Current Topics in Acoustical Research

Mini-Review

The consequences of cochlear damages on auditory scene analysis

N. Grimault* and E. Gaudrain

UMR CNRS 5020 “Neurosciences et Systèmes Sensoriels”, Université Claude Bernard Lyon 1, 50 Avenue Tony Garnier 69366 Lyon Cedex 07, France

ABSTRACT

In ecological complex acoustical environments in which several acoustic sources interact, for example several speakers, the input of the auditory system is an acoustical mixture that is the summation of waves from all acoustic sources. Most of the time, it remains possible to split the mixture in auditory streams corresponding to individual acoustic sources. Thus, acoustic mixtures are processed by the auditory system using the mechanisms of Auditory Scene Analysis that have been extensively reviewed by Albert Bregman [Bregman A, *Auditory scene analysis*, MIT Press, 1990]. Several mechanisms underlying auditory scene analysis have already been identified but spectral pitch-cues have been evidenced to be one of the strongest factors of segregation. Cochlear damages are usually related to a loss of cochlear resolution which is responsible for an alteration of the primitive spectral pitch-cues involved in the segregation. The specific difficulties of hearing-impaired listeners in Cocktail-Party situations could be related to this lack of cochlear resolution leading to a deficit in auditory scene analysis mechanisms. Some previous works about the consequences of hearing loss on auditory scene analysis will be reviewed.

KEYWORDS: auditory scene analysis, hearing impairment, segregation, speech

*Corresponding author
nicolas.grimault@olfac.univ-lyon1.fr

INTRODUCTION

In everyday life situations, several sound sources interact to form a complex acoustical sound mixture that must be interpreted by the auditory system. Auditory Scene Analysis (ASA) refers to the ability of the human auditory system to segregate sounds issued from different acoustical sources in different perceptual streams and to amalgamate sounds issued from the same acoustical source in a single perceptual stream. As such, a stream is defined as the perceptual auditory object that corresponds to a single acoustic sound source.

ASA is reputed to make use of perceptual differences between sounds for segregation. The amount of perceptual differences between sounds that can be perceived is then of paramount importance for ASA. Perceptual differences can be reduced either in situations in which the sounds are degraded *per se* (like in radio transmission) or more probably in situations in which the reception of the sounds by the ear is degraded (like for hearing-impaired listeners). The consequences of cochlear damages for ASA are reviewed in the following sections.

Intelligibility of sentences in noise for hearing impaired listeners

The mechanisms of ASA are of particular interest in ecological situations involving several talkers. The ability of Normal Hearing listeners (NH) to perceptually segregate a single voice from other concurrent voices is commonly referred to as the *cocktail party effect* [1] and has been extensively described in the literature. The acoustic cues contributing to the segregation of speech presented against a background of competing steady-state noise, fluctuating noise or speech have been reviewed by [2]. Speech to noise (or speech to speech) ratio, binaural, localisation, and pitch cues appear to be some of the most influent factors of perceptual segregation. The specific effect of a mild-to-severe sensorineural hearing impairment for segregation has also been described in a few papers. Clinical observations as well as psycho-acoustic studies involving Hearing Impaired listeners (HI) consistently evidenced that the ability to segregate a speech source from an interfering noisy or speech-like background is dramatically impaired by a sensorineural hearing loss. Table 1 reviews and provides the main results from the literature evidencing the detrimental effect of hearing impairment for segregation.

In conclusion, the results of these studies all evidenced that the speech reception threshold (STR), related to the intelligibility of sentences in noise, is affected for HI [3,4,5,6,8,9,10]. In particular, HI clearly failed to take advantage of the temporal fluctuations of the maskers [3,8,9]. They have also reduced abilities to take advantage of F0 cues [11], binaural and localization cues [3,5,6,9,10]. This latter effect is not reported in Table 1.

Sensorineural hearing losses are mainly characterized by three deficits all related to the loss of outer hair cells in the cochlea (for a review see [12]). First, the absolute sensitivity of the auditory system is dramatically reduced, which leads to detection thresholds larger than normal. Second, the natural cochlear compression of the auditory system is reduced, which leads to a linear cochlear input/output function. This is responsible for loudness recruitment. Third, the auditory filters, characterizing the frequency selectivity of the auditory system, have been described to be three or four times broader for mild-to-severe hearing losses [13]. As a consequence, the frequency selectivity of the auditory system is reduced. Each of these three deficits might contribute to the reduced ASA performances for HI evidenced in Table 1.

