Reception Threshold and Psychometric Slope Measurements for Speech masked by Multiple Noise Sources in Virtual Reverberant Rooms

Juan-Pablo Ramirez, Alexander Raake
Quality and Usability Lab, Deutsche Telekom Laboratories, TU Berlin, E-Mail: Juan-Pablo.Ramirez@telekom.de

Introduction

The intelligibility of speech in the presence of noise is conditioned by the ability of the listener to segregate the target signal from the distracting sources. Recognition scores are shown to improve significantly when the sources are spatially distributed for two known reasons [1]. Firstly, spatialization induces differences in the signal-to-noise ratio across the ears, allowing the listener to focus on the better side of the head. Secondly, head shadowing introduces interaural level differences (ILD) for frequencies typically above 1.3 kHz while, while for frequencies below, the distance between the ears induces phase differences that lead to interaural time differences (ITD).

Binaural intelligibility is typically measured through the spatial distribution of a target speech and a single masking source azimuth a listener that provides recognition scores. Several models were successfully proposed in the recent past [2] [3] [4], but they were corroborated on experimental set-up where the target speech is located in the front only. The present study proposes a set of intelligibility measurements where the position of both target and masking sources varies azimuth around the listener. Results are displayed and compared to predictions according to [4]. Intelligibility measurements performed in virtual reverberant room are proposed as well.

SRT measurements in anechoic room

The first test consisted measured the speech reception threshold (SRT) in anechoic room for an extensive set of spatial distribution. The SRT is the target-to-masker energetic ratio that yields to an intelligibility of 50%. The corpus was composed of semantically unpredictable sentences of 4 keywords in German. The masking source was speech-shaped noise. The speech and noise were spatially separated by convolution with head related transfer function (HRTF) and presented through headphones. 32 students of the TU Berlin presenting no hearing impairment took part to the test. Results are presented in table 1. The spatial release from masking (SRfM) in a given distribution is the decrease in SRT relative to the case where target and masker are in the front. Three conditions were not assessed because they can be deduced from symmetrical distribution.

The SRfM measured range from -1.30 dB to 13.56 dB. For speech and noise in the front, an SRT of -4.1dB was measured in a previous study [5]. The standard deviation along subjects is comprised between 0.6 dB and 3.5 dB. It is on average equal to 1.5 dB and generally increases with the SRfM. The following part proposes to predict these thresholds.

Table 1: SRfM in dB for speech and noise located resp. at x and y azimuth, positive angles on the right of the listener.

	0°	60°	90°	120°	180°
0°	0	9.86	7.26	9.36	2.29
60°	5.21	-0.97	0.49	1.17	9.00
90°	5.27	2.00	-1.01	3.38	8.58
120°	3.85	-0.84	0.18	0.83	7.07
180°	-1.30	7.64	6.52	8.43	-0.81
-120°	5.96	12.22	9.95	12.23	NaN
-90°	6.48	13.08	10.60	12.61	NaN
-60°	5.99	12.15	11.53	13.56	NaN

Predictions of binaural hearing

The Speech Intelligibility Index [6] enables the prediction of speech intelligibility in diotic listening. In dichotic presentation, it can be adapted by selecting between the ears the highest apparent signal-to-noise ratio in each 21 critical frequency band [3] [7]. For target and masker respectively located at angles of x and y, the level of speech is adjusted to have the "better ear per band" SII_{xy} equal to the diotic SII. This adjustment predicts the SRfM resulting from listening at the better hear.

The contribution of binaural hearing is calculated by the method exposed in [4] and summed to the benefit of listening at the better ear to provide prediction of the SRfM. These are compared with the data of table 1 on Fig. 1.

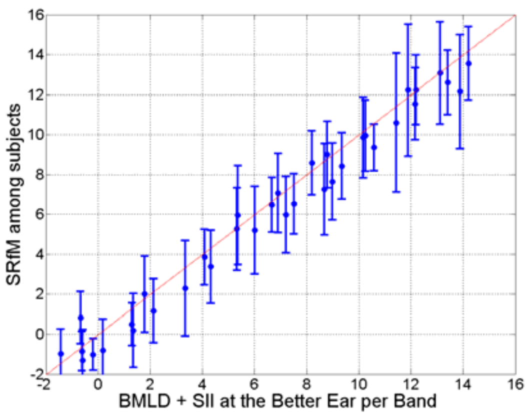

Figure 1: SRfM and standard deviation across subjects against predictions from SII at the "better ear per band" summed to Binaural Masking Level Differences.

Predictions are close to observations are close with a Pearson correlation factor and a root-mean-squared error of respectively 0.99 and 0.84.

