Intelligibility of spatially reproduced speech over headphones under ambient noise

Noam R. Shabtai[1], Jonathan Sheaffer[1], Zamir Ben-Hur[1],
Itai Nehoran[2], Matan Ben-Asher[2], and Boaz Rafaely[1]

[1] *Acoustics Lab, Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva 8410501, Israel*
[2] *Waves Audio Ltd., Azrieli Center 3, The Triangle Tower, 32nd Floor, Tel-Aviv 6701101, Israel*

Introduction

In natural environments, the intelligibility of a speech signal from a target source is degraded by noise signals that may be generated from additional acoustic sources. However, if the target and the noise signals arrive from different directions, spatial hearing mechanisms are triggered in the human auditory system and the intelligibility of the target signal is improved. This effect is referred to as the *cocktail party effect*. The improvement of the intelligibility due to the separation of the target and noise signal in the space domain is measured using the difference of the *speech reception threshold* (SRT) and referred to as the *spatial release from masking* (SRM) [1]. The SRM was studied using loudspeakers that generate signals from different locations [2, 3], and using *binaural sound reproduction* (BSR) via headphones, either by directly filtering the signals with the *head related transfer functions* (HRTFs) [4], or in a more general manner in the spherical harmonics domain [5, 6]. In the later case, the spherical harmonics order up to which the *plane-wave amplitude density function* and the HRTFs where decomposed was investigated for its effect on the SRM .

Since BSR was shown to result in SRM, beamforming and BSR were generalized into a unified form, known as the *generalized sound-reproduction beamformers* (GSB) [8], where the target signal is enhanced over the noise signals and, at the same time, all the sources are perceived as arriving their actual directions, by that improving the intelligibility of the target signal as well as improving the realism of the acoustic scene. In this framework, the order of the amplitude density function, the spatial selectivity weight functions, and the HRTFs can be tuned and form a trade-off between spatial selectivity and BSR [9]. A later representation is used to optimize the weight function of the GSB under the maximum directivity criterion, where it is assumed that the incorporation of the HRTFs has an effect on the output directivity of the beamformer [10].

As opposed to systems in which both target and noise signals are generated in the surrounding environment of the listener or binaurally reproduced and presented to the listener via headphones as in Fig. 1, a hybrid configuration in which the target signal is binaurally reproduced and the noise signal is generated using external sources, as shown in Fig. 2 hasn't been yet investigated for its effect on SRM. In this work, an adaptive hearing in noise test (HINT) is designed to study speech perception with this hybrid configuration.

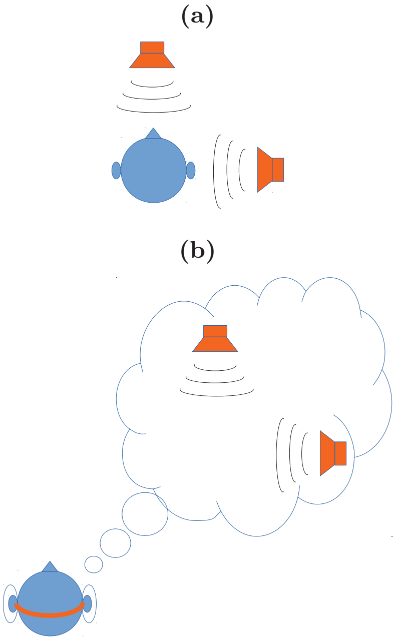

Figure 1: SRM using (a) external sources or (b) BSR.

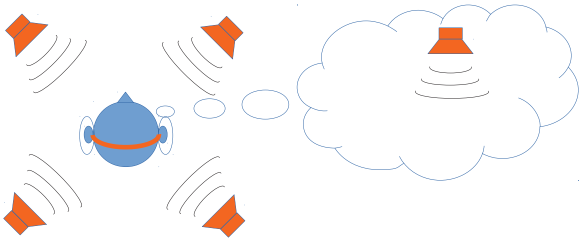

Figure 2: Hybrid configuration: target is binaurally reproduced and the noise is external.

