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

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.

© 2014 Elsevier Ltd. All rights reserved.

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,

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: mu=−10 and 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 inter-stimulus 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.
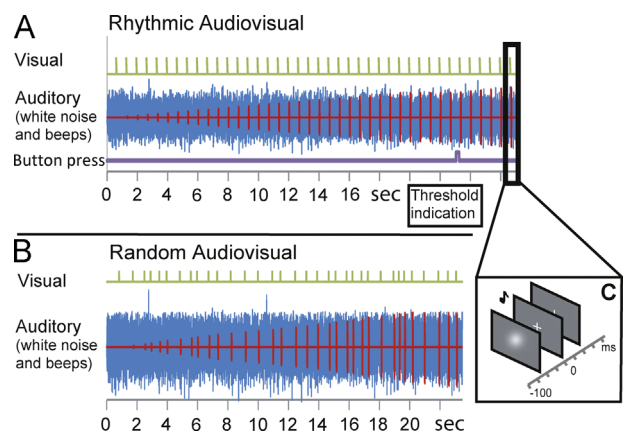

Fig. 1. Illustration of a trial in the rhythmic audiovisual condition (A) and a trial in the random audiovisual condition (B), both in the 'increasing' paradigm. In the auditory channel, beeps (red) were embedded in white noise (blue), with their intensity increasing monotonically over the trial. In 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.

2.3.2. Experiment 2

One drawback of Experiment 1 is that perceptual thresholds calculated using the method of limits approach confounds perceptual sensitivity with response-bias (Green & Swets, 1966). To control for the possibility that the results of Experiment 1 were driven purely by response-bias, we ran a second experiment, using the same four conditions, in which detection thresholds were determined using a staircase procedure in a two alternative forced choice task (2AFC; Green & Swets, 1966). In this task participants were presented sequentially with two 3-second-duration intervals of white noise at 53 dB. In one of the intervals (randomly chosen on every trial) five sounds were embedded in the noise, and the subjects were instructed to indicate in which interval (first or second) they heard the sounds The four conditions were the same as in Experiment 1 (RaAu, RhAu, RaAV, and RhAV). For the two visual conditions, both intervals contained visual stimuli. Since it is difficult to create a temporally-random sequence of stimuli within a finite interval of 3 s, a constant set of 4 ISIs was chosen (350, 500, 814, and 1000 ms) to maximize the temporal variability within each trial. The order of these ISI was randomized in each trial. In addition, under all conditions the onset of the first sound was jittered between 200, 300, 400 and 500 ms after the white noise onset, to reduce expectation effects.

To obtain a measure of the detection threshold we implemented four independent weighted staircase procedures (Kaernbach, 1991) in which the order of the conditions was randomized. In these procedures, every correct response led to a decrease in sound intensity in the next trial of the same condition and every incorrect response led to an increase in sound intensity. Since correct responses can be achieved via 1) actually hearing the stimulus or 2) guessing, the decrease in sound intensity was three times smaller than the increase in sound intensity, which corresponds to a detection threshold of 75% at staircase convergence. Volume increases were approximately 7.5% SNR in the beginning, after the second reversal 3% SNR, and after the fourth and later reversals 0.75% SNR. A reversal was defined as a change from correct to incorrect responses or vice versa for one specific condition. After 12 reversals the staircase of that specific condition was terminated. Starting intensity was 19% SNR, which was for all participants above detection threshold. If for three of four conditions the staircase was finished, additional trials of the other conditions were randomly added to remove predictability about condition type. Participants were encouraged to take a break after every 30 trials.

If not stated otherwise procedures were the same as in Experiment 1.

2.4. Data analysis

2.4.1. Experiment 1

First, we constructed psychometric functions for detection thresholds for each condition. To construct these functions, we calculated the mean detection rate at each intensity level (in bins 1% SNR wide), separately for each condition. A cumulative Gaussian was fitted to the individual data with the psychometric fitting toolbox modelfree v1.1 (Zchaluk & Foster, 2009). To eliminate effects of hysteresis (see e.g. Fender & Julesz, 1967; Palmer, 1999; Ratlifl et al., 1986) it is common to average the 50% detection values over the two paradigms (increasing and decreasing). These calculated averages per condition were used as the dependent variable in a two-way repeated measures ANOVA with factors Rhythmicity (Rhythmic versus Random) and Modality (Auditory versus AudioVisual).

