Computer simulation of the acoustics of the Ear of Dionysius (Syracuse-Italy) in order to evaluate objective descriptors of speech intelligibility

Gino Iannace[1], Elvira Ianniello[2], Amelia Trematerra[3]
[1] *Department of Architecture– Second University of Naples, Italy, gino.iannace@unina2.it*
[2] *phd, via S. Rocco 8, Parco Domus 84016 Pagani (SA) Italy, eianniello@unisa.it*
[3] *Department of Architecture– Second University of Naples, Italy, amelia.trematerra@unina2.it*

Introduction

A myth handed down for a long time says that the tyrant of Syracuse Dionysius I (the Elder, ca. 432–367 BC) could understand the speech of war prisoners through a window in the upper part of the innermost wall of the grotto named Ear of Dionysius. Sabine himself who had visited this cave had cast serious doubts on this possibility [1]. Few years ago the authors carried out a set of acoustic measurements to characterize the acoustics of this site known worldwide for its peculiar reverberation [2]. Unfortunately, they could perform only a single measurements of the sound transmission from the floor level of the grotto to the listening spot of the tyrant to estimate a degree of truth of the myth. This paper reports a numerical study of the acoustics of the Ear of Dionysius aimed at the evaluation of current room-acoustics parameters related to speech intelligibility when a speaker is located at the ground level of the cave and the listener occupies the supposed spot of the tyrant. The room-acoustics simulation was carried out with the aid of the specialized software Odeon.11 [3]. The results of the simulation still suggest reasonable doubts about the possibility the tyrant could understand the speech of prisoners

Numerical model and calibration

In the lack of detailed plans and sections of the grotto, the authors performed the survey of the internal geometry of the cave with a suitable laser-based equipment. The collected data were used to implement the 3D geometrical model for the software Odeon. Other necessary information was input into the software and a trial and error calibration of the numerical model was performed on the basis of values of T30 measured at ground level as reported in ref [2]. The chosen data regarded 24 space-averaged values of T30 (twelve receivers and two sound source locations at the height of 1.6 m over the floor) for each one-octave band from 63 Hz to 4 kHz. Most internal surfaces were smooth and hard so they were given reasonably low values of the sound absorption coefficient. A scattering coefficient about 0.05, and thermal (t = 15 °C) and hygrometric (RH = 50 %) properties of air were also assigned. By a little tweaking of the sound absorption coefficients the differences between the space-averaged octave band values of T30 were reduced satisfactorily. Figure 1 shows the sound source and receivers locations in the plan of the cave. Figure 2 displays the comparison between the calculated space averaged values ± one standard deviation and the corresponding measured values

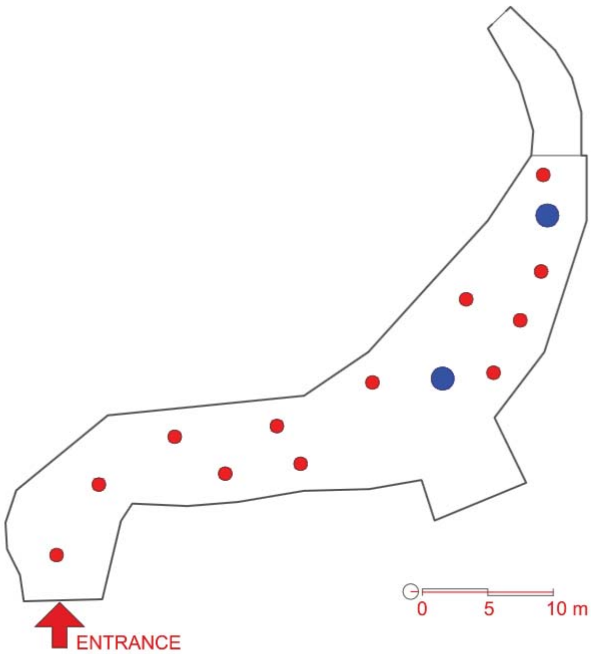

Figure 1: Plan of the Ear of Dionysius showing two sound source locations (blue dots) and twelve receiver points (red dots).. Measured responses for 24 source/receiver pairs were used for the calibration of he Odeon model.