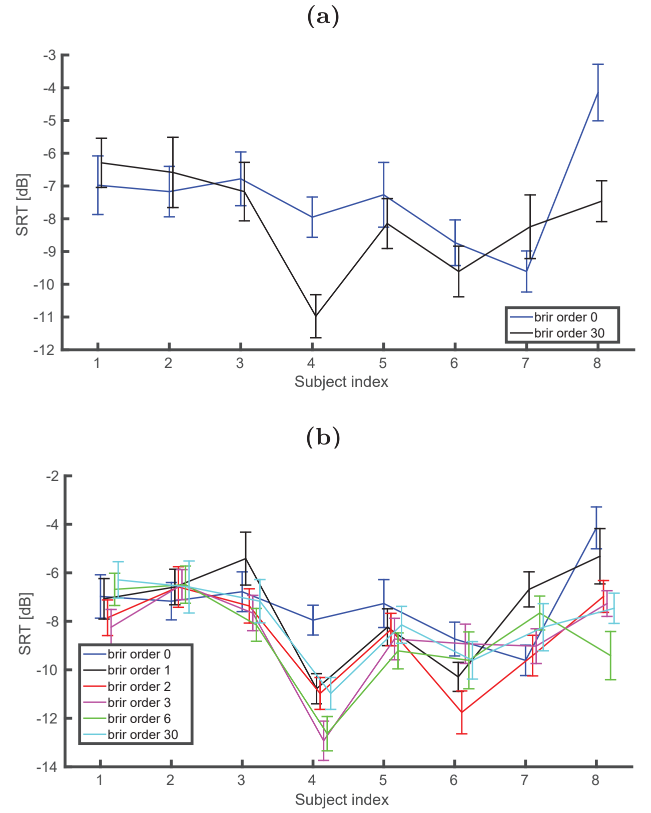

Figure 4: SRT results in (a) Binaural vs. Mono and (b) different spherical harmonics order configuration.

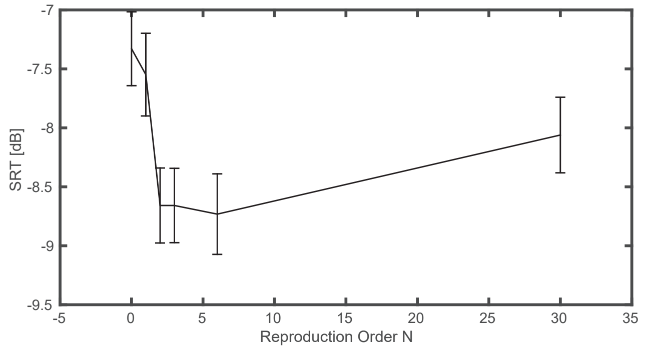

Figure 5: SRT calculated by averaging last 40 SNR values of all subjects.

Experimental Results

A HINT has been performed in the control room of the Acoustics Lab at the Ben Gurion University. The control room has the dimensions of 2.9 × 3.1 × 3.2m^3. A target signal was binaurally reproduced and displayed to the listener via headphones, under the presence of externally surrounding noise that was generated using loudspeakers, as had already been displayed in Fig. 2. As shown in Fig. 3, the target signal contains triplets of digits from 0 to 9, and 4 Loudspeakers generate speech babble noise at the level of 70 dBA at the listener position. Eight Subjects participated in the test and each one of them was presented with 50 digit triplets. The *signal to noise ratio* (SNR) level started from -30 dB and, in an adaptive manner, increased by 2 dB for each wrong response of the listener, and decreased by 2 dB for each correct response.

For each subject, the SRT was calculated by averaging over the SNR values at which the last 40 triplets were displayed. The BSR is performed using 6 different configurations, each corresponds to a different order in the spherical harmonics domain up to which the plane-wave amplitude density function and the HRTFs are decomposed, for $N = 30, 6, 4, 3, 2, 1, 0$. The *Sound Scape Renderer* (SSR), developed by TU Berlin was used along with a SparkFun 9DOF Razor AHRS IMU head-tracker. The order of $N = 30$ is assumed high enough to represent the reference case of a fully reproduced binaural signal, whereas the order of $N = 0$ represents a mono target signal.

