DOI: 10.61782/fa.2023.0257

MONAURAL AND BINAURAL PERCEPTION OF SPEECH IN COMPLEX SCENES: RECENT RESULTS ON THE ROLE OF CORRELATION

Nicola Prodi[1*] **Chiara Visentin**[1] **Matteo Pellegatti**[1]
[1] Department of Engineering, University of Ferrara, Italy

ABSTRACT

The perception of speech and its intelligibility have long been studied in relation to monaural factors (i.e. reverberation and clarity). On the other hand, especially in more complex settings, binaural hearing is necessary to achieve performance. The deployment of effective binaural processes is strongly influenced by the layout of source, masker(s) and receivers, by reverberation and also by the type of reflections in the impulse response. Scattered reflections are well-known to provide a spatially spread wavefront, a time-distributed response and uncorrelated binaural signals compared to specular ones. This work discusses the various effects mentioned above and presents some recent results on the role of correlation in this context.

Keywords: *speech perception, scattering, correlation, binaural hearing.*

1. INTRODUCTION

Intelligibility has strong foundations in monaural signal analysis. A single impulse response provides information on both reverberation and the temporal distribution of energy [1,2]. On the latter aspect, 50 ms has been indicated for decades as the useful time limit for the integration of the first reflections (see standard [3]). In recent years a systematic investigation and verification of how single reflections are actually aggregated to the direct sound has begun. In the work [4] it was shown that, for a single reflection, the delay and direction are largely independent, even if frontal reflections are more effective for spectral reasons [5] and it was found that reflections much later than 50 ms can be integrated as long as they are in phase with the direct sound [6]. Finally, the effective time limit seems to depend on the room [7]. Simple, essentially monaural, models such as C50, U50 and STI do not capture this complexity as they provide energetic predictions that are not specific either by position (a given value of the indicator is always associated with the same intelligibility disregarding location details apart distance from the source) or by listener. Furthermore, even limiting oneself to energetic masking noise (informative masking introduces a further degree of complexity), it is very complicated with the previous models to manage noises fluctuating over time and in the spectrum. On the other hand, it is known that for the phenomena of "listening in the dips (LD)" [8] and "glimpsing (GP)" [9] human hearing works best when it can benefit from short time-spectral windows with favorable signal-to-noise ratios (SNR).

2. BINAURAL EFFECTS

Hearing with two ears is essential in many realistic contexts and always produces better results than a single channel [10]. At the basis of this effectiveness are the mechanisms of comparison between the right and left channels which are well explained (although not entirely) by the so-called E-C equalization-cancellation theory [11] and by the "better ear" effect. Hearing processes occurring in the olivary complex structures extract delay times, levels and estimates of the correlation between signals, using this information for target signal and masker in order to improve SNR. In particular, the system works as if it realigned the masking signal thanks to the estimation of the delay, compensated it

[*] ***Corresponding author***: *nicola.prodi@unife.it*

for amplitude and finally subtracted it from one channel to the other; by so doing the portion of noise which is correlated between the two ears can be eliminated. The effectiveness of the process therefore depends on how well the noise signals are correlated between channels. The "better ear" is instead directly attributable to the shielding effect of the head with respect to lateral noise sources. The manifestation of this ability, which takes place over not-too-short time intervals, is experimentally measured with the spatial release from masking (SRM) quantity [12]. This quantity is the difference in dB between the threshold values for a certain intelligibility measured with co-located source and masker and shifting the masker by a certain angle. Since the threshold with an angled masker is lower (therefore better performance) then SRM is positive (except in rare cases which are not of interest here) and can even reach about ten dB in anechoic conditions for openings of 90° - 100 ° [13].

3. REVERBERATION AND THE RELATED EFFECTS

Reverberation in a room has multiple effects on the signals received. Reflections with an excessive delay corrupt the signal because they cannot be integrated: in practice they become part of the masker, even in the monaural hypothesis alone. Added to this is that the reverberant tail "fills" any gaps available for fluctuating noise and therefore the LD and GP mechanisms also offer reduced support. In binaural hearing, the right-left level differences decrease making the "better ear" less effective; above all, reverberation makes the extraction of the binaural cues on which the E-C mechanism is based more critical. It is therefore not surprising that intelligibility is poorer with inadequate reverberation. In particular, it has been discussed how the correlation of the masking decreases with the increase of reverberation and this makes the E-C process almost ineffective for a completely uncorrelated noise [14, 15]. The precise evaluation of the disadvantage in a reverberated context compared to an anechoic one also in terms of binaural performance can be detected with worse SRM values with the same geometric condition between the target and masking source [16]. It should be noted that the performances are still better than monaural listening in the same context.

4. CORRELATION AND THE TYPE OF REFLECTIONS

The emphasis on the correlation of the masker probably derives from the fact that the frontal target speech at a short distance maintains a sufficient correlation even in not too reverberant conditions. Recently, the roles of the correlation of the target and the masking noise have been studied together. Favorable monaural conditions have been established in terms of reverberation and clarity and the typology of the reflections has been changed [17] to manipulate the correlation of signals. This was obtained by changing three first reflections from specular to diffuse, the latter obtained from direct measurements on a Schroeder diffuser [18]. Working with a diffuse isotropic noise (and therefore with low correlation) it was seen that the condition which ensured greater correlation of the source also provided more intelligibility. In a further study [16] concentrated noise was used, working with the SRM. Two notable results have been obtained: both the correlation of the source and of the masker is important and the former is prevailing over the latter, at least as far as it is associated with a larger correlation value. The best performance is obtained (higher SRM) when both the target and the masker are associated with specular reflections, while, if the reflections are scattered, the performance is worse. This result provides a fairly direct indication at the room acoustical design level pointing to the provision of specular reflections to support speech via correlation.

