DOI: 10.61782/fa.2025.0177

BINAURAL SPEECH INTELLIGIBILITY IN PRACTICE

Nicola Prodi[1*] **Michele Begotti**[2]
[1] Dipartimento di Ingegneria, Università di Ferrara, Ferrara, Italy
[2] ESG Group, RCF Spa, Reggio Emilia, Italy

ABSTRACT

Speech intelligibility is crucial in many room acoustics applications and its control requires robust and accessible means of qualification and prediction. Most of the current practice is based on speech intelligibility index (STI) which has gained widespread use due to its simple underlying concepts, a rich collection of experts' advice and an acceptable precision in many applications. Such a powerful general purpose tool has of course limitations which are well depicted in the technical norm IEC60268-16. In particular, being essentially a monaural indicator, STI is not fit to mimic the binaural performance of the hearing apparatus. In this work alternative modelling schemes natively based on binaural listening will be briefly recalled and their points of merit compared to STI will be outlined. Later, a case study will be presented where discrepancies between such models and STI can be appreciated from a practical point of view. In particular, the prediction of binaural speech intelligibility in applications involving sound systems will be considered and the benefits of a binaural approach will be discussed.

Keywords: speech intelligibility, binaural, sound system, amplification

1. INTRODUCTION

Most of the common practice related to the assessment of speech intelligibility (SI) in various contexts of room acoustics (public spaces, PA systems, classrooms, conference rooms etc..) is based on the measure of the well-known speech transmission index (STI) [1]. This indicator, introduced over 50 years ago, is essentially a monaural measure of modulation reduction as caused primarily by reverberation and noise. STI is especially fit for conditions where masking is both temporally and spatially homogeneous. Leaving temporal issues apart, it happens that seldom noise sources are localized at a single place or there are several maskers at different positions. Under such circumstances STI is not able to warrant entirely reliable results. To overcome this limit, variants have been proposed, called binaural STI or BSTI henceforth, where binaural processing has been considered [2], but they have not reached widespread use. Indeed binaural processing is crucial for inhomogeneous/concentrated distributions of maskers: under such circumstances experience tells us that the SI that can be achieved binaurally is always higher than the monaural one. Audiologically inspired models for binaural SI often include a processor whose work is twofold; first selects the better ear signal-to-noise (BE-SNR) and secondly implements the E-C mechanism [3] whereby the binaural masking level differences are translated into SNR improvements (Binaural unmasking – BU). A useful binaural SI family of models based on the previous concepts are readily available [4] and are systematically presented in [5] but they also are not become a practical tool yet. It is believed that some more knowledge has to be accumulated in various forms as for instance experimental data and comparisons; such information would help researchers and practitioners to familiarize with alternative means of SI modelling. In this view the present work has the practical aim of providing a basic estimates of the discrepancies that are encountered when assessing SI in spatially critical conditions if STI, BSTI or a native binaural model are used. The work is based on acoustical measures in a set of loudspeakers combinations where a fixed target source is paired with single or multiple localized interfering sources. By systematic measuring of monaural and binaural impulse responses the SI – related indicators are computed derived from both STI and an alternative model. No

*Corresponding author: nicola.prodi@unife.it

listening tests are pursued whereas the comparison between the indicators is achieved solely by adapting the SNR.

2. MATERIALS AND METHODS

A demo room where multiple sound systems are installed was used to collect a set of acoustical measures as shown in Fig. 1. The room is located in the RCF Spa premises and its reverberation time is short (approximately 0.30 s). Several sound sources (S01 – S07) close to the room's walls faced four listening positions (R01 – R04) which were arranged in the central area. By doing so, multiple directions of arrival of the masker were tested while the target source was always delivered from the S03 loudspeaker. Two receivers (R01 and R03) were on the line connecting S03 and S06 while other two (R02 and R04) were displaced laterally on the left facing the target source by 2.50 m.

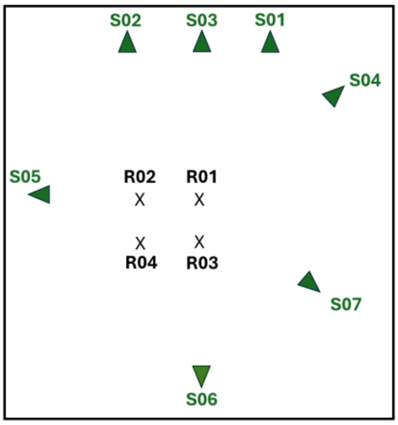

Figure 1. Layout of the positions of the seven sound sources (S01 – S07) and of the four receivers (R01-R04). S03 is fixed as the target source.