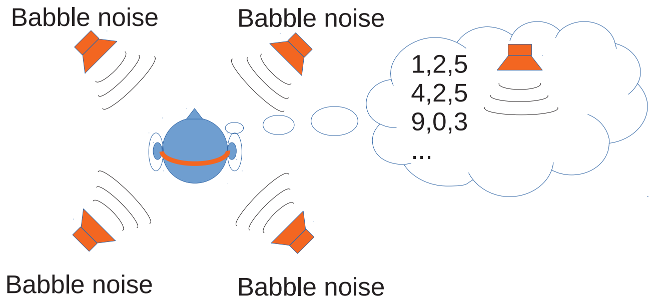

Figure 3: Hearing in noise test using the hybrid system.

Figure 4 displays the SRT results in the different configurations for each subject, first comparing the reference binaural system to the mono system in Fig. , and then displaying the results for all orders in Fig. . The vertical bars in Fig. 4 display the 95% confidence interval. From Fig. 4 it may be seen that, for most subjects, there is no significant difference between the different configurations in terms of intelligibility. Averaging the last 40 SNR values over all subjects (Fig. 5) shows that any significant difference in the intelligibility is in the range of the adaptive test SNR level increase-decrease step, i.e., 2dB.

Conclusion

A hybrid version of an adaptive HINT has been performed, where the target signal is binaurally reproduced via headphones and an ambient speech babble noise was externally generated using loudspeakers. Using the current setup and calibration method of the HINT, no significant difference was shown in the intelligibility of digit triplets utterance, when either binaurally reproduced using the SSR system, or displayed as a mono signal. The reason is probably that a SRM is already obtained with the mono signal, since it is perceived inside the head and at a distinctively different spatial position compared to all the noise sources. Future work may investigate the SRM effect when a single direction noise signal is presented, and the target signal is either mono or at a different direction that may increase the SRM. Other future experiments may include the binaural reproduction for the ambient noise signal, calibration methods that consider the attenuation of external signals by the headphones, and using other binaural reproduction tools.

References

[1] R. Y. Litovsky. Spatial release from masking. *Acoustics Today*, 8(2):18–25, Apr. 2012.

[2] N. Marrone, C. R. Mason, and G. Kidd Jr. Tuning in the spatial dimension: Evidence from a masked speech identification task. *J. Acoust. Soc. Am.*, 124(2):1146–1158, Aug. 2008.

[3] G. L. Jones and R. Y. Litovsky. A cocktail party model of spatial release from masking by both noise and speech interferers. *J. Acoust. Soc. Am.*, 130(3):1463–1474, Sep. 2011.

[4] A. Ihlefeld and B. Shinn-Cunningham. Spatial release from energetic and informational masking in a selective speech identification task. *J. Acoust. Soc. Am.*, 123(6):4380–4392, Jun. 2008.

[5] R. Duraiswami, D. Zotkin, Z. Li, E. Grassi, N. Gumerov, and L. Davis. High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues. In *119th Convention of the AES*, pages 1–16, New York, NY, USA, 2005.

[6] B. Rafaely and A. Avni. Interaural cross correlation in a sound field represented by spherical harmonics. *J. Acoust. Soc. Am.*, 172(2):823–828, Feb. 2010.

[7] A. Avni, H. Wierstorf, M. Geier, J. Ahrens, S. Spors, and B. Rafaely. Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution. *J. Acoust. Soc. Am.*, 133(5):2711–2721, Jun 2013.

[8] N. R. Shabtai and B. Rafaely. Generalized spherical array beamforming for binaural speech reproduction. *IEEE Trans. Audio, Speech, Lang. Process.*, 22(1):238–247, Jan. 2014.

[9] M. Jeffet, N. R. Shabtai, and B. Rafaely. Theory and perceptual evaluation of the binaural reproduction and beamforming trade-off in the generalized spherical array beamformer. *IEEE Trans. Audio, Speech, Lang. Process.*, 24(4):708–718, Apr. 2016.

[10] N. R. Shabtai. Optimization of the directivity in binaural-sound-reproduction beamforming. *J. Acoust. Soc. Am.*, 138(5):3118–3128, Nov. 2015.