2.4.2. Experiment 2

An exponential decay was fitted for the four conditions separately for all the SNR values presented during the whole experiment (Treutwein, 1995), using the lsqnonlin function implemented in MATLAB. The function was as follows:

$$SNR(x) = 19e^{\lambda x} + C$$

where λ corresponds to the decay constant, C to the convergence value, and x to the trial number. The starting quantity was fixed at 19 (identical to the starting SNR in Experiment 2). To ensure that the lsqnonlin estimation did not result from a local minimum we repeated the procedure 30 times and took the final estimate as the fit with the most variance explained. The final convergence values of the exponential decay were used as the dependent variables in two-way repeated measures ANOVA with the factors Rhythmicity (Rhythmic versus Random) and Modality (Auditory versus AudioVisual).

3. Results

3.1. Experiment 1

For all participants, the mean percentage detection rate distribution had a shape typical of detection paradigms and could be reliably fit with a cumulative Gaussian function (Fig. 2; average explained variance 98.6%; see e.g. Florentine, Buus, & Geng, 1999; Green, 1995; Nachmias, 1981). The analysis using average 50% detection levels of the fitted psychometric functions (Fig. 3) revealed a main effect for Modality (Fig. 3b; $F(1,11) = 40.41$, $p < 0.001$, partial $\eta^2 = 0.786$), that indicated that Audiovisual stimuli yielded lower thresholds than Auditory stimuli. Also a main effect for Rhythmicity was found [$F(1,11) = 62.62$, $p < 0.001$, partial $\eta^2 = 0.851$], that showed lower threshold for Rhythmic stimuli compared to Random stimuli. The interaction effect was not significant [$F(1,11) = 1.30$, $p = 0.279$].

3.2. Experiment 2

The results of the 2AFC task are shown in Fig. 4. The fitted exponential decay explained on average 60.2% of the variance and converged to the threshold (Fig. 4a). A main effect for Modality was found [$F(1,19) = 38.68$, $p < 0.001$, partial $\eta^2 = 0.671$] in which Audiovisual stimuli yielded lower thresholds than Auditory stimuli. In addition, a main effect for Rhythmicity was found [$F(1,19) = 6.22$, $p = 0.022$, partial $\eta^2 = 0.247$], in which Rhythmic stimuli yielded lower threshold than Random stimuli. The interaction effect was not significant [$F(1,19) = 0.45$, $p = 0.510$].

4. Discussion

The aim of the current study was to investigate how temporally-predictive visual cues and within-modality temporal regularities might change the detection of near-threshold auditory stimuli. As anticipated, we found that both types of predictive information improve auditory perception, such that sounds at lower intensity levels are judged as audible if they are preceded by visual input and/or are part of a rhythmic sequence. Critically, we further show that the combination of two types of predictive information lowered detection intensities even further (Trommershauser et al., 2011), indicating that different sources of temporal information can be combined to optimize perception. We replicated the same pattern of results using two independent methods for quantifying perceptual thresholds, and show that these effects are not explained merely by participants' internal determinants such as response bias. Rather, our results suggest that temporal predictability can change perceptual sensitivity, and provides strong evidence for the 'attention in time' hypothesis (Barnes & Jones, 2000; Large & Jones, 1999; Schroeder et al., 2008). This is consistent with electrophysiological results showing increased neuronal excitability at moments in time that stimuli are expected (Cravo, Rohenkohl, Wyart, & Nobre, 2011; Besle et al., 2011; Cravo, Rohenkohl, Wyart, & Nobre, 2013; Lakatos et al., 2008, 2009, 2013).