Fig. 2 shows pictures of the location of the sources and some of the receiver's positions. A binaural HATS (B&K4100) and a monaural microphone (B&K4189) were used as receivers and were moved in turn across the listening positions. The ESS technique was used to obtain impulse responses for each source-receiver combination. A level calibration was accomplished on the IRs in order to have all of them sharing the same energy left-right average value: this was done to compensate for the different source-receiver distances that could make the results not directly comparable. Secondly, monaural and binaural IRs were filtered to match the long term spectrum of speech. This ensured that both target and masker had the same spectrum and any spectral mismatch was ruled out.

Once calibrated and equalized IRs were available, calculations of STI were accomplished from monaural IRs and also BSTI was retrieved from binaural ones. In particular a simplified "best ear" STI was used. Then a native binaural model [6], which is suitable for the tested reverberated conditions, was selected and used according to the specifications provided in [5]. Differently from STI, this model does not provide an "absolute" measure that can be directly translated into SI percentage once the psychometric curve of the given speech material is provided. Rather, the model outputs are essentially dB estimates of the improvements gained with respect to a monaural evaluation based on SNR. In particular, the model provides two estimates: the BE-SNR which is the selection for the best ear, and the BU which is the estimate of the unmasking due to the genuine binaural interaction. The two dB values are then added algebraically and the final estimate is given. To achieve an SI percentage value a specific back-end should be attached to the model, which was not the case for the model implementation used here. So, an issue of comparability between STI / BSTI and the Leclere et al. model [6] had to be addressed and was tackled as follows.

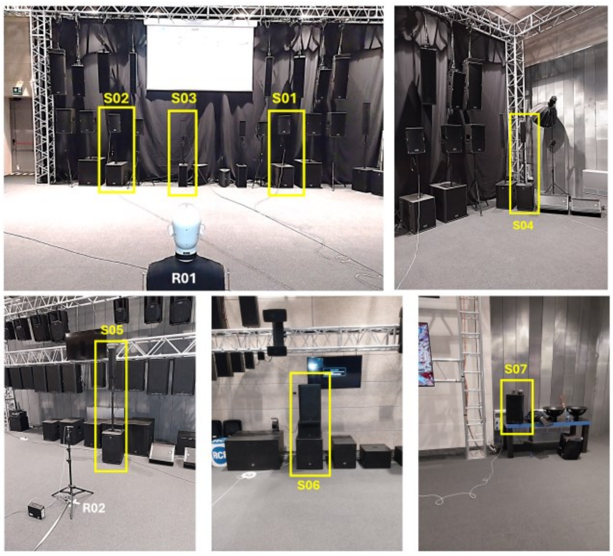

Figure 2. Pictures of the demo room with installed target and masker loudspeakers and binaural (R01) and monaural (R02) receivers.

The calculation of BSTI was considered as a basis because it already included the effect of the better ear SNR. The masking due to reverberation was rather limited in the room so an equivalent BSTI ($BSTI_{eq}$) was obtained by 1) adding the BU in dB to the SNR of the best ear STI and then 2) recalculating the best ear STI. Clearly, this process has limits but in the present conditions was deemed as a simple and viable way of comparing quantities that otherwise could not be jointly assessed.

3. RESULTS

3.1 Better-ear and binaural unmasking

Fig. 3 reports the measured values of the BE-SNR from the Leclere et al. model while Fig. 4 shows the respective values of the BU expressed as binaural masking level differences (BMLD). Consistently with expectations, the BE-SNR values encompass large variations especially for more asymmetric configurations but surprisingly in some cases the discrepancies are only within 1 dB (R03). Yet a precise prediction of left-right sound level gaps is complex and it also depends on the specific directional performances of the loudspeakers which was not controlled.
On the other hand data in Fig. 4 are more regular and directly explained by geometry. Although the measured gaps are at most slightly larger than only 2 dBs the trend is highly consistent, because higher values are obtained for S04, S05 and S07 that are the spatially uneven sound sources.

