

Understanding communication sounds processing in adverse acoustic conditions: Psychoacoustical and neurophysiological findings

Souffi Samira¹, Lorenzi Christian², Huetz Chloé¹ and Edeline Jean-Marc^{1*}

¹Department of Cognition and Behaviour, Paris-Saclay Institute of Neurosciences (Neuro-PSI), CNRS UMR 9197, Université Paris-Sud, Bâtiment 446, 91405 Orsay cedex, France

²Department of Cognitive Studies, Laboratory of Perceptive Systems, UMR CNRS 8248, Ecole Normale Supérieure (ENS), Paris Sciences & Lettres University, Paris, France

Abstract

Psychoacoustic studies revealed that speech intelligibility can be preserved in conditions of severe acoustic degradations such as those induced by the presence of masking noise or by the processing of speech by a vocoder which removes the temporal fine structure and preserves partially the temporal envelope. At the neuronal level, many studies have pointed out that auditory cortex neurons display robust discrimination, but it seems crucial to investigate how subcortical auditory neurons respond in adverse conditions.

Introduction

Our ears are constantly bombarded by a complex sound mixture, which often generates challenging acoustic conditions for understanding speech. These acoustic conditions can be the presence of background noise, with the particular case of the “cocktail party” noise» (requiring segregating a sound source from a mixture of different sources), the internal room acoustics which can potentially cause reverberation phenomena, but also the environmental conditions for example the attenuations of some frequencies by the environment [1-3] and spectro-temporal masking by other competing sounds [4]. In human, partial eliminations of acoustic features crucial for speech perception such as the temporal fine structure (TFS) or the slow temporal envelope (E) are also adverse acoustic situations [5-8]. As a consequence, they make it increasingly difficult perceiving target sounds such as speech, communication sounds and music in normal-hearing subjects. In addition, these acoustic conditions impair speech understanding for subjects with mild to moderate hearing loss and are very penalizing for subjects with cochlear implants (neuroprosthetic which restores hearing in people suffering from profound deafness).

Understanding what are the spectro-temporal acoustic cues used by human subjects in adverse conditions and the neuronal mechanisms allowing the auditory system to extract relevant cues for discriminating sounds in these acoustic conditions are major aims in psychoacoustic and auditory neuroscience.

Psychoacoustical studies: impact of acoustic degradations on speech intelligibility

A large number of studies have used vocoders [9] that is signal-processing devices designed to selectively alter specific acoustic features in speech signal [5-7,10,11]. Vocoders decompose incoming

sounds into frequency bands mimicking the spectral decomposition performed by the cochlea. For each band, the temporal-fine-structure component (i.e., a frequency-modulated signal) is degraded (replaced either by a pure tone or by a broadband noise) and then amplitude modulated by the corresponding temporal-envelope component (i.e., an amplitude modulator, AM). The resulting AM carriers are finally added up and presented for discrimination or identification to human subjects. This literature has largely documented that slow AM fluctuations (<16 Hz) in a limited number of frequency bands (4-8) are sufficient to maintain an almost perfect identification of speech in quiet [6,7]. Despite the fact that vocoding reduced the spectral content and the harmonic structure of speech (leading to the loss of pitch and timbre information), slow temporal cues (16Hz) are sufficient to produce 90% correct recognition of consonants, vowels and words [6].

On the other hand, elderly persons often experience important difficulties in understanding speech in adverse listening situations, sometimes in the absence of elevated audiometric thresholds [12,13]. Potentially, this can result from an alteration of supra-threshold auditory processing [14-17], which can be explained if high-threshold auditory-nerve fibers are the first to be impacted during

*Correspondence to: Edeline Jean-Marc, Department Cognition and Behaviour, Paris-Saclay Institute of Neurosciences (Neuro-PSI), CNRS UMR 9197, Université Paris-Sud, Bâtiment 446, 91405 Orsay cedex, France, Tel: (+33) 1 69 15 49 72; Fax: (+33) 1 69 15 77 26; E-mail: jean-marc.edeline@u-psud.fr