4.1. Audiovisual effects

Visual cues lowered the auditory detection intensity in both experiments. Since there was a constant temporal delay between visual and auditory stimuli, we suggest that the temporal predictability between the stimuli can be used to temporally prepare for the auditory stimulus, as has been shown in previous studies (Lange & Röder, 2006; Los & Van der Burg, 2013). Indeed, in cued reaction time tasks, the largest decrease in reaction times is typically found when there is a constant delay between the cue and the target, and this advantage is reduced as the delay becomes more variable (Niemi & Näätänen, 1981). Moreover, there is a crucial temporal window during which audiovisual stimuli are integrated (Lindström, Paavilainen, Kujala, & Tervaniemi, 2012; Van Atteveldt, Formisano, Blomert, & Goebel, 2007; Van Wassenhove et al., 2007; Zampini, Shore, & Spence, 2003) and although the width of the window varies, the point of maximal integration is consistently when visual stimuli precede auditory stimuli (Thorne & Debener, 2008; Van Wassenhove et al., 2007). Similarly, electrophysiological recordings show an enhancement of the neural response to auditory tones when they are preceded by a somatosensory or visual stimulus (Lakatos, Chen,
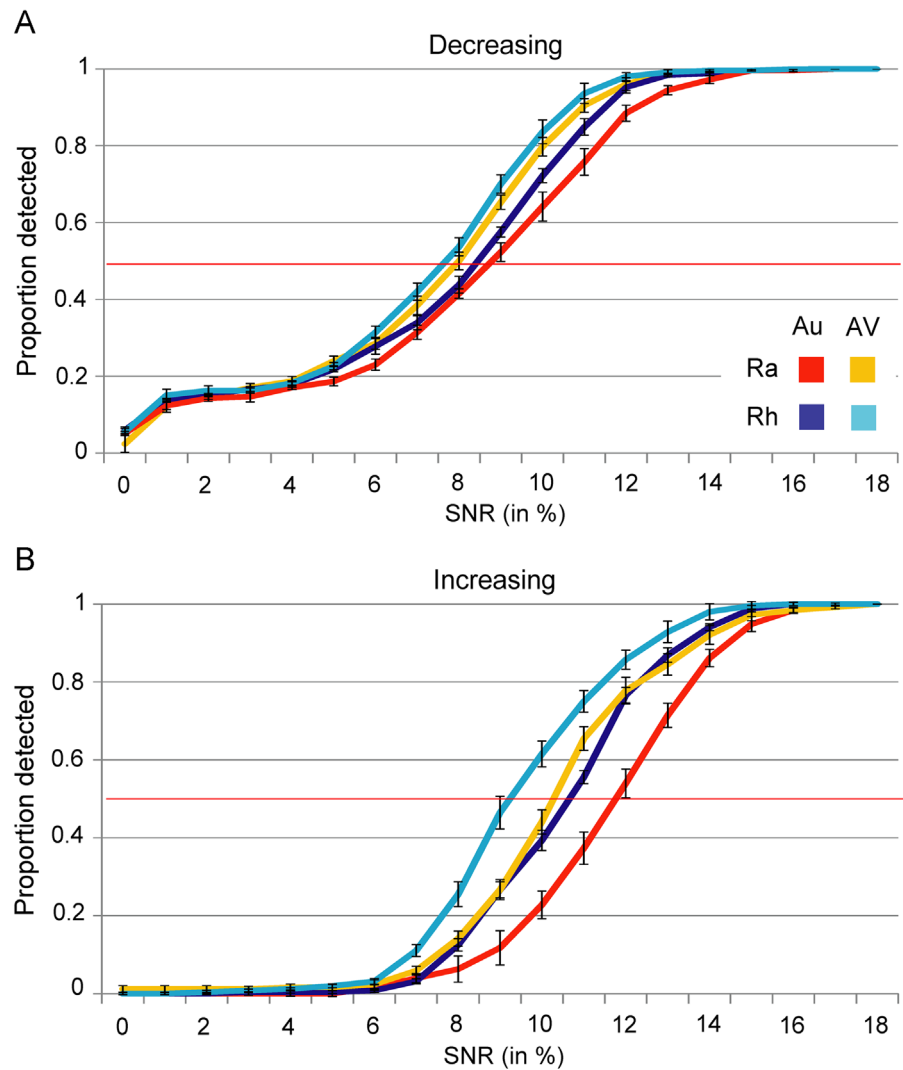
Fig. 2. Averaged detection percentages. Lines represent the average detection rate per intensity bin for the decreasing (A) and increasing (B) paradigm. All error bars represent the within-subjects standard error of the mean as described by (Morey, 2008).
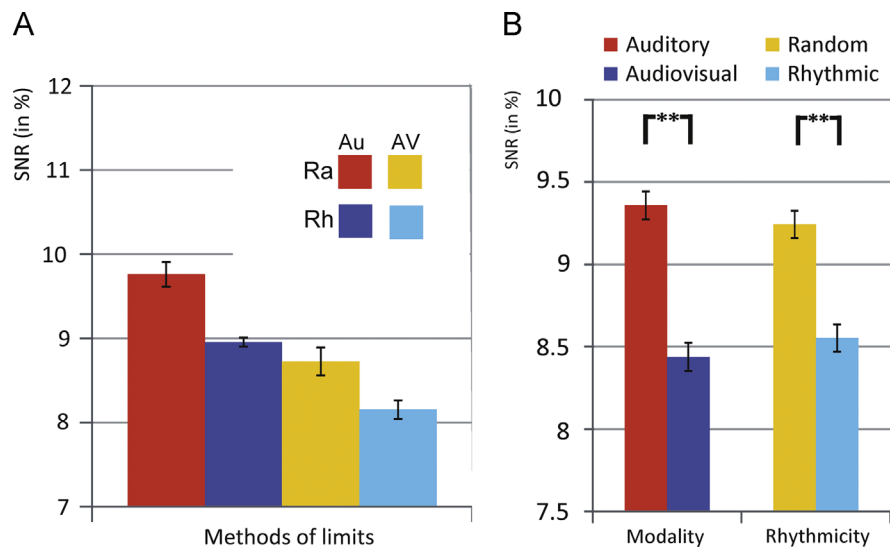