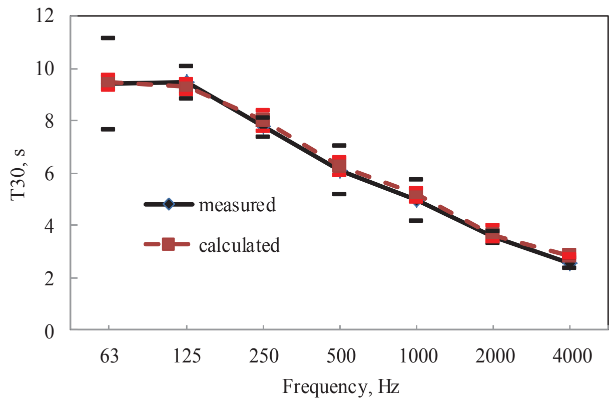

Figure 2: Calculated space-averaged T30 ± 1 standard deviation vs corresponding measured values.

Evaluation of speech intelligibility parameters

Actually speech intelligibility in rooms is a complex topic related to various factors such as the response of the system between the mouth of the speaker and the ears of the listener, the background noise, the modalities used by the speaker to issue his talk, the degree of familiarity of the listener with the spoken language and other factors as well. However, descriptors of speech intelligibility of different nature have been proposed which are correlated statistically with the degree of intelligibility of various types of speech material, e.g [4]. Their primary aims are the prediction for design purposes and uniform (standard) check of the intelligibility of speech transmitted through existing systems. Such state of the matter prevents a truly reliable answer to a question like: "*which degree the tyrant could understand the talks of prisoners even whispered?*". Notwithstanding the limitations of conventional parameters with respect to the answer to the above mentioned question, the authors deemed that a computerized simulation of the transmission of speech from the ground level of the grotto to the window of tyrant Dionysius could reveal further acoustic features of the Ear of Dionysius. The analysis was carried out in terms of STI (Speech Transmission Index)[5] and D50% (Definition)[6]. The calibrated model of the Ear of Dionysius was used with the consideration of an array of 90 point sources at the height of 1.6 m above the floor of the cave and a single receiver at the supposed location of the spying tyrant up in the top window at the end of the cave. Figure 3 shows a plan and a section illustrating the locations of the sound sources and receiving point.

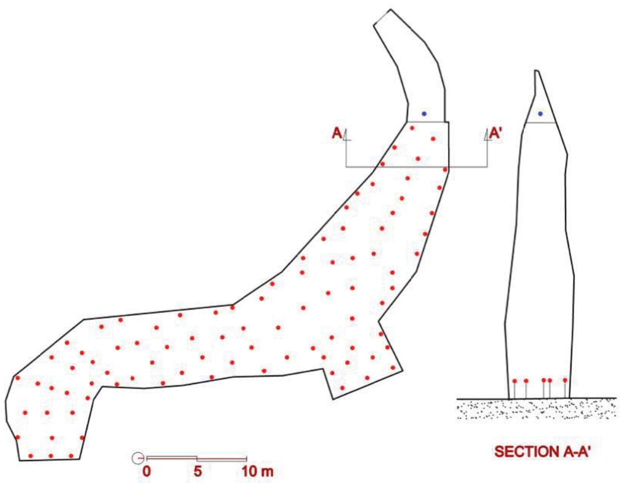

Figure 3: Plan of the Ear of Dionysus showing the location of 90 point sources at ground level (red dots) and a vertical section A-A' with the receiving point (blue dot).

A run with the transition order TO = 2 and 200.000 rays yielded impulse responses to calculate the considered intelligibility parameters. The values of STI were obtained by assigning a rather low background noise level at the receiving point for each octave band. The over-all sound level summed up at about 30 dBA which seems appropriate for the grotto in a quiet rural area. The calculated octave band values of D50% were not corrected for the background noise. This choice corresponds to the condition of maximum speech intelligibility at the tyrant spot because in this instance the received signal depends only on the room acoustic transmission from the ground level of the grotto to the spot of the tyrant.

Results

The mean value of calculated STI is 0.43 with a standard deviation of 0.07. Given the large variation of STI caused by the sound source location, it was deemed worth to report the results in a map of values of STI (at the receiving point) associated with the sound source location at ground level responsible of the mapped STI. This representation highlights the degree of goodness of each transmission channel (Figure 4).