In order to distinguish between the contribution of these deficits to the reduction of ASA performances, Baer and Moore [14,15] and Moore and Glasberg [16] have separately simulated these deficits for groups of NH. All simulations were conducted testing NH with degraded (processed in order to mimic a particular deficit) sentences presented in a concurrent background. First, Moore and Glasberg [16], who simulated only the elevation threshold and the loudness recruitment, evidenced a 11-13 dB detrimental effect of these deficits on the intelligibility of sentences in a background of speech. The reduced abilities of the HI to take advantage of the temporal fluctuations of the maskers (*cf.* Table 1) could then come from elevation threshold and loudness recruitment. Second, Baer and Moore [14,15], who simulated only the enlargement of the auditory filters, evidenced a 3-9 dB detrimental effect of the frequency selectivity deficit on the intelligibility of sentences in a background of noise or speech.

All studies reviewed in Table 1 involved an environmental methodology which consists in measuring the intelligibility of speech in various speech or speech-noise background for NH and HI. This methodology is very valuable to make an observation of the specific difficulties of HI in cocktail party situations. However, this experimental methodology is not adapted to go insight the involved mechanisms.

In order to go insight the neural mechanisms involved in ASA, Bregman [17] suggested distinguishing between the ASA processes that are aimed to segregate simultaneous acoustic events and those aimed to segregate sequential acoustic events. This suggestion has inspired the subsequent psychoacoustical researches about ASA involving HI reviewed in the following sections. All these pilot studies (from 1997 to 2005) hypothesized, as a first assumption, that concurrent voices would mainly differ by their pitches.

Segregation of simultaneous auditory events for hearing impaired listeners

Double-vowel identification task is the most common methodological approach that is used for the study of simultaneous segregation. The results in this task for NH have been extensively reviewed and modelled (see [18] for a review and a suggested model). The results from HI remain

Table 1. Speech Reception Threshold (SRT). Only approximate values extrapolated from the figures and tables are provided

Ref.	Test	Type of masker	Subjects	SRT	Hearing loss effect
[3]	SRT: Masker level fixed at 55 dBA	• Steady state speech noise • Speech	• NH (young) • Mild HI (aged) • NH (young) • Mild HI (aged)	-10.7 dB -5.3 dB -17.6 dB -4.8 dB	5.4 dB 12.8 dB
[4]	SRT: Masker level fixed (from 40 to 85 dBA)	• Steady state speech noise	Presbyacousic HI (aged) • Hearing loss 30-36 dB HL • Hearing loss 54-60 dB HL	 -3.6 at 85 dBA -1.2 at 85 dBA	 2.5 dB 4.9 dB
[5]	SRT: Target sentence level fixed at 70 dB SPL	• 12 talker babble	• NH over 55 years • HI over 55 years	-0.5 dB +3 dB	3.5 dB
[6]	SRT: Masker level fixed (from 75 to 95 dBA depending of the degree of hearing loss)	• Steady state speech noise	• NH from [7] • HI symmetrical (18-45 y.) • HI asymmetrical (21-52 y.) Good ear / Bad ear	-6.4 dB -3.9 dB -3.4 dB / -1.4 dB	 2.5 dB 2 dB
[8]	SRT (male or female voice) Masker level fixed at 80 dBA	• Steady state speech noise • Speech-modulated speech noise • Reverse speech	• NH (16-36 year) • HI (21-77 years) • NH (16-36 year) • HI (21-77 years) • NH (16-36 year) • HI (21-77 years)	-4.7 dB -0.7 dB -9 dB -0.8 dB -11.4 dB -1.1 dB	4 dB 8.2 dB 10.3 dB
[9]	SRT: Masker level fixed (from 75 to 90 dBA depending of the degree of hearing loss and 65 dBA for NH)	• Between 1 to 6 mixed speech-modulated speech noises	• NH (23-39 year) • HI (18-59 years)	1 masker NH -12 dB HI -4.9 dB 6 maskers NH -7.7 dB HI -3.2 dB	7.1 dB 4.5 dB
[10]	SRT: Masker level fixed (NH: 75 dB SPL; HI: 75-95 dB SPL)	• Steady state speech noise	• NH (25-32 year) • HI (18-59 years)	-13.85 dB -11.81 dB	2 dB
[11]	SRT: (50%) Target sentence level fixed at 90 dB SPL with F0=120 Hz Double-vowel Identification (Table 2) F0DL	• Speech(1 voice at various F0s) ΔF0 expressed in semitons.	• NH (49-74 year) • HI (59-77 years)	**Same F0 SRT (50%):** 7.9 dB 12.4 dB **benefit of ΔF0 at SRT:** +/-4 NH +11% HI +7% +/-2 NH +7% HI +7%	4.5 dB

much more marginal. Nevertheless, several studies, reviewed in Table 2, provide valuable data that leads to first conclusions about the consequences of hearing impairments for the segregation of simultaneous vowels.