Spatial hearing in reverberant room

A second test was performed with target and masker respectively located at 0°, 60° and 120° azimuth and convolved with reverberation times of 0, 0.4 and 1.5 seconds. For the anechoic condition, the target-to-masker ratio (*TMR*) was set 3 dB below and above the SRT given in table 1. For reverberant rooms, the noise level was set 3 dB lower. The aim of such set-up was to obtain recognition scores neighboring 25% and 75 % and to obtain the *SRT* and the slope *s* of the psychometric function depicting intelligibility (*Intel*) given in equation (1) [8].

$$Intel = \frac{1}{1+e^{4s\cdot(SRT-TMR)}} \quad (1)$$

Results of the test are shown in Fig. 2. For the two conditions displayed on the top right of the figure, the TMR was set to too low levels and thus were no considered in the following discussion.

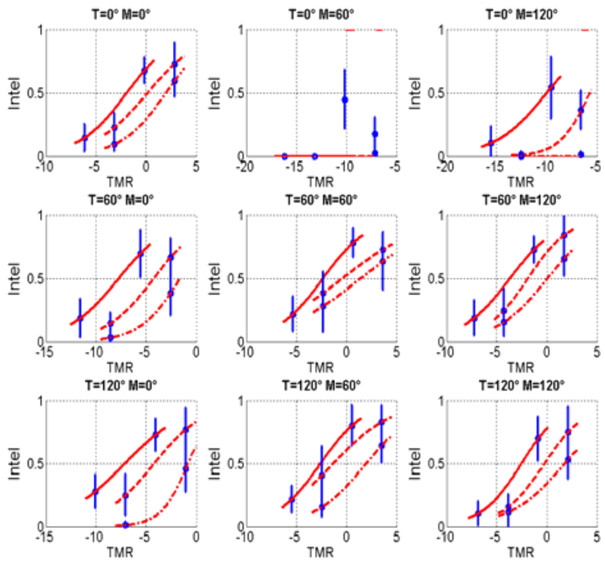

Figure 2: Recognition scores with standard deviation across subjects. Target T and masker M are located azimuth around the listener. Each line corresponds to a reverberation time resp. from left to right of 0, 0.4 and 1.5 seconds.

On average across subjects and conditions, reverberation times of 0.4 and 1.5 seconds increase the SRT of respectively 2.11 dB and 4.53 dB. *s* is believed to be an index of the complexity of the task performed by subjects. For most of the conditions considered in this test, reverberation does not significantly correlate with the variations of *s* which is equal on average to 11 $\%.dB^{-1}$. However, this observation does not apply to the configuration where both target and masker are located at 60°, as *s* decreases from 10.6 $\%.dB^{-1}$ to 6.1 $\%.dB^{-1}$ with reverberation. It is not excluded that higher reverberation time and other spatial distributions may lead to variations in the slope of the psychometric function of intelligibility.

The question of the prediction of intelligibility in reverberant rooms arises at this point. The scope of application of the SII being limited to anechoic conditions, the present contribution will end on proposing to adapt the method of the "better ear per band" SII described in the previous part to an equivalent "better ear per band" STI.

Conclusion

Speech intelligibility is enhanced by the spatial distribution of the competing sources, a phenomenon known as the spatial release from masking (SRfM) and attributed to the so-called binaural hearing. Several models propose to predict subjective recognition scores for a target speech and competing sources located in the horizontal plan around the listener [2] [3] [4]. The accuracy of these predictions is questioned in the present paper by the fact that all experimental conditions confronted to these models had the target speech located in the front of the listener.

In a first experiment, the present study proposes measurements of the speech reception threshold (SRT) [8] for an exhaustive set of distributions of a target speech and competing noise. Observations are successfully predicted by the binaural masking level differences [4] added to a "better ear per band" SII. In a second experiment, the SRfM is measured in reverberant rooms along with the slope of the psychometric function of intelligibility.

References

[1] Bronkhorst, A. W.: The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. Acustica, 86 (2000), 833-843

[2] Beutelmann, R. and Brand, T.: Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am., 120 (2006), 331–342

[3] Wijngaarden, S. J. and Drullman, R.: Binaural intelligibility prediction based on the speech transmission index. J. Acoust. Soc. Am., 123 (2008), 4514-4523

[4] Lavandier, M. and Culling, J. F.: Prediction of binaural intelligibility against noise in rooms. J. Acoust. Soc. Am., in press (2010)

[5] Ramirez, J-P., Raake, A. and Reusch, D.: Intelligibility assessment method for semantically unpredictable sentences in German. Proceedings of the NAG/DAGA conference (2009), 1153-1156

[6] Methods for calculation of the speech intelligibility index. ANSI Report No. S3.5-1997, American Standards Institute, New York (1997).

[7] Edmonds, B. A. and Culling, J. F.: The spatial unmasking of speech: Evidence for better-ear listening. J. Acoust. Soc. Am., 120 (2006), 1540-1545

[8] Brand, T., and Kollmeier, B.: Efficient adaptive procedure for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. J. Acoust. Soc. Am., 111 (2002), 2801-2810

[9] Steeneken, H. J. M. and Houtgast, T.: A physical method for measuring speech transmission quality. J. Acoust. Soc. Am., 67 (1980), 318–326