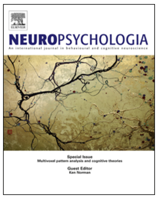
Neuropsychologia 63 (2014) 43–50

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



Rhythmicity and cross-modal temporal cues facilitate detection

Sanne ten Oever [a], Charles E. Schroeder [b,c], David Poeppel [d], Nienke van Atteveldt [a,e], Elana Zion-Golumbic [b,c,f,*]

[a] *Faculty of Psychology and Neuroscience, Maastricht University, 6200 MD, Maastricht, The Netherlands*
[b] *Departments of Psychiatry and Neurology, Columbia University Medical Center, New York, NY 10032, USA*
[c] *The Nathan Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA*
[d] *Department of Psychology, New York University, New York, NY 10003, USA*
[e] *Department of Educational Neuroscience, Faculty of Psychology and Education and Institute Learn, VU University Amsterdam, The Netherlands*
[f] *Gonda Brain Research Center, Bar Ilan University, Ramat Gan, Israel*

ARTICLE INFO

Article history:
Received 20 May 2014
Received in revised form
14 July 2014
Accepted 6 August 2014
Available online 13 August 2014

Keywords:
Temporal prediction
Audiovisual integration
Rhythmicity
Detection

ABSTRACT

Temporal structure in the environment often has predictive value for anticipating the occurrence of forthcoming events. In this study we investigated the influence of two types of predictive temporal information on the perception of near-threshold auditory stimuli: 1) intrinsic temporal rhythmicity within an auditory stimulus stream and 2) temporally-predictive visual cues. We hypothesized that combining predictive temporal information within- and across-modality should decrease the threshold at which sounds are detected, beyond the advantage provided by each information source alone. Two experiments were conducted in which participants had to detect tones in noise. Tones were presented in either rhythmic or random sequences and were preceded by a temporally predictive visual signal in half of the trials. We show that detection intensities are lower for rhythmic (vs. random) and audiovisual (vs. auditory-only) presentation, independent from response bias, and that this effect is even greater for rhythmic audiovisual presentation. These results suggest that both types of temporal information are used to optimally process sounds that occur at expected points in time (resulting in enhanced detection), and that multiple temporal cues are combined to improve temporal estimates. Our findings underscore the flexibility and proactivity of the perceptual system which uses within- and across-modality temporal cues to anticipate upcoming events and process them optimally.

1. Introduction

Increasingly, the brain is thought of as intrinsically proactive, not merely relying on bottom-up sensory information to interpret perceptual information. Instead, even low-level sensory cortices are thought to be constantly creating and updating internal models of the external world, to anticipate and predict upcoming events (Bar, 2011; Friston, 2011; Nobre, Correa, & Coull, 2007; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010; Schubotz, 2007; Summerfield & Egner, 2009; Summerfield et al., 2006). In addition to predicting the content of upcoming stimuli - e.g. features or location – recent research indicates that anticipating the timing of upcoming sounds significantly improves perceptual judgement. Specifically, at least two types of temporal expectations are shown to improve behavioral performance: Rhythmic regularity within a stimulus sequence decreases reaction times and improves accuracies of responses to supra-threshold stimuli when target stimuli occur at an anticipated moment, compared to stimuli occurring randomly or at unanticipated times (Ellis & Jones, 2010; Jones, Moynihan, MacKenzie, & Puente, 2002; Mathewson, Fabiani, Gratton, Beck, & Lleras, 2010; Niemi & Näätänen, 1981), as well as improving stimulus sensitivity (Rohenkohl, Cravo, Wyart, & Nobre, 2012). In addition, temporal cueing within- and across modalities has been used extensively to show that a constant time-interval between a cue and target can improve the speed of target detection (Correa, Lupiáñez, Milliken, & Tudela, 2004; Coull & Nobre, 1998; Lange & Röder, 2006) and recognition (Griffin, Miniussi, & Nobre, 2001) by means of temporal preparation (Los & Van der Burg, 2013). In particular, visual cues appear to be a natural temporal cue for audition (Thorne & Debener, 2008; Van Wassenhove, Grant, & Poeppel, 2005, 2007). A prominent example is speech, since observed lip movements and facial gestures are temporally correlated with, and precede, the auditory input (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; Ten Oever, Sack, Wheat, Bien, & Van Atteveldt, 2013; Van Wassenhove et al., 2005, 2007). Moreover, lip movements and facial gestures have intrinsic rhythmic regularities (Giraud & Poeppel, 2012; Greenberg, Carvey,

---

[*] Corresponding author at: Gonda Brain Research Center, Bar Ilan University, Ramat Gan, Israel.
E-mail address: elana.zion-golumbic@biu.ac.il (E. Zion-Golumbic).

http://dx.doi.org/10.1016/j.neuropsychologia.2014.08.008
0028-3932/© 2014 Elsevier Ltd. All rights reserved.