5. THE SPECIAL CASE OF ANCIENT THEATRES

In ancient theaters such as Greek or Roman theatres the binaural impulse response between the source in the orchestra and a listener in the cavea is characterized, after the direct sound, by the scattered components produced by the steps. Although such type of reflections typically reduces the correlation at the ears, in this case the correlation between the right-left channels is instead still high [19]. This is explained by the centrally symmetrical orientation of the audience towards the source which minimizes the differences between the ears and, above all, by the fact that scattering occurs primarily on the listener's sagittal plane and less on the coronal one. The result is that the two ears receive reflections resulting from scattering, but almost equal to each other on the left and right sides. This allows the listener to trigger the binaural mechanisms seen previously. The correlation of the target source can be a factor which, together with others already known [20], ensures good intelligibility even in these large outdoor environments.

6. REFERENCES

[1] J. S. Bradley, R.D. Reich, and S.G. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility", *Journal of the Acoustical Society of America,* 106(4), 1820-1828, 1999.

[2] J. S. Bradley, H. Sato, M. Picard, *"On the importance of early reflections for speech in rooms"*, *Journal of the Acoustical Society of America*, 113(6), 3233-3244, 2003.

[3] *ISO3382 Acoustics — Measurement of Room Acoustic Parameters — Part 1: Performance spaces*. Geneva, International Organization for Standardization, 2009.

[4] A. Warzybok, J. Rennies, T. Brand, S. Doclo, B. Kollmeier, "Effects of spatial and temporal integration of a single early reflection on speech intelligibility", *Journal of the Acoustical Society of America*, 133(1), 269–282, 2013.

[5] I. Arweiler, J.M. Buchholz, "The influence of spectral characteristics of early reflections on speech intelligibility", *Journal of the Acoustical Society of America*, 130(2), 996-1005. 2011.

[6] J. Rennies, A. Warzybok, T. Brand, T., B. Kollmeier, "Measurement and prediction of binaural-temporal integration of speech reflections", *Trends in Hearing*, 23, 2331216519854267, 2019.

[7] T. Leclere, M. Lavandier, J.F. Culling, "Speech intelligibility prediction in reverberation: Towards an integrated model of speech transmission, spatial unmasking and binaural de-reverberation", *Journal of the Acoustical Society of America*, 137 (6), 3335-3345, 2015.

[8] J. M. Festen, R. Plomp, "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing", *Journal of the Acoustical Society of America*, 88, 1725 – 1736, 1990.

[9] M. Cooke, "A glimpsing model of speech perception in noise", *Journal of the Acoustical Society of America*, 119(3), 1562-1573, 2006.

[10] A.W. Bronkhorst, "The cocktail-party problem revisited: early processing and selection of multi-talker speech", *Attention, Perception, & Psychophysics*, 77, 1465–1487, 2015.

[11] N. I. Durlach, "Binaural signal detection: equalization and cancellation theory", in *Foundations of Modern Auditory Theory vol. II*, Academic, New York, 371-462, 1972.

[12] J. Culling, M. Lavandier, "Binaural unmasking and spatial release from masking". In, *Binaural Hearing,* 209-241, ASA Press Springer Nature Switzerland AG, Cham, Switzerland, 2021.

[13] J. Peissig, B. Kollmeier, "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners", *Journal of the Acoustical Society of America*, 101, 1660–1670, 1997.

[14] M. Lavandier, J. F. Culling, "Speech segregation in rooms: effects of reverberation on both target and interferer", *The Journal of the Acoustical Society of America*, 122 (3), 1713-1723, 2007.

[15] M. Lavandier, J.F. Culling, "Speech segregation in rooms: monaural, binaural, and interacting effects of reverberation on target and interferer", *Journal of the Acoustical Society of America*, 123(4), 2237-2248, 2008.

[16] N. Prodi, M. Pellegatti, C.Visentin, "Comparing the effects of scattered and specular sound reflections on speech intelligibility in rooms", *Building and Environment*, 228, 109881, 2023.

[17] N. Prodi, M. Pellegatti, C.Visentin, "Effects of type of early reflection, clarity of speech, reverberation and diffuse noise on the spatial perception of a speech source and its intelligibility", *Journal of the Acoustical Society of America*, 151(5), 3522-3434, 2022.

[18] C. Visentin, M. Pellegatti, N. Prodi, "Effect of a single lateral diffuse reflection on spatial percepts and speech intelligibility", *Journal of the Acoustical Society of America*, 148(1), 122 – 140, 2020.

[19] S. Sato, S. Sakai, N. Prodi, "Acoustical measurements in ancient Greek and Roman theatres", in *Proc. of Forum Acusticum 2002*, (Siville, Spain), 2002,

[20] S. Giron, A. Álvarez-Corbacho, T. Zamarreno, "Exploring the Acoustics of Ancient Open-Air Theatres", *Archives of Acoustics*, 45(2), 181–208, 2020.