ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Scalp Potentials of Tonal Brain's Speaker in Audiovisual Perception

Wichian Sittiprapaporn*

Faculty of Medicine, Mahasarakham University, Maha Sarakham, Thailand.

ABSTRACT

The purpose of the present study was to study the tonal brain's speaker related to the audiovisual integration mechanisms for letters, i.e., for stimuli that have been previously associated through learning. The brain activations were detected with electroencephalogram (EEG), which is well suited for noninvasive identification of cortical activity and its accurate temporal dynamics. The audiovisual interaction was used as an indicator for investigating the automatic processing of simultaneously audiovisual integration related to the suprasegmental information in the tonal brain's speaker. Multisensory integration of letters (orthography) and speech sounds of tonal language in the human auditory association cortex of the tonal brain's speaker showed a strong dependency on the relative timing of the inputs. The critical role of input timing on multisensory integration has been demonstrated before at the neuronal level for naturally related visual and auditory signals. The critical role of input timing on multisensory integration has been demonstrated before at the neuronal level for naturally related visual and auditory signals.

Key words: Tonal language; Brain; Audiovisual integration; Multisensory; Scalp potentials

Corresponding author

INTRODUCTION

Nowadays most written languages are speech-based alphabetic scripts, in which speech sound units (phonemes) are represented by visual symbols (letters, or graphemes). Associations between sensory events in different modalities can either be defined by natural relations (e.g., the shape and sound of a natural object) or by more artificial relations. A neural mechanism for the integration of audiovisual speech has been suggested by Calvert and colleagues [1, 2] and supported by other neuroimaging findings on audiovisual speech perception [3-5] and lip reading [6-8]. Results of these studies suggested that the perceptual gain experienced when perceiving multimodal speech was accomplished by enhancement of the neural activity in the relevant sensory cortices. The left posterior superior temporal sulcus (STS) has been advanced as the heteromodal site that integrated visual and auditory speech information and modulated the modality-specific cortices by back projections [1, 2]. Modality-specific regions involved in this mechanism were the visual motion processing area V5 and auditory association areas in superior temporal cortex.

In addition to this interplay between STS and sensory cortices, frontal and parietal regions seem to be involved, although activation of these regions was less consistent between the different studies. Interestingly, some reports showed that the STS have a more general role in the integration of cross-modal identity information. The involvement of the left posterior STS in the integration of auditory and visual nonlinguistic information has also been reported recently [9, 10]. The purpose of the present study was then to study the tonal brain's speaker related to the audiovisual integration mechanisms for letters, i.e., for stimuli that have been previously associated through learning.

MATERIALS AND METHODS

Participants

In this study, subjects were healthy and had normal hearing and vision (self reported). Fourteen adult, native speakers of Mandarin Chinese (7 males; 7 females) participated in the event-related potential (ERP) experiment. All participants were closely matched with respect to age (mean = 23.2, SD = 4.1) and years of formal education (mean = 17.2, SD = 2.3). None of the participants had more than three years of formal musical training and none had any musical training within the past five years. All participants were paid for their participation. Informed consents were distributed to all participants involved in this study.

Stimuli

Stimuli consisted of a set of four Mandarin Chinese words that were distinguished minimally by tonal contour (*pinyin* Roman transliteration): yi^1 'clothing' [T1]; yi^2 'aunt' [T2]; yi^3 'chair' [T3]; yi^4 'easy' [T4]. Only three of the four Mandarin Chinese tones (T1, T2, T3) were chosen for presentation in a oddball paradigm. This limitation restricted electroencephalogram (EEG) recording time to 90 mins, thus minimizing the risk of subject fatigue. The experiment consisted of an oddball condition. The duration of the stimuli were 300 msec. The audiovisual experiment included four stimuli: congruent / yi^1/ (acoustic /yi^1/

+ visual / yi^1/), congruent /yi^2/ (acoustic /yi^2/ + visual /yi^2/), incongruent /yi^1/ (acoustic /yi^1/ + visual /yi^2/) and congruent /yi^4/ (acoustic /yi^4/ + visual /yi^4/) (see Figure 1 and Figure 2).