Hitchcock, & Chang, 2003; Luo, Liu, & Poeppel, 2010; Zion Golumbic, Poeppel, & Schroeder, 2012). Thus, in natural situations, such as speech, we are faced with intermixed temporal information to predict upcoming events, provided by cross-modal as well as rhythmic temporal cues.

The behavioral advantages afforded by these two types of temporal expectations – stimulus rhythmicity and cross-modal temporal cueing – imply that attentional resources can be dynamically allocated to points in time when input is expected (Jones, Johnston, & Puente, 2006; Jones, et al., 2002; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Large & Jones, 1999; Nobre, et al., 2007; Nobre & Coull, 2010). However, it is not clear whether multiple types of cues are used jointly to improve temporal prediction and optimally allocate attention. Since many naturalistic stimuli, such as speech, music and biological motion combine both cross-modal temporal cues and intrinsically rhythmic properties (Zion Golumbic et al., 2012), investigating the joint contribution of temporal cues from these two sources bears substantial ecological relevance.

Here, we used two complementary auditory detection paradigms to investigate the influence of temporal cues on threshold intensities, since the above-described ‘attention in time’ framework predicts that reliable temporal prediction can enhance perceptual sensitivity to subtle stimuli. We manipulated both the temporal structure within the sound stream as well as the presence of cross modal (visual) cues, and investigated the influence of each cue on detection intensities, as well as the combination of both cues. Our hypothesis was that both types of temporal predictions – rhythmicity and cross-modal cueing - would lower sound detection intensities. Rhythmic prediction during the auditory only conditions might not have a strong effect on detection thresholds since, by definition, sounds are “below threshold” before participants indicate that they have heard them. Adding visual input could significantly improve the rhythm percept, thus enriching the temporal prediction. Therefore, we expect an interaction effect in which the combination of cross-modal and rhythmic temporal cues would provide the lowest detection thresholds (Trommershauser, Kording, & Landy, 2011).

2. Materials and Methods

2.1. Participants

Twelve volunteers participated in Experiment 1 (age 20–40; average age: 23.5, 5 male) and twenty volunteers participated in Experiment 2 (age 21–33; average age 25.4, 7 male). All had normal or corrected to normal vision. Informed consent was obtained before the study, which was approved by the New York University Committee on Activities Involving Human Subjects (NYU UCA/HS; Experiment 1) and by the Local Ethical Committee at the Department of Psychology and Neuroscience at the Maastricht University (Experiment 2). Participants were randomly selected and were unaware of the purpose of the study during the experiment. For taking part in the experiment participants received monetary compensation.

2.2. Stimulus material

Auditory stimuli were sinusoidal 1 kHz beeps of 50 ms duration (including a linear rise and fall time of 5 ms) embedded in continuous white noise (53 dB) and presented diotically via headphones (Sennheiser HD 380 Professional, Sennheiser Electronic Corporation, Wedemark, Germany in Experiment 1, Sennheiser HDM25-1 in Experiment 2). The visual stimuli were Gaussian white circles of 50 ms duration (generated using the Gaussian generator of the Visual Stimulus Generation Toolkit implemented in the software Presentation used for stimulus delivery, with parameters: $\text{mu}=-10$ and $\text{sigma}=60$; Neurobehavioral Systems, Inc., Albany, NY), presented foveally on a gray background (rgb: 115,115,115). The visual angle of the Gaussian was 3.1° (corresponding to the width of the 95% contrast interval relative to the center intensity). Both experiments were run in dimly lit sound shielded rooms and participants were seated approximately 57 cm from the screen.

2.3. Experimental procedure

In order to investigate the influence of temporal cues on auditory detection we ran two experiments, using complementary approaches for evaluating detection thresholds.