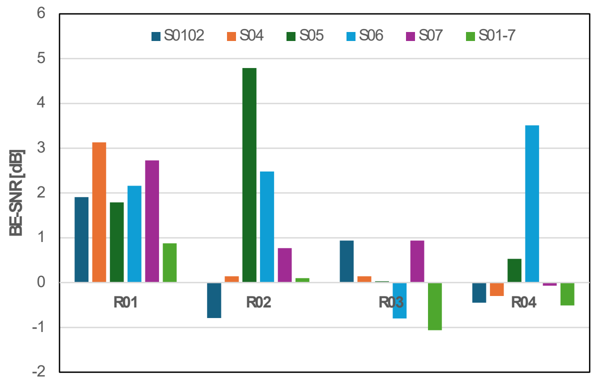

Figure 3. Values for the better-ear SNR evaluation in the various combinations of sources and receivers tested. Target source is S03 hence no improvement is achieved for the co-located masker in S03 and is thus not reported.

To appreciate the findings and the implications they may have in practical applications it has also to be recalled that, depending on the slope of the psychometric curve, a bias as small as 1dB can turn into an SI gap of even 11% as for instance it occurs for the Italian matrix test [7] in the region of 50% accuracy. Should the target accuracy be raised to 80% or more then the gap would reduce significantly due to the shape of the psychometric curve, whose slope flattens with the increase of the accuracy.

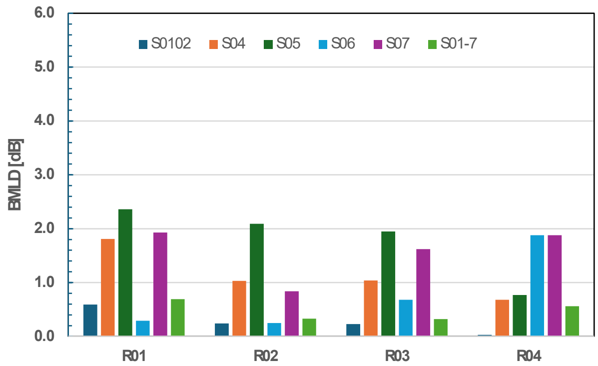

Figure 4. Values for the binaural masking level difference (BMLD) calculated in the various combinations of sources and receivers tested.

3.2 Monaural and binaural STI and BSTIeq

Then a direct comparison of STI, BSTI with the calculated BTI_{eq} was pursued and is shown in Fig. 5. The plot reports differences between the quantities and also the conventional JND of STI set to 0.03 is included for interpretation of the results. Data refer to position averages where the contribution of the symmetric maskers (S01S02, S06 and S03) was excluded. In fact such maskers have a minimal impact on the E-C processing and hence do not provide a binaural release from masking. Very large discrepancies between STI and $BSTI_{eq}$ are shown and they were surely expected due to the lack of any binaural information in the STI calculations, but also the gap between BSTI and $BSTI_{eq}$ is always equal or larger than the JND. This means that the improvement brought about by binaural listening is relevant and can be decisive for specific applications. To complete the analysis also a more direct estimate in terms of SI was developed based on the psychometric curves provided in the assessment of STI, with the hypothesis that they would be applicable also for the quantities used in the present assessment.

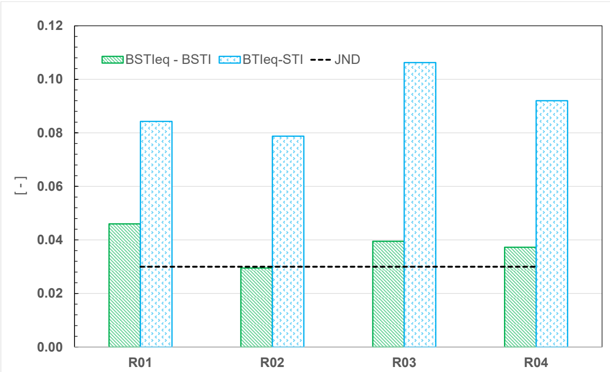

Figure 5. Differences of the average values of STI, BTI and $BSTI_{eq}$ per position. Masker positions S03 and S0102 are not included.

In Fig. 6 one can see that, for instance, if the "sentences" psychometric curve is considered and a critical value of 0.50 is achieved with BTIeq then the previous gaps of Fig. 5 can be turned into specific SI negative biases for BSTI and STI. In particular, in grand average conditions the biases amount to nearly 30% with respect to STI and to 12% compared to BSTI.

