



# Journal of Applied Sciences

ISSN 1812-5654

**science**  
alert

**ANSI***net*  
an open access publisher  
<http://ansinet.com>

## Scalp Potentials of Human Brain's Audiovisual Integration

Wichian Sittiprapaporn

Faculty of Medicine, Mahasarakham University, Maha Sarakham, Thailand

---

**Abstract:** The aim of the present study was to investigate the human brain's audiovisual integration mechanisms for letters, i.e., for stimuli that have been previously associated through learning. The subjects received audiovisual (AV) letters of the Chinese characters and were required to identify them, regardless of stimulus modality. The brain activations were detected with electroencephalogram (EEG), which is well suited for noninvasive identification of cortical activity and its accurate temporal dynamics. The present study was able to find evidence of both non-phonetic and phonetic audiovisual interactions in the Event-Related Potentials (ERPs) to the same AV stimuli. In addition, the differences in the ERPs to the meaningful and meaningless of AV stimuli probably reflect multisensory interactions in phonetic processing. When acoustic and visual phonemes were meaningful, they formed a natural multisensory interaction stimulus. The present study demonstrates that the audiovisual interaction is an indicator for investigating the automatic processing of suprasegmental information in tonal language. Multisensory integration of letters (orthography) and speech sounds of tonal language in the human auditory association cortex 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.

**Key words:** Letters, sounds, audiovisual, language, multisensory

---

### INTRODUCTION

Reading is essential to social and economic success in the present technological society (National Reading Council, 1998). In contrast to spoken language, which is a product of biological evolution, reading and writing are cultural inventions from the last few thousand years and are only relevant for most people since a few hundred years (Liberman, 1992). An intriguing question is, therefore, how it is possible that most people acquire literacy skills with such remarkable ease even though a naturally evolved brain mechanism for reading is unlikely to exist. An interesting hypothesis is that evolutionarily adapted brain mechanisms for spoken language provide a neural foundation for reading ability, which is illustrated by the low literacy levels in deaf people (Perfetti and Sandak, 2000).

Nowadays most written languages are speech-based alphabetic scripts, in which speech sound units (phonemes) are represented by visual symbols (letters, or graphemes). Learning the correspondences between letters and speech sounds of a language is, therefore, a crucial step in reading acquisition, failure of which is thought to account for reading problems in developmental dyslexia (Frith, 1985). However, in the normal situation, letter-speech sound associations are learned and used with high efficiency. At least 90% of school children learn

the letter-sound correspondences without exceptional effort with a few months (Blomert, 2002), which is a remarkable achievement, since our brains are not phylogenetically adapted to the requirements for acquiring written language.

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. In contrast to the culturally defined associations between letters and speech sounds (Raij *et al.*, 2000), lip reading is based on naturally developed associations of speech with visual information (Paulesu *et al.*, 2003). Therefore, it seems a plausible assumption that the perception of speech and the inherently linked lip movements (hereafter referred to as audiovisual speech) emerged simultaneously during evolution, shaping the brain for integrating this audiovisual information.

At the behavioral level, it has been reported that speech perception can be influenced both by lip movements and by letters. Sumby and Pollack (1954) showed that lip reading can improve speech perception, especially in situations when auditory input is degraded. More extremely, lip reading can also change the auditory speech percept, as is shown in the McGurk effect (McGurk and MacDonald, 1976). Improvement of speech perception by simultaneous presentation of print has

been demonstrated at the level of words and syllables (Massaro *et al.*, 1988). Dijkstra *et al.* (1989) reported facilitation and inhibition effects on auditorily presented phoneme identity decisions by congruent and incongruent letter primes, respectively suggesting activation phoneme representations by letters.