Fig. 3. (A) Average SNR for 50% detection rate derived from the two psychometric curves are shown for all conditions separately. (B) Rhythmicity and modality main effects were significant in both the increasing and decreasing paradigms. Error bars indicate the within-subjects standard error of the mean described by Morey (2008). Double asterisks indicate significance at *p*-values of 0.01.
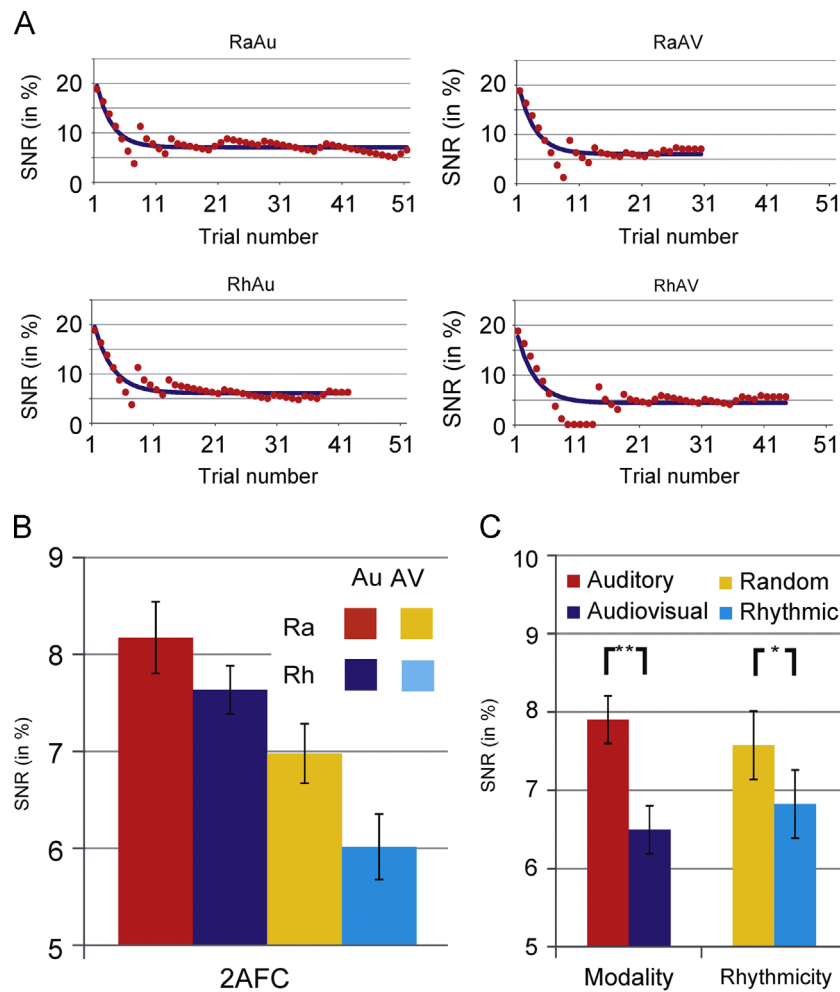
Fig. 4. (A) Example of fitted decay functions for all conditions for a representative participant. Red dots indicate the actual SNR values and blue lines indicates the fitted decay. (B) Average SNR for the convergence values of the fitted exponential decay shown for all conditions separately. (C) The results for the two main effects of Rhythmicity and Modality. Error bars indicate the within-subjects standard error of the mean described by Morey (2008). Asterisks and double asterisks indicate significance at *p*-values of 0.05 and 0.01, respectively.

