available at www.sciencedirect.com
ScienceDirect
www.elsevier.com/locate/brainres
ELSEVIER
BRAIN RESEARCH

Research Report

# Visual cues release the temporal coherence of auditory objects in auditory scene analysis

*Torsten Rahne*[a,b,*], *Martin Böckmann-Barthel*[b]

[a]*Department of Otolaryngology, Martin-Luther-University Halle-Wittenberg, Germany*
[b]*Department of Experimental Audiology and Medical Physics, Otto-von-Guericke-University Magdeburg, Germany*

ARTICLE INFO

*Article history:*
Accepted 31 August 2009
Available online 9 September 2009

*Keywords:*
Auditory perception
Auditory scene analysis
Cross-modal perception
Audiovisual interaction
Mismatch negativity (MMN)
Ambiguity

ABSTRACT

Auditory scene analysis can arrange alternating tones of high and low pitch in a single, integrated melody, or in two parallel, segregated melodies, depending on the presentation rate and pitch contrast of the tones. We conducted an electrophysiological experiment to determine whether an inherently stable sound organization can be altered by a synchronous presentation of visual cues. To this end, two tones with different frequencies were presented in alternation. Frequency distance was selected as narrow or wide, inducing an inherently stable integrated or segregated organization, respectively. To modulate the integration or segregation organization, visual stimuli were synchronized to either the within-set frequency pattern or with a superimposed intensity pattern. Occasional deviations of the regular frequency pattern were introduced. Elicitation of the mismatch negativity (MMN) component of event-related brain potentials by these deviants indexed the presence of a segregated organization. MMN was elicited by tone sequences with wide frequency distance irrespective of the presence of visual cues. At a narrow frequency distance, however, an MMN was elicited when the visual pattern promoted segregation of the sounds showing a release of the inherently stable integrated organization due to visual stimulation. The results demonstrate cross-modal effects on auditory object perceptual organization even on an inherently stable auditory organization.



## 1. Introduction

In a typical environment many concurrent sound sources contribute to the acoustic signal that enters the ears. As an example, in a cocktail-party situation the voice of a dialog partner must be isolated from the noisy surrounding. Solving this problem is part of the auditory scene analysis. So, different sound sources produce overlapping acoustic signals and must be distinguished. Further, sequential elements with a similar pitch, timbre, spatial location, or loudness must be grouped, even if they do not follow one another directly (Bregman, 1990). The entity of sound elements belonging together is usually termed as an auditory stream.

Auditory scene analysis has been studied behaviorally by using sequences of alternating tones with frequencies A and B, presented in an ABA_ or ABAB design where “A” and “B” denote short tones with different frequency spectra and “_” represents a silent interval (Bregman, 1990; Carlyon, 2004;

* *Corresponding author*. Universitätsklinikum Halle (Saale), Klinik für Hals-, Nasen-, Ohrenheilkunde, Magdeburger Str. 12, D-06112 Halle (Saale). Fax: +49 345 557 1859.
E-mail address: torsten.rahne@medizin.uni-halle.de (T. Rahne).

*Abbreviations:* ANOVA, analysis of variance; ECD, equivalent current dipole; EEG, electroencephalogram; ERP, event-related potential; MMN, mismatch negativity; SOA, stimulus onset asynchrony

0006-8993/$ – see front matter 
doi:10.1016/j.brainres.2009.08.086

Rose and Moore, 1997, 2000; van Noorden, 1975). Three different perceptual domains can be distinguished, depending on the frequency distance and the presentation rate of A and B tones (van Noorden, 1975). For a wide range of presentation rates, tones with a frequency distance below the fission boundary are perceived as only a single stream. On the other hand, at a frequency distance above the temporal coherence boundary, the tones are assigned to two segregated streams. In between these boundaries the percept is ambiguous and can involuntarily switch between one and two streams. However, the absolute values of the fission boundaries and temporal coherence boundaries vary across subjects and depend on factors like attention or task demands (Bregman, 1990; Moore and Gockel, 2002; Rahne et al., 2008).

