristics.

REFERENCES

[1] Zuliani P, Gamba R. and Legros C. , "Intelligibilité de la parole dans le bruit: élaboration d'une méthode de diagnostic in-situ", LAUTM Rapport 85 333 (1988).

[2] Bradley J.S., "Predictors of speech intelligibility in rooms", J.Acoust.Soc.Am 80 (3), 837-845 (1986).

[3] Faiget L., "Etude comparative de critères d'estimation de l'intelligibilité: RASTI et Méthode directe", DEA Université Paul Sabatier (1993).

[4] Fletcher H. and Galt R.H., "The perception of speech and its relation to telephony", J.Acoust.Soc.A 19 (1), 90-119 (1950).