O'Connell, Mills, & Schroeder, 2007; Kayser & Logothetis, 2009; Lakatos et al., 2009; Thorne, De Vos, Viola, & Debener, 2011; Wallace, Wilkinson, & Stein, 1996), with the largest AV effect found at an audiovisual SOA of ~65 ms. This SOA has also been found to have the largest behavioral AV facilitation effect (Thorne & Debener, 2008) and therefore we choose this SOA in the current study. Although in other studies 300 ms has been found as optimal reaction time facilitation effect (Niemi & Näätänen, 1981), we did not choose this SOA since then the visual stimulus would be exactly in anti-phase of the auditory rhythm. In addition, when audiovisual information is presented repeatedly with a fixed SOA temporal 'recalibration' occurs such that the audiovisual stimuli are more often perceived as synchronous at that SOA (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004). Although we did not specifically test for this, the fixed SOA used here probably induced this recalibration effect, which may have increased the integration of the audiovisual stimulus pair.

We cannot rule out the possibility that additional factors known to promote multisensory processing, such as spatial proximity (Plank, Rosengarth, Song, Ellermeier, & Greenlee, 2012; Wallace et al., 1996) or content congruency (Beauchamp, Lee, Argall, & Martin, 2004; Van Atteveldt, Formisano, Goebel, & Blomert, 2004) also contribute to the observed effects. For example, visual input was also found to increase detectability for low intensity auditory stimuli presented simultaneously (Lovelace,

Stein, & Wallace, 2003). Nonetheless, it seems plausible that the temporal relationship of the audiovisual stimulus pair used here, with the visual stimulus leading with a consistent, effective SOA contributed to reduction of the detection intensities by promoting temporal preparation.

4.2. The additive effects of rhythmicity and audiovisual cues

In the audiovisual conditions, Rhythmicity further reduced auditory detection compared to the Random condition. This result suggests that even though the visual input could perfectly predict the timing of auditory stimuli (since there was a constant lag of 65 ms), having temporal regularity within the sequence provides a significant additional benefit for perception. One reason for this finding could be that, since people's temporal estimates are not entirely accurate (Eisler, 1976), particularly in the visual domain (Welch & Warren, 1980), having two sources of predictive information in different time scales (666 ms between two sequential visual cues and 65 ms from the visual cue to the sound) sharpens temporal predictions. This is consistent with previous studies showing that complementary cues about a stimulus can be expressed by integrating the reliability of the perceptual estimates of individual cues, with the combination yielding higher reliability than each single cue alone (e.g. Yuille & Bülthoff, 1996; Landy, Maloney, Johnston, & Young, 1995; Oruç, Maloney, & Landy, 2003). Moreover, our data show an almost perfect additive

effect of rhythm and visual cues which suggest that the two temporal prediction processes might convey two separate and possibly independent mechanisms (Sternberg, 2001, Woodman, Kang, Thompson, & Schall, 2008). One possibility for how these two types of information may work in concert to additively lower detection thresholds is a 'winner take all' approach, in which the occurrence of upcoming events is predicted by parallel mechanisms separately utilizing either the rhythmic structure or the known cross-modal SOA, and behavioral detection occurs when a stimulus is detected through one of those mechanisms. Alternatively, it could be that the type of benefit afforded by rhythmic-visual cues is synergistic in essence and is qualitatively different from that provided by non-rhythmic visual cues (Correa & Nobre, 2008; Nobre, et al., 2007; Schroeder & Lakatos, 2009). According to this view, visual cues occurring at random times invoke a 'vigilance mode' of operation (Schroeder & Lakatos, 2009), since the participant cannot anticipate when the cue will occur, and once it has occurred only has 65 ms to orient attention towards a potential auditory stimulus, requiring the swift allocation of computational resources. However, if the visual cues themselves are presented rhythmically, participants can enter a 'rhythmic mode' in which the timing of all stimuli – both visual and auditory – is completely predictable. It has been suggested that such a 'rhythmic mode' is a more automatic and implicit process, requiring less metabolic demand, whereas a 'vigilance mode' requires explicit and controlled processing (Capizzi, Sanabria, & Correa, 2012; Correa, 2010; Schroeder et al., 2010; Van Atteveldt et al., 2011). Additional research is required to further characterize how these two modes of attention in time interact and work together to affect perceptual processing, particularly since the interpretation of additive effects on perceptual thresholds as reflecting contribution of sequential or parallel processes is not straightforward (Dubois, Poeppel, & Pelli, 2013; Sternberg, 2001; Miller, van der Ham, & Sanders, 1995).