Key words: vocoder, masking noise, speech intelligibility, neuronal responses, auditory system

Received: November 06, 2018; **Accepted:** November 23, 2018; **Published:** November 27, 2018

aging [18]. Initially, it was found that older listeners showed deficits in using complex amplitude modulation (AM) patterns to identify speech accurately [19,20]. This is possibly due to reduced sensitivity to AM cues, or reduced central ability to make optimal use of AM cues [12,21]. However, several studies did not find any effect of aging on AM sensitivity [13,22,23], suggesting that the processing of AM information is roughly comparable for younger and older listeners exhibiting similar audiometric thresholds.

When elderly subjects exhibit sensorineural hearing loss, AM sensitivity is generally improved compared to normal [24], presumably because of the loss in the fast-acting amplitude compression applied by outer hair cells in the cochlea [25]. Thus, the potentially detrimental effects of aging on AM processing may be confounded with (and counter-balanced by) the loss of compression associated with mild hearing loss.

Neurophysiological studies: impact of acoustic degradations on neuronal responses in the auditory pathway

The auditory system successfully maintains a detailed and precise neuronal representation of target acoustic stimuli such as speech, communication sounds and music, in the presence of important acoustic alterations. Many studies describing consequences of acoustic degradations have been performed in the primary auditory cortex (AI) with stimuli that differ in length, spectral content, and other acoustic parameters. All the studies using vocoded vocalizations showed little changes in AI responses: in several species, cortical responses were not drastically reduced even when the number of frequency bands was reduced down to two bands [26-29]. At the level of the secondary auditory cortex (SRAF area), Carruthers et al. [30] showed that neuronal populations code invariant representations of conspecific vocalizations despite important spectro-temporal degradations. In the only study performed at the subcortical level with vocoded stimuli, it was reported that, in terms of firing rate, the responses of IC neurons were resistant to drastic spectral degradations [31].

Acoustically, the vocoded stimuli become spectrally more similar as the number of frequency bands decreases, whereas their temporal envelopes remain quite different but are partially degraded. The auditory neurons remain thus sensitive to temporal envelope fluctuations still present in the vocoded stimuli. If this interpretation is correct, masking the amplitude modulations of natural stimuli by noise addition, should largely reduce the neuronal discriminative abilities of auditory neurons. In fact, this might be the best adverse condition to evaluate the abilities of auditory neurons to discriminate communication sounds.

Most of the studies describing the consequences of background noise on neuronal responses to target stimuli have been also performed at the level of the primary auditory cortex. Initially, Nagarajan et al. [26] reported that white noise addition reduced neuronal responses to communication sounds only at a 0dB SNR. In bird field L (homologous to AI), neuronal responses to song motifs were strongly reduced by three different types of masking noises [4]. However, recent results revealed a more complex picture [32]. The responses of cortical neurons can be classified in four classes named robust, balanced, insensitive or brittle when vocalizations were embedded in a broadband white or a babble noise. However, a given neuron can fall into a class or another depending on the type of noise, demonstrating the existence of contextual effects. In fact, the initial results of Bar-Yosef et al. [33] in the cat primary auditory cortex have already pointed out that some cortical neurons are more sensitive to the noise background than to the

actual communication sounds. In the bird homologous of a secondary auditory area (area NCM), cortical inhibitory microcircuits, which contributes to sparsify the evoked discharges of pyramidal cells, allows the emergence of invariant neural representations of communication sounds in noise conditions [34]. In a quite interesting study, Shetake et al. [35] quantified neuronal discriminative abilities of AI responses to similar speech sounds with and without noise addition and found that the discrimination abilities of cortical cells can closely match the behavioral performance. The discrimination performance of neuronal populations was not affected at a SNR of +12dB, but the performance felt close to the chance level with a SNR of -12dB [35]. This resistance of cortical discrimination is at variance with the strong impact of the noise observed in auditory thalamus. Indeed, a massive reduction in evoked firing rate and temporal reliability of evoked responses was observed in auditory thalamus in the easier noise condition (SNR of +10dB; [36]).