A neural mechanism for the integration of audiovisual speech has been suggested by Calvert and colleagues (Calvert *et al.*, 1999, 2000) and supported by other neuroimaging findings on audiovisual speech perception (Sams *et al.*, 1991; Sekiyama *et al.*, 2003; Wright *et al.*, 2003) and lip reading (Calvert *et al.*, 1997; Calvert and Campbell, 2003; Paulesu *et al.*, 2003). Results of these studies suggest that the perceptual gain experienced when perceiving multimodal speech is 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 integrates visual and auditory speech information and modulates the modality-specific cortices by back projections (Calvert *et al.*, 1999, 2000). Modality-specific regions involved in this mechanism are 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 is less consistent between the different studies. Interestingly, the involvement of the left posterior STS in the integration of auditory and visual nonlinguistic information has also been reported recently (Calvert *et al.*, 2001; Beauchamp *et al.*, 2004). These results suggest that the STS has a more general role in the integration of cross-modal identity information. Therefore, the aim of the present study was to study the human brain's audiovisual integration mechanisms for letters, i.e., for stimuli that have been previously associated through learning.

**MATERIALS AND METHODS**

**Subjects:** In this study, subjects were healthy and had normal hearing and vision (self reported). Fourteen adult, native speakers of Korean (7 males; 7 females), were participated in the ERP experiment. Age mean was 25.2 (±3.2) and years of formal education was 18.2 (±2.2). No subject has any previous exposure to Chinese or for that matter any other tone languages. None of the subjects had more than three years of formal musical training and none had any musical training within the past five years. All subjects were paid for their participation. They gave informed consent in compliance with a protocol approved

by the Institutional Review Board of the Seoul National University Hospital, Korea.

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

- Congruent/yi<sup>1</sup>/(acoustic/yi<sup>1</sup>+visual/yi<sup>1</sup>/),
- Congruent/yi<sup>2</sup>/(acoustic/yi<sup>2</sup>+visual/yi<sup>2</sup>/),
- Incongruent/yi<sup>1</sup>/(acoustic/yi<sup>1</sup>+visual/yi<sup>2</sup>/) and
- Congruent/yi<sup>4</sup>/(acoustic/yi<sup>4</sup>+visual/yi<sup>4</sup>/) (Fig. 1, 2)

Audiovisual paradigm (meaningful stimuli)				
	Standard	Congruent deviant	Incongruent deviant	Target
Visual	衣	姨	姨	易
Acoustic	/yi <sup>1</sup> /	/yi <sup>2</sup> /	/yi <sup>1</sup> /	/yi <sup>4</sup> /
	300 events (60%)	75 events (15%)	75 events (15%)	50 events (10%)

Fig. 1: The meaningful audiovisual experiment included four stimuli. Standard: Congruent/yi<sup>1</sup>/(acoustic/yi<sup>1</sup>+visual/yi<sup>1</sup>/); Deviant: Congruent/yi<sup>2</sup>/(acoustic/yi<sup>2</sup>+visual/yi<sup>2</sup>/); Deviant: Incongruent/yi<sup>1</sup>/(acoustic/yi<sup>1</sup>+visual/yi<sup>2</sup>/) and Target: Congruent/yi<sup>4</sup>/(acoustic/yi<sup>4</sup>+visual/yi<sup>4</sup>/)

Audiovisual paradigm (meaningless stimuli)				
	Standard	Congruent deviant	Incongruent deviant	Target
Visual	/●/	/▲/	/●/	/◆/
Acoustic	/□□ <sup>1</sup> /	/□□ <sup>2</sup> /	/□□ <sup>1</sup> /	/□□ <sup>4</sup> /
	300 events (60%)	75 events (15%)	75 events (15%)	50 events (10%)

Fig. 2: The meaningless audiovisual experiment included four stimuli, Standard: Congruent/yi<sup>1</sup>/(acoustic/□□<sup>1</sup>+visual/●/), Deviant: Congruent/yi<sup>2</sup>/(acoustic/□□<sup>2</sup>+visual/▲/), Deviant: Incongruent/yi<sup>1</sup>/(acoustic/□□<sup>1</sup>+visual/●/) and Target: Congruent/□□<sup>4</sup>/(acoustic/yi<sup>4</sup>+visual/◆/)

**Stimulus presentation:** 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 (60%) congruent/yi<sup>1</sup>/stimuli and congruent (15%) and incongruent (15%)/yi<sup>2</sup>/stimuli. Congruent/yi<sup>4</sup>/stimuli were presented as target (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). Acoustic stimuli were delivered binaurally to the subjects through plastic tubes and earpieces. 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. 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.