4.3. The effect of rhythmicity in the absence of visual cues

Although not as effectively as the visual input, adding rhythmicity to the auditory stream also decreased detection thresholds. What is striking is that this effect occurred even when no visual stimulus was presented and auditory stimuli were not yet detected. This exciting finding implies that the rhythmic pattern of tones is detected at lower intensities than each individual tone. It suggests a central role for temporal integration and detection of temporal regularities in near-threshold perception. This pattern is in line with previous findings that auditory detection thresholds are reduced for low intensity auditory stimuli when they are presented for a longer duration, supposedly brought about by the aggregation of subthreshold information (Florentine, Fastl, & Buus, 1988; Lütkenhöner, 2011; Yrttiaho, Tiitinen, Alku, Miettinen, & May, 2010) (similar effects are found in the visual system, see: Anstis, Geier, & Hudak, 2012; Daikhin & Ahissar, 2011; Minelli, Marzi, & Girelli, 2007). Since it has been shown that 3–4 stimuli are sufficient for a neuronal population to identify a rhythmic structure (Lakatos et al., 2008; Thorne et al., 2011), it seems plausible that the (even subliminal) processing of a few rhythmic stimuli provides enough information about the temporal structure of the auditory stream to decrease perceptual thresholds. One intriguing question is how precisely isochronous a stimulus trains needs to be to still provide a perceptual benefit. Neuronal entrainment has been previously demonstrated for tone sequences with a temporal jitter of up to 20% (SOAs distributed between 666 ± 150 ms; Lakatos et al., 2008). This suggests that the system can tolerate some degree of jitter and still maintain temporal predictions that can facilitate perception, which would be beneficial from an ecological perspective since many real-life stimuli such as speech and music have temporal regularities but are not perfectly isochronous. The robustness of temporal prediction to

jitter, and its influence on both perception and on neural processing needs to be systematically studied in future work.

5. Conclusion

We show that both temporal regularity within a stimulus stream and cross-modal temporal cueing decrease auditory detection thresholds. Moreover, both types of temporal information are used in combination to prepare our system for incoming stimuli and may play complementary roles in focusing 'attention in time'. These findings are a testament to the flexibility and proactivity of the perceptual system (Schroeder et al., 2010; Zion Golumbic et al., 2012; Van Atteveldt, Murray, Thut, & Schroeder, 2014), in that thresholds for reporting auditory detection are not necessarily fixed but rather are strongly influenced by contextual factors, like those tested here. Our findings have implications for understanding the role of temporal prediction in processing more complex and natural stimuli, such as speech, which contain both intrinsic regularities (Giraud & Poeppel, 2012; Greenberg et al., 2003; Luo & Poeppel, 2007) as well as temporally predictive cross-modal cues such a facial and head movements (Schroeder et al., 2008; Chandrasekaran et al., 2009; Grant & Seitz, 2000; Munhall & Vatikiotis-Bateson, 2004; Schwartz, Berthommier, & Savariaux, 2004; Zion Golumbic, Cogan, Schroeder, & Poeppel, 2013), both of which are likely to influence the fundamental operations of auditory cortex (Lakatos et al., 2013; Schroeder et al., 2008). Although additional research is needed to understand the usage of multiple contextual factors during perception, we show that contextual information can be combined from different sources to allocate our attention in time, thereby sensitizing and optimizing perception.