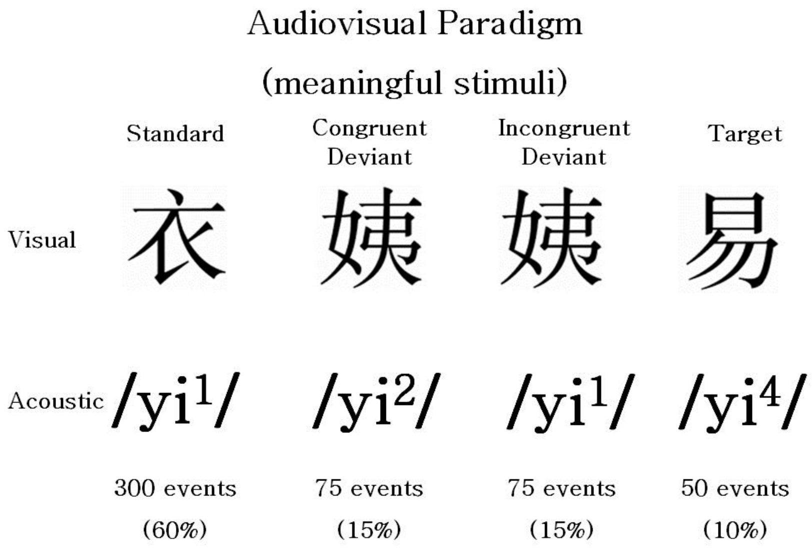

Figure 1: The meaningful audiovisual experiment included four stimuli. Standard: congruent /yi^1/ (acoustic /yi^1/ + visual / yi^1/); Deviant: congruent /yi^2/ (acoustic /yi^2/ + visual /yi^2/); Deviant: incongruent /yi^1/ (acoustic /yi^1/ + visual /yi^2/); and Target: congruent /yi^4/ (acoustic /yi^4/ + visual /yi^4/).

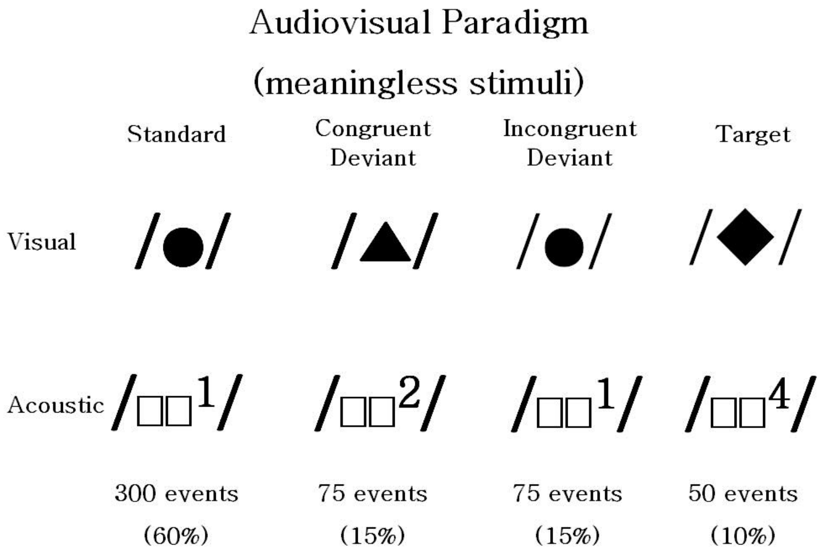

Figure 2: The meaningless audiovisual experiment included four stimuli. Standard: congruent /yi^1/ (acoustic /□□^{1}/ + visual /●/); Deviant: congruent /yi^2/ (acoustic /□□^{2}/ + visual /▲/); Deviant: incongruent /yi^1/ (acoustic /□□^{1}/ + visual /●/); and Target: congruent /□□^{4}/ (acoustic /yi^4/ + visual /♦/).

Procedure

Subjects sat in an electrically shielded and soundproofed room with the response buttons under their hands. The subject had to press the button on the response pad when the target was presented and ignore any other types of stimuli. Prior to the experimental session, a practice block was administrated to ensure that the subjects understood the task. In an initial practice run, the task difficulty (i.e. target discriminability) was individually adjusted to about 75% correct responses for both audiovisual target stimuli. Each experiment was consisted of 2 blocks and each block had 300 trials. There were 6 blocks of all experiments. Every stimulus was presented with 300 msec exposure duration and inter-stimulus interval was 1,000ms in every condition. Stimulus sequences were presented to the subjects with STIM2 software. The stimulus onset asynchrony was 1300 msec (from acoustic/visual speech onset to onset). Stimulus sequences consisted of frequent (P = 0.60) congruent /yi^1/ stimuli and congruent (P = 0.15) and incongruent (P = 0.15) /yi^2/ stimuli. Congruent /yi^4/ stimuli were presented as target (P = 0.10) to be able to check that subjects were attending the stimuli. Randomized stimulus sequences were presented consisting of equiprobable audiovisual stimuli (a simultaneous combination of auditory and visual). Sound density was adjusted to be 85 dB above the subject's hearing threshold (defined for the audiovisual stimulus sequence). Visual stimuli were presented on the computer screen and acoustic stimuli were simultaneously presented in audiovisual experiment.