2.3.1. Experiment 1

In the first experiment we employed the “method of limits” approach to evaluate perceptual thresholds (Gescheider, 1997), using an ‘increasing’ paradigm followed by a ‘decreasing’ paradigm. In the ‘increasing’ paradigm participants heard a stream of auditory beeps embedded in continuous white noise (Fig. 1).

The signal to noise ratio (SNR) of the tone targets was initially below threshold, and the intensity of the tones increased monotonically over the trial. Participants were asked to indicate via button press when the target signals were first detected. In the first four trials, the starting SNR was 0.25% (none of the participants were able to detect the stimulus with this SNR). SNR was defined as the maximal amplitude in the presented sound divided by the maximal amplitude of the white noise. In subsequent trials, the starting intensity was set to be 7.5% SNR lower than the lowest intensity previously-detected, and this level was monitored throughout the experiment to ensure a minimum of 5% SNR difference with the lowest detected intensity judgment. Over the trial, sound intensities increased incrementally in steps of either 0.5 or 1% SNR. The two different incremental steps were randomized to ensure that the sequence of sounds and length of the trials were not identical across trials. After participants indicated detection of auditory stimuli, 4–6 additional beeps were presented at the same intensity level. The ‘decreasing’ paradigm paralleled the ‘increasing’ paradigm, but the sounds started well above detection threshold and decreased in intensity over the trial. Participants had to indicate when they could no longer hear the sounds. Here too, the first four trials were used to determine the individual starting intensities per trial (starting intensity of the first four trials was 17.5% SNR), and ensured that the starting intensity was at least 5% above the highest intensity of the detection judgment.

We manipulated the temporal structure of each trial by changing the interstimulus interval (ISI) between the tones. In half of the trials there was a constant ISI of 666 ms (Rhythmic condition), whereas in the other half the ISI was randomized among one of 21 evenly spaced time points between 300 and 1000 ms, maintaining an average ISI of 666 ms (Random condition). In addition, in half of the trials the Gaussian white circle preceded every auditory stimulus, with a fixed audio-visual stimulus onset asynchrony (SOA) of 65 ms (AudioVisual condition). We choose this interval since it has previously been shown to give optimal cross-modal effects for audiovisual tasks (Thorne & Debener, 2008). Thus, in total there were four conditions: Random Auditory (RaAu), Rhythmic Auditory (RhAu), Random AudioVisual (RaAV), and Rhythmic AudioVisual (RhAV). Designing the paradigm in this way served the purpose of implementing a distinct rhythmic or random temporal structure to a continuous stream of stimuli, which is closer to natural listening conditions. It also mimics natural situations in which visual information is salient, but auditory stimuli vary in intensity over time, for example when listening to a person in a noisy environment. Under all conditions, participants were explicitly instructed to maintain fixation on a gray cross in the middle of the screen when no visual input was presented. Trials were randomized across conditions (20 trials per condition) and the experiment was divided in four blocks of approximately seven minutes each. After every block participants were encouraged to take a break.

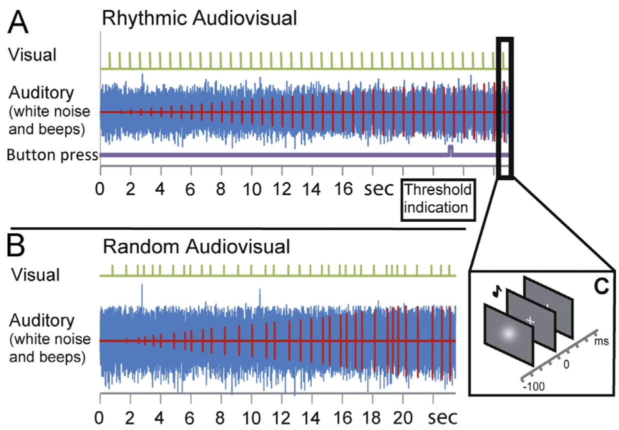


Fig. 1. Illustration of a trial in the rhythmic audiovisual condition (A) and a trial in the random audiovisual condition (B), both in the ‘increasing’ paradigm. In the auditory channel, beeps (red) were embedded in white noise (blue), with their intensity increasing monotonically over the trial. In the audiovisual conditions, a white Gaussian circle was presented 65 ms prior to each beep (C). The button press (purple) indicates the moment that the participant indicates hearing the sound for the first time.