References

Anstis, S., Geier, J., & Hudak, M. (2012). Afterimages from unseen stimuli. *i-Perception*, 3(8), 499–502.
Bar, M. (2011). *Predictions in the Brain: Using Oour Past to Generate a Future*. USA: OUP.
Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time. *Cognitive Psychology*, 41(3), 254–311.
Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, 41(5), 809–823.
Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., et al. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *The Journal of Neuroscience*, 31(9), 3176–3185.
Capizzi, M., Sanabria, D., & Correa, Á. (2012). Dissociating controlled from automatic processing in temporal preparation. *Cognition*, 123(2), 293–302.
Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, 5(7), e1000436.
Correa, A. (2010). Enhancing behavioral performance by visual temporal orienting. *Attention and Time*, 357–370.
Correa, Á., Lupiáñez, J., Milliken, B., & Tudela, P. (2004). Endogenous temporal orienting of attention in detection and discrimination tasks. *Attention, Perception, & Psychophysics*, 66(2), 264–278.
Correa, Á., & Nobre, A. C. (2008). Neural modulation by regularity and passage of time. *Journal of Neurophysiology*, 100(3), 1649–1655.
Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *The Journal of Neuroscience*, 18(18), 7426–7435.
Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2011). Endogenous modulation of low frequency oscillations by temporal expectations. *Journal of Neurophysiology*, 106(6), 2964–2972.
Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *The Journal of Neuroscience*, 33(9), 4002–4010.

Daikhin, L., & Ahissar, M. (2011). Responses to deviants are modulated by subthreshold variability of the standard. *Psychophysiology*, *49*(1), 31–42.

Dubois, M., Poeppel, D., & Pelli, D. G. (2013). Seeing and hearing a word: combining eye and ear is more efficient than combining the parts of a word. *PloS One*, *8*(5), e64803.

Eisler, H. (1976). Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin*, *83*(6), 1154.

Ellis, R. J., & Jones, M. R. (2010). Rhythmic context modulates foreperiod effects. *Attention, Perception, & Psychophysics*, *72*(8), 2274–2288.

Fender, D., & Julesz, B. (1967). Extension of Panum's fusional area in binocularly stabilized vision. *JOSA*, *57*(6), 819–826.

Florentine, M., Buus, S., & Geng, W. (1999). Psychometric functions for gap detection in a yes–no procedure. *The Journal of the Acoustical Society of America*, *106*, 3512.

Florentine, M., Fastl, H., & Buus, S. (1988). Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *The Journal of the Acoustical Society of America*, *84*, 195.

Friston, K. (2011). Prediction, perception and agency. *International Journal of Psychophysiology*, *83*(2), 248–252.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. y. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778.

Gescheider, G. (1997). *The classical psychophysical methods* (Chapter 3:). *Psychophysics:Tthe Fundamentals*. Hillsdale, NJ: L. E. Associates (3rd ed.).

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517.

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, *108*, 1197.

Green, D. M. (1995). Maximum-likelihood procedures and the inattentive observer. *The Journal of the Acoustical Society of America*, *97*, 3749.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* Vol. 1974. New York: Wiley.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics*, *31*(3), 465–485.

Griffin, I. C., Miniussi, C., & Nobre, A. C. (2001). Orienting attention in time. *Frontiers in Bioscience*, *6*, 660–671.

Jones, M. R., Johnston, H. M., & Puente, J. (2006). Effects of auditory pattern structure on anticipatory and reactive attending. *Cognitive Psychology*, *53*(1), 59–96.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, *13*(4), 313–319.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception, & Psychophysics*, *49*(3), 227–229.

Kayser, C., & Logothetis, N. K. (2009). Directed interactions between auditory and superior temporal cortices and their role in sensory integration. *Frontiers in Integrative Neuroscience*, *3*.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*(2), 279–292.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*(5872), 110–113.

Lakatos, P., Musacchia, G., O'Connell, M., Falchier, A., Javitt, D., & Schroeder, C. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, *77*(4), 750–761.

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron*, *64*(3), 419–430.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, *35*(3), 389–412.

Lange, K., & Röder, B. (2006). Orienting attention to points in time improves stimulus processing both within and across modalities. *Journal of Cognitive Neuroscience*, *18*(5), 715–729.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychological Review*, *106*(1), 119.

Lindström, R., Paavilainen, P., Kujala, T., & Tervaniemi, M. (2012). Processing of audiovisual associations in the human brain: dependency on expectations and rule complexity. *Frontiers in Psychology*, *3*.

Los, S. A., & Van der Burg, E. (2013). Sound speeds vision through preparation, not integration. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 1612–1624.

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research*, *17*(2), 447–453.

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, *8*(8), e1000445.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001–1010.