**Experiment:** Each experiment 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 1000 msec in every condition. 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 administered to ensure that the subjects understood the task.

**Event-Related Potential (ERP) Recordings:** EEG data were collected in an electrically and acoustically shield room. 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 electrodes were at mastoids. The signals were bandpass filtered at 0.05-100 Hz and digitized at 1000 Hz. The impedance of the electrodes was below 5 k $\Omega$ . Eye movements were monitored with two Electrooculogram (EOG) electrodes. Four electrodes monitored horizontal and vertical eye movements for off-line artifact rejection. Vertical and horizontal EOG was recorded by electrodes situated above and below the left eye and on the outer canthi of both eyes, respectively. Epochs with EEG or EOG with a large (>100  $\mu$ V) amplitudes were automatically rejected. The artifact-free epochs were filtered at 0.1-15 Hz, baseline corrected and averaged.

**Data analysis:** After the data recordings, the EEG was segmented into 1000 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  $\pm 100 \mu$ V at any EEG channels were automatically excluded from the averaging. The epoch was separately averaged for the standard, deviant and the target stimuli. 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. The N1 that elicited at approximately 100 msec after the onset of auditory stimulus was visually inspected from waveform of standard and deviant stimuli. Cross-modal interaction was identified as the peak voltage between 100-250 msec after stimulus onset.

**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 (Wackermann *et al.*, 1993). Centroids are 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:** 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 (Lehmann and Skrandies, 1980, 1986; Brandeis and Lehmann, 1986; Lehmann, 1987) were used. These methods identify time periods during which there is a quasi-stable map configuration or "landscape" in the following way: Two spatial windows are set around the location of the positive and the negative centroid of the first map. The centroids of the next map are read in and

assigned to the appropriate "negative" and "positive" window. Both windows are then shifted to accommodate the new centroid locations while still including all the earlier ones. If this is possible, the accommodation procedure is repeated with the next pair of centroids. If a centroid cannot be accommodated within its window unless an earlier one has to be left out, the segment is terminated and the next segment is initiated.

**Assessment of map landscape:** 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 (Wackermann *et al.*, 1993). 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 that included the peak against the -100-0 msec pre-stimulus base line. Five sites were prefrontal line, frontal line, central line, parietal line and occipital line, respectively. 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**

Figure 3 and 4 show the mean locations of the landscape centroids of the 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 ( $p < 0.03$ ) and the negative (anterior) centroid was more posterior ( $p < 0.07$ ). There were no statistically relevant differences along the left-right axis (Table 1).

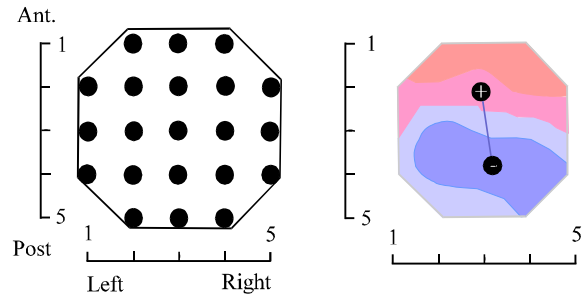


Fig. 3: The locations of the mean (across subjects) map centroids at the peak latencies for the meaningful stimuli per AV integration experimental condition, anterior (Ant.) centroids are negative (-), posterior (Post) centroids are positive (+)

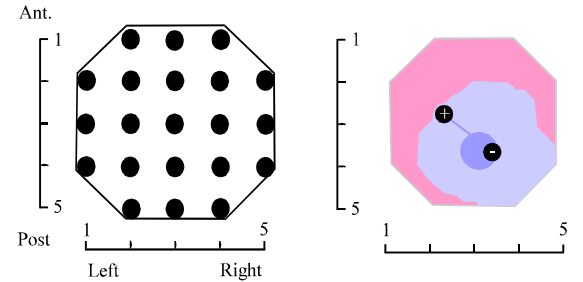


Fig. 4: The locations of the mean (across subjects) map centroids at the peak latencies for the meaningless stimuli per AV integration experimental condition, Anterior (Ant.) centroids are negative (-), posterior (Post) centroids are positive (+)