Among the simple parameters describing speech intelligibility in rooms D50% has been used because it was conceived and studied since the fifties [7]. Originally its average value in the frequency range from 340 to 3500 Hz was correlated with the percent intelligibility of "syllables" (phonetically balanced logatomes).

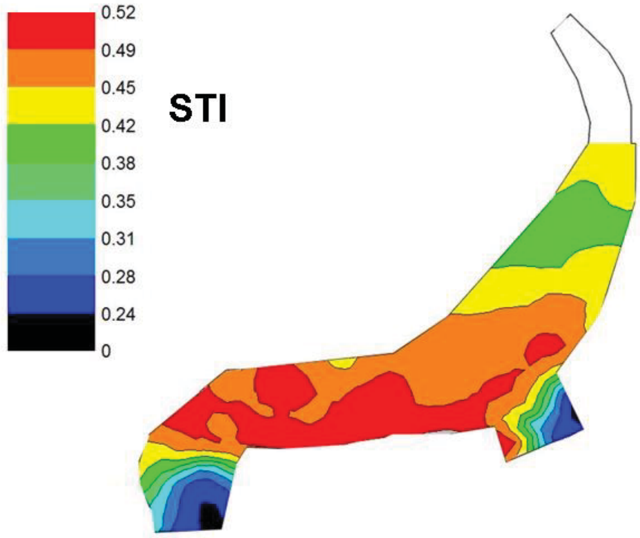

Figure 4: Map of STI calculated at the receiving point (tyrant spot) associated with the sound source location at ground level in the cave. Best speech transmission corresponds to red areas.

The Figure 5 displays the space-averaged 1- octave-band values of D50% and the relevant interval ± 1 standard deviation calculated at the tyrant spot with the same distribution of sound sources on the ground of the cave.

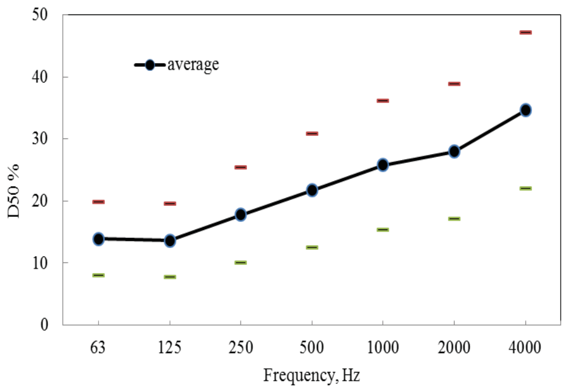

Figure 5: 1-octave-band mean values of D50% and relevant two standard deviations interval. Average calculated for 90 source receiver pairs.

The calculated D50% shows a marked dependence on the sound source location especially in the frequency range of speech. Therefore a map similar to the one reported in Figure 4 was implemented also for D50% averaged in the octave bands from 500 Hz to 4 kHz (Figure 6)..

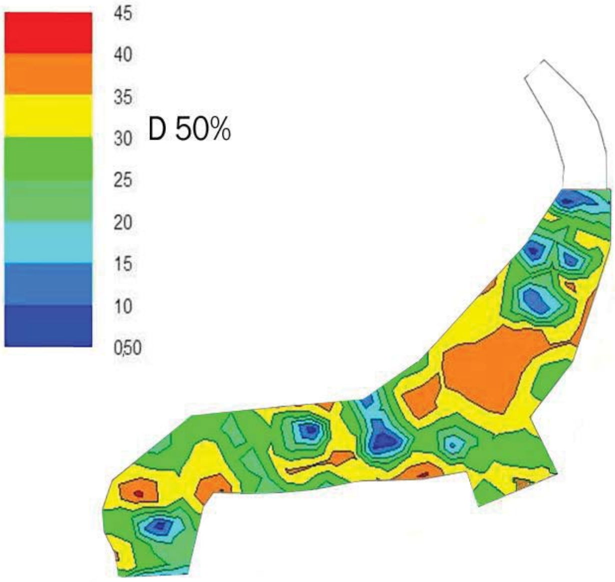

Figure 6: Map of averaged D50% (500-4k Hz) calculated at the receiving point (tyrant spot) associated with the sound source location at ground level in the cave. Best speech transmission corresponds to red areas.