Lütkenhöner, B. (2011). Auditory signal detection appears to depend on temporal integration of subthreshold activity in auditory cortex. *Brain Research*, *1385*, 206–216.

Mathewson, K. E., Fabiani, M., Gratton, G., Beck, D. M., & Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition*, *115*(1), 186–191.

Miller, J., van der Ham, F., & Sanders, A. F. (1995). Overlapping stage models and the additive factor method. *Acta Psychologica*, *90*, 11–28.

Minelli, A., Marzi, C. A., & Girelli, M. (2007). Lateralized readiness potential elicited by undetected visual stimuli. *Experimental Brain Research*, *179*(4), 683–690.

Morey, R. D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Reason*, *9*(45.4), 49.46.

Munhall, K., & Vatikiotis-Bateson, E. (2004). Spatial and temporal constraints on audiovisual speech perception. In: S. C. Calvert, & G. A. Stein B. (Eds.), *The Handbook of Multisensory Processing* (pp. 177–188). Cambridge, Massachusetts: The MIT Press.

Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research*, *21*(2), 215–223.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, *89*(1), 133.

Nobre, A., Correa, A., & Coull, J. (2007). The hazards of time. *Current Opinion in Neurobiology*, *17*(4), 465–470.

Nobre, K., & Coull, J. T. (2010). *Attention and time*. Oxford, UK: Oxford University Press.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23), 2451–2468.

Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. Cambridge, MA: The MIT press.

Plank, T., Rosengarth, K., Song, W., Ellermeier, W., & Greenlee, M. W. (2012). Neural correlates of audio-visual object recognition: effects of implicit spatial congruency. *Human Brain Mapping*, *33*(4), 797–811.

Ratlifl, F., Comsweet, T., Ross, J., Holt, J., Johnstone, J., Ratliff, F., et al. (1986). Hysteresis in the perception of motion direction as evidence for neural cooperativity. *Nature*, *324*, 20.

Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *The Journal of Neuroscience*, *32*(24), 8424–8428.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, *12*(3), 106–113.

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology*, *20*(2), 172–176.

Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in Cognitive Sciences*, *11*(5), 211–218.

Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, *93*(2), B69–B78.

Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica*, *106*, 147–246.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science*, *314*(5803), 1311–1314.

Ten Oever, S., Sack, A., Wheat, K. L., Bien, N., & Van Atteveldt, N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Frontiers in Psychology*, *4*.

Thorne, J. D., De Vos, M., Viola, F. C., & Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. *The Journal of Neuroscience*, *31*(10), 3853–3861.

Thorne, J. D., & Debener, S. (2008). Irrelevant visual stimuli improve auditory task performance. *NeuroReport*, *19*(5), 553.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, *35*(17), 2503–2522.

Trommershauser, J., Kording, K., & Landy, M. S. (2011). *Sensory Cue Integration*. New York: Oxford University Press.

Van Atteveldt, N., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*, *17*(4), 962–974.

Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*(2), 271–282.

Van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron*, *81*(6), 1240–1253.

Van Atteveldt, N., Musacchia, G., Sehatpour, P., Zion Golumbic, E. M., Lakatos, P., Gaspar, P., et al. (2011). A combined EEG-fMRI investigation of "rhythmic" versus "vigilant" processing II: functional neuro-anatomy and slow dynamics, *In Proceedings of the International Conference on Cognitive Neuroscience (ICON XI), Mallorca, Spain*.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, *22*(1), 32–35.

Wallace, M., Wilkinson, L., & Stein, B. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology*, *76*(2), 1246–1266.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638.

Woodman, G. F., Kang, M.-S., Thompson, K., & Schall, J. D. (2008). The effect of visual search efficiency on response preparation. *Psychological Science*, *19*, 128–136.

Yrttiaho, S., Tiitinen, H., Alku, P., Miettinen, I., & May, P. J. (2010). Temporal integration of vowel periodicity in the auditory cortex. *The Journal of the Acoustical Society of America, 128*(1), 224–234.

Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In: D. C. Knill, & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123–161). Cambridge University Press.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research, 152*(2), 198–210.

Zchaluk, K., & Foster, D. H. (2009). Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics, 71*(6), 1414–1425.

Zion Golumbic, E. M., Cogan, G. B., Schroeder, C. E., & Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "Cocktail Party". *The Journal of Neuroscience, 33*(4), 1417–1426.

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. Brain and Language.