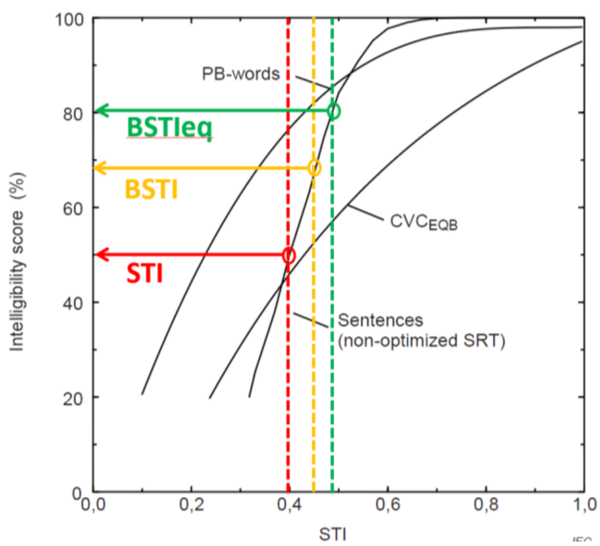

Figure 6. Estimates of the SI gaps derived by the different metrics for typical types of speech materials according to psychometric curves. Sentences are targeted in the graph. The target STI value is close to 0.50 and it is seen how monaural STI and BSTI would provide an underestimate of SI.

The $BSTI_{eq}$ estimates are always favorable due to the exploiting of binaural information which is not fully processed in the other cases. From this perspective one may also argue that STI estimates are precautionary when technical assessment inside difficult environments are to be developed. This argument is surely robust, but it disregards the extra-efforts that may be necessary to reach a certain satisfactory level of SI while the same SI level would be achievable by relying more on the natural binaural processing capacities.

4. REMARKS

The conditions employed to develop the comparisons were deliberately chosen as being critical for the usage of STI. On the other hand it is not uncommon that this indicator, due to its widespread usage, is reported without an accurate preliminary evaluation on the spatial distribution of interferes so that the present evaluations work as worst case scenarios for a misusage of STI in those cases. On the other hand to resolve the limit it appears that manipulation of the STI with some binaural advancement would be hardly sufficient. The present data are in line with previous literature and in addition they try to bridge the gap between the current room acoustics practice and an advanced and more effective binaural modelling of SI. In conclusion the results add experimental evidence to the different performance of the different models under challenging conditions: it is hoped that such approach will be considered in room acoustics practice. Clearly, within this set of measurements there cannot be any hint to the contribution of informational masking which, on the contrary, is a paramount source of interference when SI is assessed in a more ecological experimental layout.

5. REFERENCES

[1] IEC 60268-16:2020 Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index

[2] S.J. van Wijngaarden, R. Drullman: "Binaural intelligibility prediction based on the speech transmission index," *J Acoust Soc Am.* 2008; 123(6), 4514-23, 2008.

[3] M. Lavandier, V. Best, "Modelling speech understanding in complex situations." In J. Blauert (Ed.), *The Technology of Binaural Understanding* (Cap. 19). Berlin, Heidelberg: 2019. Springer.

[4] P. Majdak, C. Hollomey, R. Baumgartner: "AMT 1.5: A toolbox for reproducible research in auditory modeling," *Acta Acustica,* 6(19), 1-17, 2022.

[5] M. Lavandier, T. Vicente, L. Prud'Homme: "A series of SNR-based speech intelligibility models in the Auditory Modeling Toolbox," *Acta Acustica*, 6(20), 1-17, 2022.

[6] S.J. van Leclère T., Lavandier M., Culling J.F. (2015), Speech intelligibility prediction in reverberation: Towards an integrated model of speech transmission, spatial unmasking, and binaural de-reverberation. J. Acoust. Soc. Am. 137(6), 3335-45. Wijngaarden and R. Drullman. 'Development of a binaural speech transmission

[7] G.E. Puglisi, A. Warzybok, S. Hochmuth, C. Visentin, A. Astolfi, N. Prodi, B. Kollmeier: "An Italian matrix sentence test for the evaluation of speech intelligibility in noise," *Int J Audiol*., 54 Suppl 2, 44-50, 2015.