Discussion

The inspection of maps in Figure 4 and Figure 6 highlights a strong dependence of the speech intelligibility parameters on the location of the speaker at ground level. The best locations for speech transmission to the window of tyrant do not occur when the speaker stands at minimal unscreened distances from the receiver where the level of the direct sound attains higher values. This outcome suggests the presence of acoustic paths that enhance useful early sound with respect to the detrimental sound in the impulse response. These paths produce also a lesser degradation of the modulation of the voiced speech. It is worth to recall that the calculations of STI were carried out with the assumption of relatively low values of the background noise level (about 30 dBA) while those for the octave band averaged D50% were performed without any influence of the background noise. These ideal conditions involve only natural room-acoustics. This instance yields the highest values of STI and D50% for a given room. Although in these optimal conditions, it is useful to get a judgement about the predictable speech intelligibility. Figure 7 displays the statistical distribution of the 90 values of calculated STI organized to fit the classification for native speakers reported in IEC 60268-16. No data falls in the bins “good” or “excellent”. For about half of data the quality of speech intelligibility is “fair” and for the remaining data it is “poor” or “bad”. The maximum predictable intelligibility of syllables is less than 67%. If non-native speakers, people with speech disorders or hard-of-hearing are involved, the probability is even lesser.

The 90 values of D50% averaged in the octave bands from 500 to 4k Hz calculated with Odeon were evaluated with reference to the correlation between the percent speech intelligibility of syllables and the measured average D50% reported in ref [7, p.224]. The histogram in Figure 8 shows the classification of the 90 values of D50% in intervals having a constant span of 5%. The adjunct numbers below the scale of the Definition D50% mark the limits of classes in terms of corresponding percent intelligibility of syllables as obtained from the above mentioned correlation curve which is not linear.

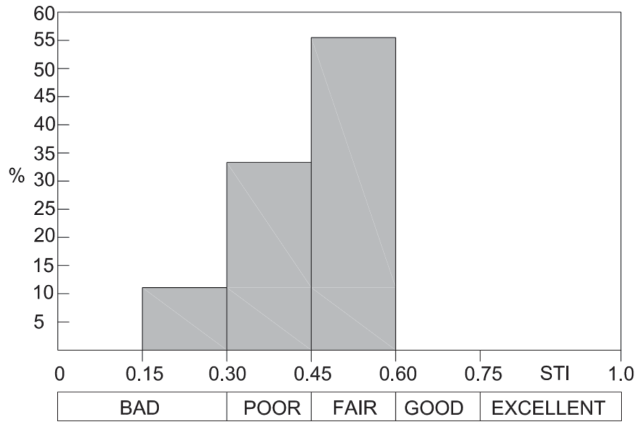

Figure 7: Statistical distribution of STI (90 calculated values) and quality of speech intelligibility.

.

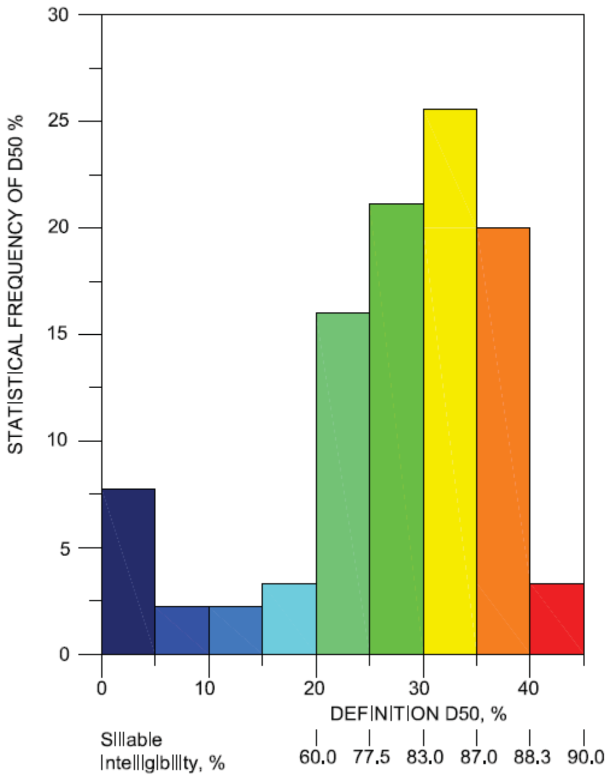

Figure 8: Statistical distribution of percent values of D50% averaged in octave bands from 500 to 4k Hz (90 calculated values). The lower numbers refer to the percent syllables intelligibility obtained from the correlation in ref [7, p.224] for the limits of classes of D50% where applicable.