Table 1: Mean amplitude ( $\mu V$ ) location of the landscape centroid for both meaningful and meaningless stimuli

Stimuli	Centroid		p-value
	Anterior	Posterior	
Meaningful	-1.92	-1.42	<0.03
Meaningless	-1.02	-0.85	<0.07

Figure 5 and 6 illustrate maps series of potential distributions evoked by both meaningful and meaningless stimuli and they are obvious that both the configuration and strength of the fields change over time: The evoked fields are strong between 91 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 116 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.

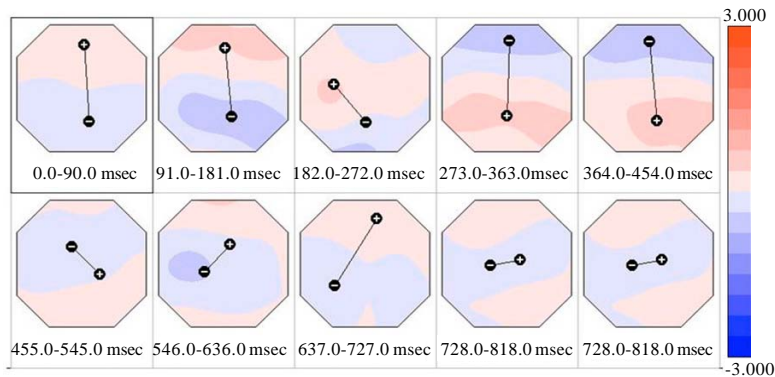


Fig. 5: Series of scalp potential fields of global field power (GFP) as function of time from 0 and 900 msec with intervals of 90 msec evoked by the meaningful stimuli are displayed per the AV integration experimental condition. Occurrence time of maximal GFP from 91 and 181 msec. Equipotential intensity lines in steps of 3  $\mu$ V, for color-coded amplitude values in the maps, refer to color scale

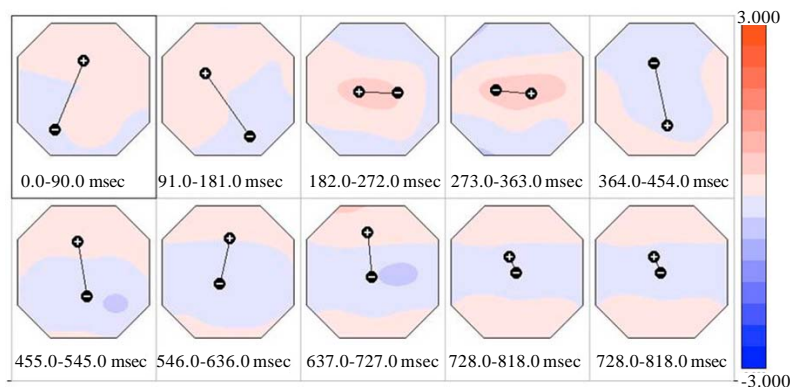


Fig. 6: Series of scalp potential fields of global field power (GFP) as function of time from 0 and 900 msec with intervals of 90 msec evoked by the meaningless stimuli are displayed per the AV integration experimental condition, occurrence time of maximal GFP from 91 and 181 msec, equipotential intensity lines in steps of 3  $\mu$ V. For color-coded amplitude values in the maps, refer to color scale

### DISCUSSION

The present study was able to find evidence of both non-phonetic and phonetic audiovisual interactions in the ERPs to the same AV stimuli. In addition, the differences in the ERPs to the meaningful and meaningless of AV stimuli probably reflect multisensory interactions in phonetic processing. When acoustic and visual phonemes were meaningful, they formed a natural multisensory interaction stimulus. A similar auditory-visual interaction was observed at occipital sites by Giard and Peronnet (1999) at 155-220 msec, which they interpreted as a modulation of the visual evoked N1 wave. This effect does indeed appear to represent an influence of auditory input on processing in a predominantly visual cortical area. The second major deflection indicative of

cross-modal interaction peaked at 220-250 msec and could be accounted for by a dipole pair in anterior temporal peri-sylvian cortex. This effect might represent an interaction in auditory association cortex or in Polymodal cortex of the superior temporal plane (Calvert *et al.*, 2000).