To what extent this stream segregation acts on an automatic level is still an issue of ongoing research. Preattentive neurophysiological measures like the mismatch negativity component of the EEG (Sussman et al., 1998, 2005) and psychophysically measures (Carlyon et al., 2001; Jones et al., 1999) can reveal how the brain resolves the problem of maintaining separately neural representations of distinct streams and how the focus of attention influences detection of change within such a stream. Other work focused on what parametric factors influence the sound object detection in the absence of attention (Rahne et al., 2007; Ritter et al., 2000; Snyder et al., 2006; Sussman et al., 1999; Sussman, 2005; Sussman and Steinschneider, 2006). The mismatch negativity (MMN) response serves as an appropriate tool to investigate the subject's preattentive organization of sound sequences (Näätänen et al., 1978, 2007; Sussman et al., 1998). This component of event-related potentials (ERPs) is generated within auditory cortices (Alho, 1995; Giard et al., 1990) and reflects the output of a change detection process: Regularities ("standard") within the auditory input are extracted. As a result of a comparison of incoming neural information with the regularity stored in the auditory echoic memory, a violation ("deviant") of such a regularity elicits MMN (Näätänen et al., 2001; Sussman, 2005). Thus, MMN can be used to determine which regularities (individual features or pattern of sounds) are represented in sensory memory at the time the "deviant" occurs.

In order to use MMN to study the preattentive organization of streams, Sussman et al. (1998, 1999) introduced sequences of ABAB tones in which rare "deviants" were introduced. The standards consisted of a repeated three-tone pattern. Occasionally, this pattern was reversed, thus forming a deviant. Since the patterns occurred within the A tone set only, the deviants were only detected when A and B tones were segregated in two separate streams. Therefore, MMN can be used to evaluate the representation of the automatic organization of a set of sounds as integrated or segregated (Sussman et al., 1999; Sussman, 2005). Based on this paradigm, Rahne et al. (2007) added visual cues that were correlated with either one of the streams or an across-stream pattern. MMN to the deviant triplets was observed only if visual cues promoted stream segregation. No MMN was observed when visual cues acted to integrate the tone sequence and also in a control condition without the visual stimuli but identical auditory stimulation. Therefore, instead of causing segregation of an inherently ambiguous tone sequence by directing the subject's attention, implicit cues of two separate streams are enough to produce preattentive segregation.

The perception of coherent objects in the environment often includes information from multiple sensory modalities. Then, cross-modal interactions occur in that way that information from one modality can influence object perception in another modality (Besle et al., 2005; Guttman et al., 2005; King and Calvert, 2001; Remijn et al., 2004; Vroomen et al., 2001). Thus, the perception of objects in the environment can be enhanced by speeding detection (Schröger and Widmann, 1998) and by resolving ambiguities (Watanabe and Shimojo, 2001). Here, we focus on the audio–visual interaction: Auditory perception can be influenced by visual information, such as in the misinterpretation of the place of articulation (the "McGurk effect," (McGurk and MacDonald, 1976)) or in the misperception of auditory spatial location that is shifted by visual input (the "ventriloquist effect," (Bertelson et al., 2000)). Vice versa, visual perception can be influenced by auditory information, such as when the number of light flashes perceived is altered by the number of simultaneous auditory beeps (Shams et al., 2000).

Recently, an attempt was made to modify the auditory perceptual organization not by instruction but by visual cues in correlation with either one stream or an across-stream pattern (Rahne et al. 2007, 2008). An ABAB sequence as introduced by Sussman et al. contained repeated triplets of tones rising in frequency ("standards") within the low-frequency "A" tones. Occasional reverse triplets falling in frequency served as "deviants." The high-frequency "B" tones were chosen from three different frequencies in a random sequence. Additionally, every third tone was enhanced in intensity, thus creating an intensity pattern alternating between the A and B set. To promote the perceptual segregation of the low from the high frequency tones, every "A" tone was synchronized to a visual stimulus (a square of a size corresponding to the pitch of the tone; see Fig. 1). In a second condition, the same tone sequence was used, but now every third tone–the louder tones–were synchronized to another visual stimulus (either a small or large circle, corresponding to the pitch of the tones). The tone sequence was ambiguous and was perceived as one or two streams with similar probability. It was found that this ambiguous auditory input was shifted towards segregated organization by the square stimuli synchronized with one stream, and towards integrated organization by the across-stream circles. Consequently, the same auditory sequence was organized differently depending on the visual input, thus demonstrating cross-modal effects on auditory object formation. In an extension of this promising cross-modal shifting, we aimed to study if the temporal coherence and fission boundaries are sensitive for synchronized visual cues, that is, if inherently non-ambiguous tone sequences can be biased to the opposite type of organization. This is suggested by behavioral measurements using a similar paradigm (Rahne et al., 2008). Therefore, we modified the tone sequence of the previous ERP experiment by increasing the frequency distance of the ABAB sequence above the temporal coherence boundary to obtain a stable segregated perceptual organization. In another condition, we lowered the frequency distance below the fission boundary to obtain a stable integrated perceptual organization.