Usually, D50% > 50 % is considered as the range of "good" and "excellent" intelligibility of speech. No calculated D50% was found beyond this boundary value. One can observe that the most frequent percent speech intelligibility of syllables is a little more than 25% in the bin 30% < D50% < 35% This means that the transmission channels with the sound source located in the yellow areas depicted in the map can be characterized with the percent speech intelligibility of syllables in the range (83%-87/%). An analogous meaning can be given to the other classified data. It is worth noting the presence of "dead spots" (blue areas) dispersed along the longitude of the cave which are characterized by very poor speech intelligibility. The values of the above reported metrics, obtained in ideal objective conditions, suggest a low probability that the tyrant Dionysius could understand the talks of prisoners issued with normal voice levels. Beside his skills about the knowledge and familiarity with spoken languages and other unknown facts, it can be presumed that unavoidable background noises may have caused further degradation of the speech intelligibility at the listening spot. The origins of these noises may have been natural, e,g. wind, rain, thunders and other external causes. The prisoners themselves may have been an internal source of noise even when sleeping with their breathing, snoring coughing and body motion. As regards the intelligibility of whispered speech several papers have been published related to this topic, especially in the area of the recognition of speech elements. However, only few information is useful to formulate a kind of answer to the question posed previously. Various authors, e.g. [8]. report that whispered phonation is produced by a turbulent air flow through partially open vocal folds which in turn produces a noise-like excitation of the vocal tract resonances (the formants) which affects principally the emission of vocals. In spite of time and spectral changes in the signal of whispered syllables with respect to the natural voiced speech, several cues are conveyed to the listener allowing a sufficient degree of intelligibility of the whispered speech in a low-level background noise. At least two aspects are important for the present discussion:

1) whispered speech is typically 15–20 dB softer than voiced speech;
2) whispered speech has a spectral tilt of approximately + 6 dB/octave with respect to normal voiced speech.

The first aspect influences the audibility. On the base of the previous discussion, unavoidable noises can increase easily the degradation of the perception and the intelligibility of whispered speech. As regards the second aspect, it can be noted that Figure 5 shows that calculated D50% presents an average slope of about + 2.5 %/octave. A high slope in the frequency range from 500 Hz to 4 kHz is also evident in the single measurement reported in Figure 9 in ref [2]. On one side this may favour the intelligibility of voiced speech because the increment of useful energy in the higher octave bands moderates the masking effect of voiced vocals on consonants. On the other side the perception of unvoiced vocals may be lost or degraded when the intelligibility of whispered speech is considered.

Conclusion

By accepting some limitations related to the computed evaluations of STI and D50% and subsequent assumptions about the real situation at the times of tyrant of Syracuse, the results of this study represent a further step toward the clarification of the myth of Dionysius. Better results could stem from direct intelligibility tests carried out with suitable human speakers and listeners. The authors are planning this future research. However, it is their opinion that this might serve the sake of scientific truth but they are also conscious that myths, legends and tales must survive as such because they are a mirror of the culture of the past.

References

[1] Sabine W. C. Collected Papers on Acoustics . Dover Publications Inc. New York, 1964, 273-276.

[2] Iannace G., Marletta L., Sicurella F., Ianniello E.: Acoustic measurements in the ear of Dionysius at Syracuse (Italy). Proceedings of Inter Noise 2010, Lisbon (Portugal) 13 – 16 June 2010 paper 580.

[3] Christensen C. L. ODEON Room Acoustics Software version 11 manual. URL: `www.odeon.dk`

[4]. Bradley J.S : Predictors of speech intelligibility in rooms, J. Acoust. Soc. Am. **80** (3), September 1986 pp.837-846,

[5] IEC 60268-16 (2003) Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index, Third edition.

[6] ISO 3382-1 (2009) Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces.

[7] Kuttruff H. Room Acoustics, Fifth Edition. Spon Press London and New York, 2009, 224.

[8] Vestergaard M. D., Patterson R. D.: Effects of voicing in the recognition of concurrent syllables (L), J. Acoust. Soc. Am. **126** (6), December 2009 pp. 2860-2863.