The present results also corroborate previous findings showing that visual speech has access to the early levels of auditory processing hierarchy (Sams *et al.*, 1991; Calvert *et al.*, 1997; Mottonen *et al.*, 2002, 2004; Klucharev *et al.*, 2003; Besle *et al.*, 2004; Van Wassenhove *et al.*, 2005) and support the auditory integration models. Electrophysiological studies in monkeys suggests, that auditory cortex responses to visual stimuli are due to projections from higher cortical regions (Schroeder and Foxe, 2002; Schroeder *et al.*, 2003).

EEG studies show that the auditory N100 amplitude is suppressed during audiovisual speech stimulation in comparison to the sum of unimodal responses (Klucharev *et al.*, 2003; Besle *et al.*, 2004; Van Wassenhove *et al.*, 2005). These studies indicate that in terms of processing time there are early (within 100 msec 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 msec from stimulus onset) (Schroeder and Foxe, 2002; Schroeder *et al.*, 2003). 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 STG from STS or other multisensory areas (Sams *et al.*, 1991; Mottonen *et al.*, 2002, 2004). 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 (Calvert *et al.*, 2001; Klucharev *et al.*, 2003).

### CONCLUSION

The current study demonstrates that the audiovisual interaction is an indicator for investigating the automatic processing of suprasegmental information in tonal language. The use of multiple language groups is important for showing language-related differences in the relative importance of perceptual dimensions that may influence the magnitude of the response to pitch contours. Multisensory integration of letters (orthography) and speech sounds of tonal language in the human auditory association cortex 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

Beauchamp, M.S., K.E. Lee, B.D. Argall and A. Martin, 2004. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, 41: 809-823.

- Besle, J., A. Fort, C. Delpuech and M.H. Giard, 2004. Bimodal speech: Early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.*, 20: 2225-2234.
- Blomert, L., 2002. Dyslexie: Stand van Zaken. [Dyslexia: State of Affairs in The Netherlands]. Report for the Dutch Ministry of Health. In *Dyslexie naar een vergoedingsregeling*, (Amstelveen: Dutch Health Care Insurance Board).
- Brandeis, D. and D. Lehmann, 1986. Event-related potentials of the brain and cognitive processes: Approaches and applications. *Neuropsychologia*, 24: 151-168.
- Calvert, G.A. and R. Campbell, 2003. Reading speech from still and moving faces: The neural substrates of visible speech. *J. Cognitive Neurosci.*, 15: 57-70.
- Calvert, G.A., E.T. Bullmore, M.J. Brammer, R. Campbell and S.C.R. Williams *et al.*, 1997. Activation of auditory cortex during silent lipreading. *Science*, 276: 593-596.
- Calvert, G.A., M.J. Brammer, E.T. Bullmore, R. Campbell, S.D. Iversen and A.S. David, 1999. Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport*, 10: 2619-2623.
- Calvert, G.A., P.C. Hansen, S.D. Iversen and M.J. Brammer, 2001. Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage*, 14: 427-438.
- Calvert, G.A., R. Campbell and M.J. Brammer, 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.*, 10: 649-657.
- Dijkstra, A., R. Schreuder and U.H. Frauenfelder, 1989. Grapheme context effects on phonemic processing. *Lang. Speech*, 32: 89-108.
- Frith, U., 1985. Beneath the Surface of Developmental Dyslexia. In: *Surface Dyslexia*, Patterson, K.E., J.C. Marshall and M. Coltheart (Eds.). Routledge and Kegan-Paul, London, UK., pp: 301-330.
- Giard, M.H. and F. Peronnet, 1999. Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J. Cognitive Neurosci.*, 1: 473-490.
- Klucharev, V., R. Mottonen and M. Sams, 2003. Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Res.*, 18: 65-75.