Despite some variability in the results, all studies reviewed in Table 2 showed a depreciative effect of hearing loss for F0-based segregation of simultaneous vowels. This effect was consistently discussed across studies to be related to a spectro-temporal processing deficit [19,20] or an increased amount of peripheral masking [21] associated to hearing loss. Altogether, the reduced frequency selectivity induced by the hearing impairment.

Table 2. Double-vowel and target-vowel identification performances. Only approximate values extrapolated from the tables and figures are provided

Ref.	Test	ΔF0 in semitones	Subjects	Identification performances in percent at ΔF0 in semitones	Hearing loss effect
[19]	Double-vowel Identification	• 0	• NH • HI	70% 45%	-25%
		• 2	• NH • HI	88.5% 61.5%	-27%
[20]	Double-vowel Identification		High Frequency Amplification		Amp. Effect
		• 0	• HI without amplification • HI with amplification	47% 43%	-4%[1]
		• 2	• HI without amplification • HI with amplification	61% 60%	-1%[1]
[11]	Double-vowel Identification.	• 0	• NH • HI	40% 26%	-13%[2]
	F0DL	• 2	• NH • HI	58% 29%	-29%[2]
	SRT (see Table 1)				
[21]	Target-vowel Identification.	• 0	• NH • HI	80% at 0 dB 50% at 0 dB	-30% at 0 dB
	S/N from -10 to +10 dB	• 2	• NH • HI	92% at 0 dB 82% at 0 dB	-10% at 0 dB
[22]	Double-vowel Identification	• 0	• NH • HI	37% 30%	-7%[1]
		• 0.5	• NH • HI	70% 59%	-11%[1]
		• 6 to 9	• NH • HI	78% 65%	-13%[1]

[1] Effect not significant.

[2] Effect of hearing loss correlated to FDL and SRT at group level but effect of hearing loss not correlated to SRT at individual level.

This interpretation is also consistent with the lack of improvement provided by a high frequency amplification that simulates a hearing aid [20]. However, the implication of these results for true cocktail party situations must be tempered as the relationship between speech reception threshold (SRT) in speech and double-vowels identification was hardly proven [11] and should be confirmed by further studies.

Segregation of sequential auditory events for hearing impaired listeners

The experimental methodology implemented to study the mechanisms underlying sequential segregation (i.e. streaming) was first introduced by van Noorden [23]. This methodology consists in the presentation of a repeated temporal sequence of two auditory events (*A* and *B* in the following). When presented with a temporal sequence *ABA*-*ABA*-… (symbol - indicates a silence between *A* and *A*) the amount of perceptual differences between *A* and *B* conducts either to the perception of a single stream (like a gallop), or to the perception of two independent streams played at different rhythms (respectively *A*-*A*-*A*… and *B*---*B*---*B*…). Based on experimental observations, van Noorden [23] defined two streaming thresholds: The Fission Boundary (FB) is the point below which it is impossible to hear two streams. This threshold is roughly independent of the tempo of the sequence. The Temporal Coherence Boundary (TCB) is the point above which it is impossible to hear a single stream. This latter threshold increases when decreasing the tempo of the sequence. Based on the tempo dependence of these thresholds, Bregman [17] suggested that these two thresholds should be sustained by different neural processes. The acoustical cues between *A* and *B* leading to a segregated percept for NH were extensively studied. In general, it seems that almost any salient perceptive difference can lead to a segregated percept. However, frequency differences, pitch (F_0) differences, differences in spectral passbands and to a lesser extent differences in loudness are generally described as the strongest cues for streaming. Streaming performances for NH have already been theorized [24] and modelled [25,26] and extensively reviewed [27]. On the contrary, only few data about HI have been collected. All studies involving HI, reviewed in Table 3, were specifically dedicated either to test for the channelling theory [24], or to evaluate the potential relationship between sequential segregation and the segregation deficit of hearing impaired listeners in concurrent listening situations (Table 1).