Event-related Potential (ERP) Recordings and Analysis

EEG was recorded with a Quick-Cap equipped with 64 channels according to the international 10-20 system using Scan system (Scan 4.3, Neurosoft, Inc. Sterling, USA). Reference electrode was at mastoids. The signals were bandpass filtered at 0.05-100 Hz and digitized at 1000 Hz. The impedance of the electrode was below 5 kΩ. Eye movements were monitored with two EOG electrodes. Four electrodes monitored horizontal and vertical eye movements for off-line artifact rejection. Vertical and horizontal electro-oculogram (EOG) was recorded by electrodes situated above and below the left eye, and on the outer canthi of both eyes, respectively. After the data recordings, the EEG was segmented into 1,000 msec epochs, including the 100 msec pre-stimulus period. The baseline was corrected separately for each channel according to the mean amplitude of the EEG over the 100 msec period that preceded stimulus onset. The EEG epochs contained amplitudes exceeding ±100 μV at any EEG channels were automatically excluded from the averaging. The epoch was separately averaged for the standard, deviant, and the target stimulus. The average waveforms obtained from the standard, deviant and target stimuli were digitally filtered by a 0.1 - 15 Hz band-pass filter and finally baseline-corrected.

Data pre-processing and feature extraction

For each subject, mean ERP map series for the audiovisual stimuli were computed over the 6 blocks where each block was weighted by the number of averaged sweeps that it consisted of. All mean map series were carefully inspected for artifacts. The grand mean map series over subjects and conditions was then computed. These mean ERP map series were recomputed to average reference, the EOG channel was removed from the data, and Fpz was linearly interpolated as mean of Fp1, Fp2 and Fz in order to be compatible with existent analysis software. For all mean ERP map series, the locations of the centroid of each map were computed [11]. Centroids were the points of gravity of the positive and the negative areas of an average reference-referred map. For each of these centroid location points, the location coordinates were determined on the left-right axis and on the anterior-posterior axis. Thus, a single map was described by four coordinate values. All subsequent analysis steps (segmentation of the data into microstates and statistical analysis) were based on these extracted spatial descriptors of the maps.

Assessment of changes of spatial map configuration and landscape

For the analysis of changes of spatial map configuration, the curves of the centroid locations over time were averaged over subjects. As the goal was to investigate whether there are periods of stable spatial map configuration in the data, methods for space based segmentation of the ERP map series [12-14] were used. The landscape of each map, i.e., the spatial configuration of its potential distribution, was assessed numerically by the locations of the centroids of the positive and negative map areas [11]. In this way four parameters were obtained for each map; the location of the centroid of the positive map on the left-right axis and on the anterior-posterior axis; and the location of the negative map area centroid on the left-right axis and on the anterior-posterior axis.

Statistical analysis

Statistical analysis was performed on the Global Field Power (GFP) area of 21 electrodes sites within the time range of difference waveform of cross-modal interaction (100-250 msec). Two conditions were speech and non-speech sounds. For statistical testing two-tailed *t*-tests were carried out comparing mean amplitudes within specified time windows. ERP was analyzed with a repeated measure (condition x electrode site). Four electrodes sites such as prefrontal line (FP1, FPz, Fp2), frontal line (F7, F3, Fz, F4, F8), central line (T7, C3, Cz, C4, T8), parietal midline (P7, P3, Pz, P4, P8), and occipital line (O1, Oz, O2) sites were used.