- Lehmann, D. and W. Skrandies, 1980. Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography Clin. Neurophysiol.*, 48: 609-621.
- Lehmann, D. and W. Skrandies, 1986. Segmentation of Evoked Potentials Based on Spatial Field Configuration in Multichannel Recordings. In: *Cerebral Psychophysiology: Studies in Event-Related Potentials*, McCallum, W.C., R. Zappoli and F. Denoth (Eds.). Vol. 38, Elsevier, New York, pp: 506-512.
- Lehmann, D., 1987. Principles of Spatial Analysis. In: *Methods of Analysis of Brain Electrical and Magnetic Signals*, Gevins, A.S. and A. Remond (Eds.). Vol. 1. Elsevier, Amsterdam, The Netherlands, ISBN-13: 9780444808042, pp: 309-354
- Lieberman, A.M., 1992. The Relation of Speech to Reading and Writing. In: *Orthography, Phonology, Morphology and Meaning*, Frost, R. and L. Katz (Eds.). Elsevier Science Publishes B.V., Amsterdam, The Netherlands, pp: 167-178.
- Massaro, D.W., M.W. Cohen and L.A. Thompson, 1988. Visible language in speech perception: Lipreading and reading. *Visible Lang.*, 22: 9-31.
- McGurk, H. and J. MacDonald, 1976. Hearing lips and seeing voices. *Nat.*, 264: 746-748.
- Mottronen, R., C.M. Krause, K. Tiippana and M. Sams, 2002. Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Res.*, 13: 417-425.
- Mottronen, R., M. Schurmann and M. Sams, 2004. Time course of multisensory interactions during audiovisual speech perception in humans: A magnetoencephalographic study. *Neurosci. Lett.*, 363: 112-115.
- National Reading Council, 1998. *Preventing Reading Difficulties in Young Children*. National Academy Press, Washington, DC.
- Paulesu, E., D. Perani, V. Blasi, G. Silani and N.A. Borghese *et al.*, 2003. A functional-anatomical model for lipreading. *J. Neurophysiol.*, 90: 2005-2013.
- Perfetti, C.A. and R. Sandak, 2000. Reading optimally builds on spoken language: Implications for deaf readers. *J. Deaf Stud. Deaf Educ.*, 5: 32-50.
- Raij, T., K. Uutela and R. Hari, 2000. Audiovisual integration of letters in the human brain. *Neuron*, 28: 617-627.
- Sams, M., R. Aulanko, M. Hamalainen, R. Haru, O.V. Lounasmaa, S.T. Lu and J. Simola, 1991. Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neurosci. Lett.*, 127: 141-145.
- Schroeder, C.E. and J.J. Foxe, 2002. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cogn. Brain Res.*, 14: 187-198.
- Schroeder, C.E., J. Smiley, K.G. Fu, T. McGinnis, M.N. O'Connell and T.A. Hackett, 2003. Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *Int. J. Psychophysiol.*, 50: 5-17.
- Sekiyama, K., I. Kanno, S. Miura and Y. Sugita, 2003. Auditory-visual speech perception examined by fMRI and PET. *Neurosci. Res.*, 47: 277-287.
- Sumby, W.H. and I. Pollack, 1954. Visual contribution to speech intelligibility in noise. *J. Acoust. Sci. Am.*, 26: 212-215.
- Van Wassenhove, V., K.W. Grant and D. Poeppel, 2005. Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci.*, 102: 1181-1186.
- Wackermann, J., D. Lehmann, C.M. Michel and W.K. Strik, 1993. Adaptive segmentation of spontaneous EEG map series into spatially defined microstates. *Int. J. Psychophysiol.*, 14: 269-283.
- Wright, T.M., K.A. Pelphey, T. Allison, M.J. McKeown and G. McCarthy, 2003. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb. Cortex*, 13: 1034-1043.