A direct comparison between the consequences of acoustic degradation in different structures is the more straightforward way for dissecting where invariant representations are generated. When measuring how different levels of noise alter neuronal coding in the auditory system, it was found that from the auditory nerve to the IC and to AI, the neural representation of natural sounds became more and more independent of the level of background noise [37]. At the population level, this tolerance to background noise was proposed to result from an adaptation to the noise statistics, which is much more pronounced at the cortical than at the subcortical level [37].

Conclusion

Both in the field of psychoacoustics and auditory neurosciences, the way by which robust representations of communication sounds are generated and allow humans and animals to react rapidly and efficiently in adverse acoustic conditions has become an intense research area. At the present time, most studies confirm that cortical neurons contribute to the invariant representation of speech-like stimuli despite very severe acoustic degradations. Almost no study has been performed so far at the subcortical level but clarifying processing mechanisms at the early stages of the auditory pathway may be crucial for understanding how these invariant features and representations emerged within the auditory system. Electrophysiological explorations combining state-of-the-art signal processing analyses and large-scale neuronal recordings at different levels of the auditory system are necessary for progressing in this field.

References

1. Bidelman GM, Davis MK, Pridgen MH (2018) Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hear Res* 367:149-160. [[Crossref](#)]
2. Fuglsang SA, Dau T, Hjortkjer J (2017) Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *Neuroimage* 156: 435-444. [[Crossref](#)]
3. Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. *Science* 343: 1006-1010.
4. Narayan R, Best V, Ozmeral E, McClaine E, Dent M, et al. (2007) Cortical interference effects in the cocktail party problem. *Nat Neurosci*. 10: 1601-1607. [[Crossref](#)]
5. Drullman R (1995) Temporal envelope and fine structure cues for speech intelligibility. *J Acoust Soc Am* 97: 585-592. [[Crossref](#)]
6. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270: 303-304. [[Crossref](#)]
7. Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416: 87-90. [[Crossref](#)]
8. Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BC (2006) Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc Natl Acad Sci USA* 103: 18866-18869. [[Crossref](#)]