RESULTS AND DISCUSSION

Figure 3 and 4 show the mean locations of the landscape centroids of the N1 maps at group centre latency for meaningful and meaningless stimuli. None of the landscape differences as represented by the centroids reached significant double-ended *p*-values in meaningful stimuli. In meaningless stimuli, there were statistical differences of interest along the anterior-posterior axis: The positive (posterior) centroid was more anterior ($P < 0.23$) in meaningful than in meaningless condition. In the same way, the negative (anterior) centroid was more posterior ($P < 0.17$) in meaningful than in meaningless condition. Thus, the anterior-posterior distance between centroids was smaller in meaningless than in meaningful condition. Tested directly, this difference was significant at $P < 0.03$. There were no statistically relevant differences along the left-right axis.

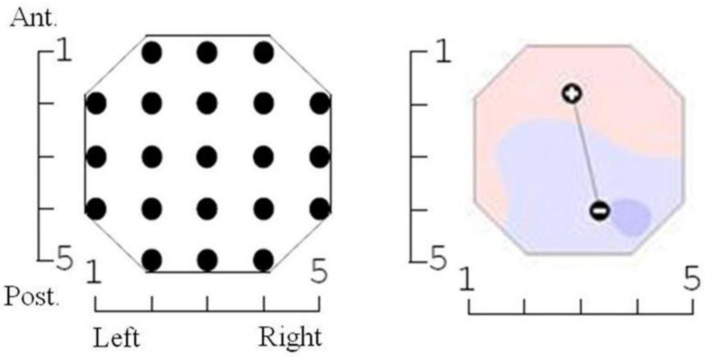

Figure 3: **The locations of the mean (across subjects) map centroids at the peak latencies for the meaningful stimuli per AV integration experimental condition. Anterior centroids are negative (-), posterior centroids are positive (+). Ant. = Anterior; Post. = Posterior**

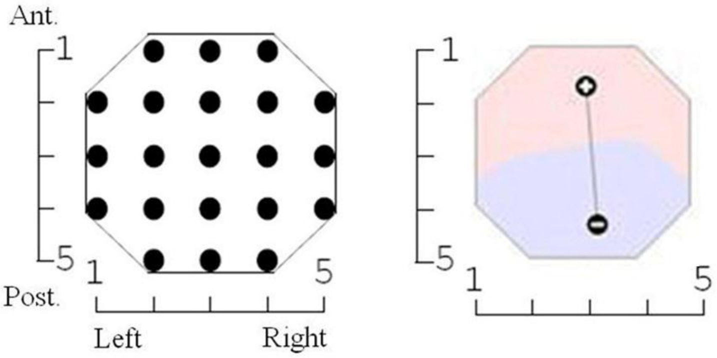

Figure 4: **The locations of the mean (across subjects) map centroids at the peak latencies for the meaningless stimuli per AV integration experimental condition. Anterior centroids are negative (-), posterior centroids are positive (+).Ant. = Anterior; Post. = Posterior**

Figure 5 and 6 illustrate maps series of potential distributions evoked by both conditions, and they are obvious that both the configuration and strength of the fields change over time: the evoked fields are strong between 91 msec and 181 msec evoked by meaningful stimuli (Fig. 5) and meaningless stimuli (Fig. 6). During this period parieto-occipital negativity gradually builds up, and the field reaches a maximal pronunciation at 89 msec with a high occipital peak surrounded by densely packed Equipotential lines. At later times this peak slowly diminished and finally is replaced by a relative negativity over the occipital.