Altogether, all streaming studies involving pure tones [28,29,31] reported high variability and no clear direct correlation between frequency selectivity impairment and streaming performances. Then, the channelling theory is only partially and inconsistently supported by these data. On the contrary, all streaming studies involving complex tones [32,33] or vowels [36] evidenced that reduced spectral cues (either pitch cues and/or passband cues) lead to a specific streaming deficit for HI (real or simulated). These studies that involved stimuli closer to speech than pure tones are in good agreement with the channelling theory. This conclusion must be tempered yet by the fact that all studies involving complex tones along with [35] also evidenced genuine temporal streaming (streaming without spectral cues) which leads to similar performances for NH and HI. This is not predicted by the channelling theory. It is worth noticing that all reviewed streaming studies involving pure tones provide an estimate of the FB and all reviewed streaming studies involving complex tones and vowels provide either an estimate of the TCB or an estimate of a subjective threshold between FB and TCB. As these thresholds could be sustained by different neural mechanisms, this methodological aspect could also contribute to the discrepancy between the results from these studies. In fact, the channelling theory might apply differently to the TCB and to the FB. Finally, Mackersie *et al.* [29] and Mackersie [30] found some correlation between pure tone streaming and speech-in-speech intelligibility. In order to get a true evaluation of the contribution of sequential mechanisms to speech-in-speech intelligibility, this correlation should be confirmed yet with streaming data involving speech-like stimuli (either complex tones or vowels) that might be processed by distinct mechanisms than pure tones.

Table 3. Fission Boundary (FB), Temporal Coherence Boundary (TCB) or subjective threshold (subj.) between FB and TCB. Mains results from the studies are reported in the last column

Ref.	Test	Stimuli	Subjects	Results
Pure Tones streaming				
[28]	Pure tone streaming (FB)	ABA- F_A from 250 to 2000 Hz	NH unilateral HI bilaterally HI	• FB expressed in ERB are constant for NH. • FB not consistent with ERB enlargement for HI (sometimes larger, sometimes normal).
[29]	Pure Tone streaming (FB) SRT (male and female) Auditory filters at 1kHz	ABA- F_A= 1000 Hz	NH mild HI	• Hearing loss is correlated with FB and SRT • Frequency selectivity is not correlated with SRT • FB is not correlated with frequency selectivity
[30]	Pure tone streaming (FB) SRT (1 male and 1 female voices)	ABA- F_A= 1000 Hz Ascending and descending patterns	HI with mild-to-moderate hearing loss	• FB for ascending patterns correlated with recognition of male voice. • FB for descending patterns correlated with recognition of female voice.
[31]	Pure tone streaming (FB) FDL	ABA- F_A from 250 to 2000 Hz	NH and HI	• NH: FB = 0.4 ERB. FDL expressed in ERB is constant below 2 kHz and increase above 2 kHz. • HI: high variability. No correlation FDL / FB.
Complex Tones streaming				
[32]	Filtered complex Tone streaming (subj.)	ABA- Complex tones filtered 1375-1875 Hz. F_0(A) from 88 to 250 Hz	NH and HI	• NH / HI: same if no spectral cues for any groups. • NH better than HI if spectral cues for NH only.
[33]	Filtered complex tone streaming (TCB and subj.) with various phase relationships (cosine, alt and random).	**Passbands:** 1250-2500, 1768-3536 and 2500-5000 Hz • TCB: rhythm task: AB- • Subj. ABA-	NH from [34] HI moderate to severe hearing loss	• Effect of passband and phase on TCB. 1.5 better thresholds than for NH. Same influence of phase but smaller influence of passband. • Effect of phase but no effect of passband. Smaller effect of phase and passband than for NH.
[35]	Temporal streaming (subj.)	ABA- Modulated broadband noises. Fm(A)=100Hz Fm(B)=100-800 Hz	NH HI with mild and flat hearing loss	• The subjective segregation thresholds for HI are similar (or slightly better) to that of NH.
Vowels streaming				
[36]	Streaming with vowels (TCB: order task)	Vowels: a e I O U y Alternating F0 pattern F01=100 Hz F02=100-238 Hz	NH simulated HI [15]	• The simulated HI performed better than NH indicating a higher TCB related to a primitive segregation deficit.

CONCLUSION

Studies reviewed in Table 1 clearly established that sensorineural hearing loss leads to a specific intelligibility deficit for speech-in-speech intelligibility. According to these studies, this deficit can be explained in terms of lower sensitivity, loudness recruitment and reduced frequency selectivity generally associated with hearing impairment. The psychoacoustical studies reviewed in the second part of this work distinguished between simultaneous and sequential segregation mechanisms. Most of them evidenced that reduced frequency selectivity, associated to hearing loss, decreases listeners' performances in a double vowel identification task (Table 2) as well as listeners' performances in streaming tasks involving complex tones or vowels (Table 3). The contribution of sequential segregation mechanisms for speech-in-speech intelligibility is consistently reported across studies by a significant correlation between pure tone streaming and speech reception thresholds. This correlation should be confirmed using speech-like stimuli. On the contrary, the contribution of simultaneous segregation to speech-in-speech intelligibility remains unclear.