9. Dudley H (1939) Remaking speech. *J Acoust Soc Am* 11: 169-177.
10. Shamma S, Lorenzi C (2013) On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. *J Acoust Soc Am* 133: 2818-2833. [[Crossref](#)]
11. Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, et al. (2005) Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci U S A* 102: 2293-2298. [[Crossref](#)]
12. Füllgrabe C, Moore BC, Stone MA (2015) Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front Aging Neurosci* 6: 347. [[Crossref](#)]
13. Schoof T, Rosen S (2014) The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Front Aging Neurosci* 6: 307. [[Crossref](#)]
14. Fitzgibbons PJ, Gordon-Salant S (2010) Age-related differences in discrimination of temporal intervals in accented tone sequences. *Hear Res* 264: 41-47. [[Crossref](#)]
15. Huetz C, Guedin M, Edeline JM (2014) Neural correlates of moderate hearing loss: time course of response changes in the primary auditory cortex of awake guinea-pigs. *Front Syst Neurosci* 8: 65. [[Crossref](#)]
16. Gourévitch B, Edeline JM (2011) Age-related changes in the guinea pig auditory cortex: relationship with peripheral changes and comparison with tone-induced hearing loss. *European Eur J Neurosci* 34: 1953-65. [[Crossref](#)]
17. Gourévitch B, Doisy T, Avillac M, Edeline JM (2009) Follow up of latency and threshold shifts of auditory brainstem responses after single and interrupted acoustic trauma in guinea pig. *Brain Res* 1304: 66-79. [[Crossref](#)]
18. Sergeyenko Y, Lall K, Liberman MC, Kujawa SG (2013) Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. *J Neurosci* 33: 13686-13694.
19. Souza PE, Boike KT (2006) Combining temporal-envelope cues across channels: effects of age and hearing loss. *J Speech Lang Hear Res* 49: 138-149. [[Crossref](#)]
20. Sheldon S, Pichora-Fuller MK, Schneider BA (2008) Effect of age, presentation method, and learning on identification of noise-vocoded words. *J Acoust Soc Am* 123: 476-488. [[Crossref](#)]
21. Takahashi GA, Bacon SP (1992) Modulation detection, modulation masking, and speech understanding in noise in the elderly. *J Speech Hear Res* 35: 1410-1421. [[Crossref](#)]
22. Paraouty N, Lorenzi C (2017) Using individual differences to assess modulation-processing mechanisms and age effects. *Hear Res* 344: 38-49. [[Crossref](#)]
23. Paraouty N, Ewert SD, Wallaert N, Lorenzi C (2016) Interactions between amplitude modulation and frequency modulation processing: Effects of age and hearing loss. *J Acoust Soc Am* 140: 121. [[Crossref](#)]
24. Wallaert N, Moore BC, Ewert SD, Lorenzi C (2017) Sensorineural hearing loss enhances auditory sensitivity and temporal integration for amplitude modulation. *J Acoust Soc Am* 141: 971. [[Crossref](#)]
25. Moore BC, Hafter ER, Glasberg BR (1996) The probe-signal method and auditory-filter shape: results from normal- and hearing-impaired subjects. *J Acoust Soc Am* 99: 542-552. [[Crossref](#)]
26. Nagarajan SS, Cheung SW, Bedenbaugh P, Beitel Re, Schreiner CE, et al. (2002) Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. *J Neurophysiol* 87: 1723-1737. [[Crossref](#)]
27. Ter-Mikaelian M, Semple MN, Sanes DH (2013) Effects of spectral and temporal disruption on cortical encoding of gerbil vocalizations. *J Neurophysiol* 110: 1190-1204. [[Crossref](#)]
28. Ranasinghe KG, Vrana WA, Matney CJ, Kilgard MP (2012) Neural Mechanisms Supporting Robust Discrimination of Spectrally and Temporally Degraded Speech. *J Assoc Res Otolaryngol* 13: 527-542. [[Crossref](#)]
29. Aushana Y, Souffi S, Edeline JM, Lorenzi C, Huetz C (2018) Robust neuronal discrimination in primary auditory cortex despite degradations of spectro-temporal acoustic details: comparison between guinea pigs with normal hearing and mild age-related hearing loss. *J Assoc Res Otolaryngol* 2018 19: 163-180. [[Crossref](#)]
30. Carruthers IM, Laplagne DA, Jaegle A, Briguglio JJ, Mwilambwe-Tshilobo L, et al. (2015) Emergence of invariant representation of vocalizations in the auditory cortex. *J Neurophysiol* 114: 2726-2740. [[Crossref](#)]
31. Ranasinghe KG, Vrana WA, Matney CJ, Kilgard MP (2013) Increasing diversity of neural responses to speech sounds across the central auditory pathway. *Neuroscience* 252: 80-97. [[Crossref](#)]
32. Ni R, Bender DA, Shanечи AM, Gamble JR, Barbour DLs (2017) Contextual effects of noise on vocalization encoding in primary auditory cortex. *J Neurophysiol* 117: 713-727. [[Crossref](#)]
33. Bar-Yosef O, Nelken I (2007) The effects of background noise on the neural responses to natural sounds in cat primary auditory cortex. *Front Comp Neurosci* 1: 3. [[Crossref](#)]
34. Schneider DM, Woolley SM (2013) Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron* 79: 141-152. [[Crossref](#)]
35. Shetake JA, Wolf JT, Cheung R J, Engineer CT, Ram SK, et al. (2011) Cortical activity patterns predict robust speech discrimination ability in noise. *Eur J Neurosci* 34: 1823-1838. [[Crossref](#)]
36. Martin EM, West MF, Bedenbaugh PH (2004) Masking and scrambling in the auditory thalamus of awake rats by Gaussian and modulated noises. *Proc Natl Acad Sci U S A* 101: 14961-14965. [[Crossref](#)]
37. Rabinowitz NC, Willmore BD, King AJ, Schnupp JW (2013) Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biol* 11: e1001710. [[Crossref](#)]