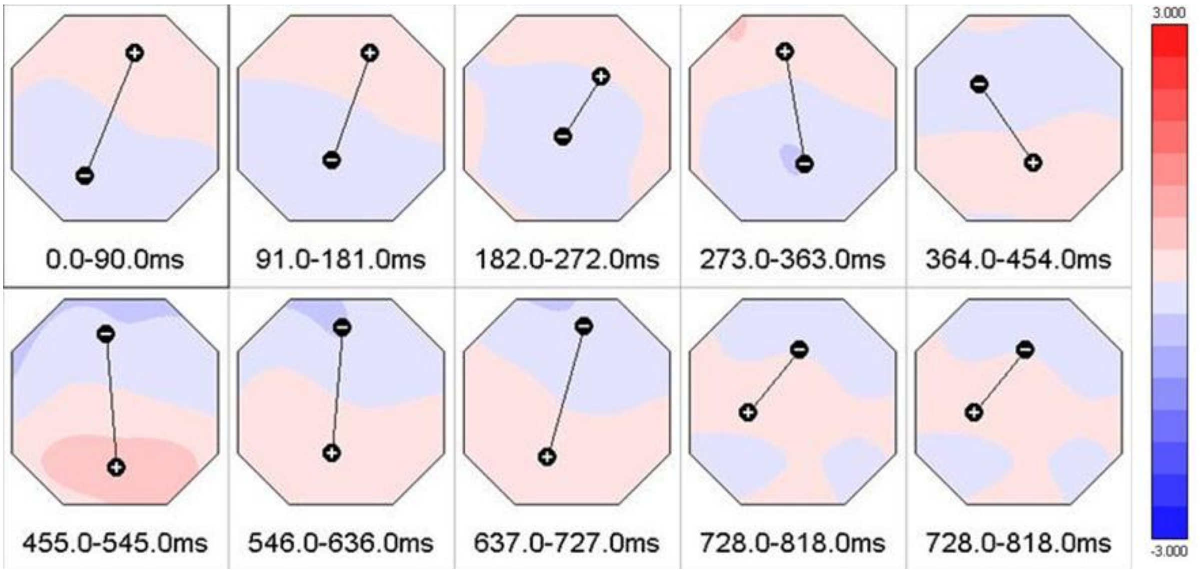

Figure 5: Series of scalp potential fields of Global Field Power as function of time from 0 and 900 msec with intervals of 90 msec evoked by the meaningful stimuli per the AV integration experimental condition. Occurrence time of maximal global filed power from 91 and 181 msec. Equipotential lines in steps of 3µV. For color-coded amplitude values in the maps, refer to color scale on the right.

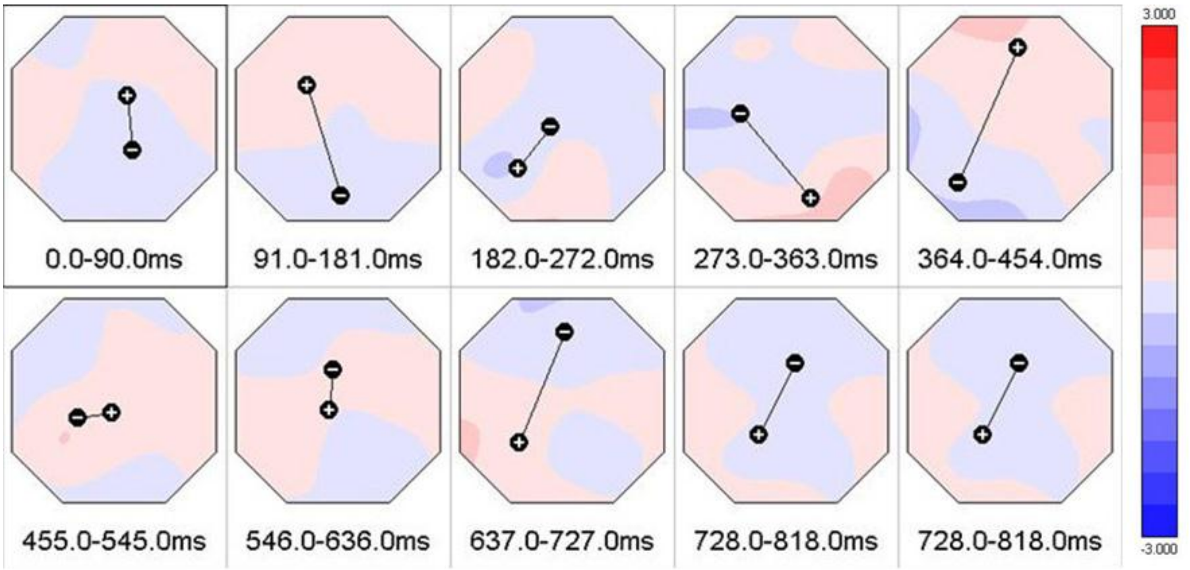

Figure 6: Series of scalp potential fields of Global Field Power as function of time from 0 and 900 msec with intervals of 90 msec evoked by the meaningless stimuli per the AV integration experimental condition. Occurrence time of maximal global filed power from 91 and 181 msec. Equipotential lines in steps of 3µV. For color-coded amplitude values in the maps, refer to color scale on the right.