REFERENCES

1. Cherry, E. C. 1953, J. Acoust. Soc. Am., 25, 975.
2. Bronkhorst, A. 2000, Acustica, 86, 117.
3. Duquesnoy, A. J. 1983, J. Acoust. Soc. Am., 74, 739.
4. Duquesnoy, A. J., and Plomp, R. 1983, J. Acoust. Soc. Am., 73, 2166.
5. Gelfand, S. A., Ross, L., and Miller, S. 1988, J. Acoust. Soc. Am., 83, 248.
6. Bronkhorst, A., and Plomp, R. 1989, J. Acoust. Soc. Am., 86, 1374.
7. Bronkhosrt, A. W., and Plomp, R. 1988, J. Acoust. Soc. Am., 83, 1508.
8. Festen, J., and Plomp, R. 1990, J. Acoust. Soc. Am., 88, 1725.
9. Bronkhorst, A., and Plomp, R. 1992, J. Acoust. Soc. Am., 92, 3132.
10. Peissig, J., and Kollmeier, B. 1997, J. Acoust. Soc. Am., 101, 1660.
11. Summers, V. and Leek, M. R. 1998, J. Speech Language Hear. Res., 41, 1294.
12. Moore, B. C. 1998, Cochlear hearing loss, Whurr, London.
13. Moore, B. C. 1985, Br. J. Audiol., 19, 189.
14. Baer, T., and Moore, B.C. 1993, J. Acoust. Soc. Am., 94, 1229.
15. Baer, T., and Moore, B. C. 1994, J. Acoust. Soc. Am., 95, 2277.
16. Moore, B. C. J., and Glasberg B. R. 1993, J. Acoust. Soc. Am., 94, 2050.
17. Bregman, A. S. 1990, Auditory Scene Analysis: The Perceptual Organization of sound, The MIT Press, Massachusetts, USA.
18. de Cheveigné, A. 1999, J. Acoust. Soc. Am., 106, 2959.
19. Arehart, K. H., King, C. A., and McLean-Mudgett, K. S. 1997, J. Speech Language Hear. Res., 40, 1434.
20. Arehart, K. H. 1998, J. Acoust. Soc. Am., 104, 1733.
21. Arehart, K. H., Rossi-Katz, J., and Swensson-Prutsman, J. 2005, J. Speech Language Hear. Res., 48, 336.
22. Rossi-Katz J., and Arehart, K. H. 2005, J. Acoust. Soc. Am., 118, 2588.
23. van Noorden, L. P. A. S. 1975, Temporal coherence in the perception of tones sequences, Ph.D. thesis, Eindhoven University of Technology.
24. Hartmann, W. M., and Johnson, D. 1991, Music Perception, 9, 115.
25. Beauvois, M. W., and Meddis, R. 1996, J. Acoust. Soc. Am., 99, 2270.
26. McCabe, S. L., and Denham, M. J. 1997, J. Acoust. Soc. Am., 101, 1611.
27. Moore, B. C., and Gockel, H. 2002, Acta Acustica united with Acustica, 88, 320.
28. Rose, M. M., and Moore, B. C. 1997, J. Acoust. Soc. Am., 102, 1768.
29. Mackersie, C. L., Prida, T. L., and Stiles, D. 2001, J. Speech Language Hear. Res., 44, 19.
30. Mackersie, C. L. 2003, J. Speech Language Hear. Res., 46, 912.
31. Rose, M. M., and Moore, B. C. 2005, Hear. Res., 204, 16.
32. Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. 2001, Br. J. Audiol., 35, 173.

33. Stainsby, T. H., Moore, B. C. J., and Glasberg, B. R. 2004, Hear. Res., 192, 119.
34. Roberts, B., Glasberg, B. R., and Moore, B. C. J. 2002**,** J. Acoust. Soc. Am., 112, 2074.
35. Grimault, N., Bacon, S. P. and Micheyl, C., 2005, Auditory streaming without spectral cues in hearing-impaired subjects, *in* Auditory signal processing: physiology, psychoacoustics, and models, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams and L. Collet. Springer Verlag: New York. pp 212.
36. Gaudrain, E., Grimault, N., Healy, E.W. and Béra, J.-C. (Submitted).