The present results corroborate previous findings showing that visual speech has access to the early levels of auditory processing hierarchy [3, 6, 15-19] and support the auditory integration models. Electrophysiological studies in monkeys suggest that auditory cortex responses to visual stimuli are due to projections from higher cortical regions [20-21]. EEG studies show that the auditory N100 amplitude is suppressed during audiovisual speech stimulation in comparison to the sum of unimodal responses [15, 18, 19]. These studies indicate that in terms of processing time there are early (within 100 ms from the stimulus onset) audiovisual interactions in the auditory cortical areas. In support, electrophysiological studies in monkeys show that responses to visual stimuli in auditory cortex neurons are very early (~50 ms from stimulus onset) [20, 21]. These results suggest that there are audiovisual interactions in auditory cortical areas before the phonetic categorization of the speech input. Interactions occur also during or after the approximate time-window of phonetic categorization (> 50 msec) possibly through feedback to Superior Temporal Gyrus (STG) from STS or other multisensory areas [3, 16, 17]. Integration of auditory and visual non-speech information is primarily based on temporal and spatial coincidence of the stimuli. These mechanisms are important in audiovisual integration of speech as well. However, seeing and hearing speech provide also phonetic information. Therefore, both general and speech-specific multisensory mechanisms might be important in audiovisual perception of speech [10, 15].

CONCLUSION

The audiovisual interaction is used as an indicator for investigating the automatic processing of simultaneously audiovisual integration related to the suprasegmental information in the tonal brain's speaker. Multisensory integration of letters (orthography) and speech sounds of tonal language in the human auditory association cortex of the tonal brain's speaker showed a strong dependency on the relative timing of the inputs. The critical role of input timing on multisensory integration has been demonstrated before at the neuronal level for naturally related visual and auditory signals.

REFERENCES

[1] Calvert G A, Brammer M J, Bullmore E T, Campbell R, Iversen S D, David A S. Neuroreport 1999; 10: 2619-2623.

[2] Calvert G A, Campbell R, Brammer M J. Curr Biol 2000; 10: 649-657.

[3] Sams M, Aulanko R, Hämäläinen M, Haru R, Lounasmaa O V, Lu S-T. Neurosci Lett 1991; 127: 141-145.

[4] Sekiyama, K., I. Kanno, S. Miura and Y. Sugita. Neurosci Res 2003 47: 277-287.

[5] Wright T M, Pelphrey K A, Allison T, McKeown M J, McCarthy G. Cerebral Cortex 2003; 13: 1034-1043.

[6] Calvert G A, Bullmore E T, Brammer M J, Campbell R, Williams S C, McGuire P K. Science 1997; 276: 593-6.

[7] Calvert GA, Campbell R. J Cog Neurosci 2003; 15: 57-71.

[8] Paulesu E, Perani D, Blasi V, Silani G, Borghese N A, De Giovanni U., Sensolo S, Fazio F. J Neurophysiol 2003; 90: 2005-2013.

[9] Beauchamp M, Lee K, Argall B, Martin A. Neuron 2004; 41: 809-823.

[10] Calvert G A, Hansen P C, Iversen S D, Braummer M J. Neuroimage 2001; 14: 427-438.

[11] Wackermann J, Lehmann D, Michel, C M, Strik W K. Int J Psychophysiol 1993; 14: 269-283.

[12] Lehmann D, Skrandies W. Electroencephalog Clin Neurophysiol 1980; 48: 609-621.

[13] Lehmann D, Skrandies W. Electroencephalog Clin Neurophysiol 1986; 38(S): 506-512.

[14] Lehmann D. Principles of spatial analysis. In Gevins A S, Remond A. (eds), Methods

of Analysis of Brain Electrical and Magnetic Signals. Handbook of Electroencephalography and Clinical Neurophysiology, revised series, Volume 1. Elsevier, Amsterdam, 1987; pp. 309-354.

[15] Klucharev V, Möttönen R, Sams M. Brain Res Cog Brain Res 2003; 18: 65-75.

[16] Möttönen R, Krause C M, Tiippana K, Sams M. Brain Res Cog Brain Res 2002; 13: 417-25.

[17] Möttönen R, Schurmann M, Sams M. Neurosci Lett 2004; 363: 112-5.

[18] Besle J, Fort A, Delpuech C, Giard M H. Euro J Neurosci 2004; 20: 2225-34.

[19] van Wassenhove V, Grant K W, Poeppel D. Proc Natl Acad Sci U S A 2005; 102: 1181-1186.

[20] Schroeder C E, Foxe J. Brain Res 2002; 14: 187-198.

[21] Schroeder C E, Smiley J, Fu K G, McGinnis T, O'Connel M N, Hackett T A. Int J Psychophysiol 2003; 50: 5-17.