Cortical Operational Synchrony during Audio-Visual Speech Integration

Andrew A. Fingelkurts[1,2*], Alexander A. Fingelkurts[1,2], Christina M. Krause[3], Riikka Möttönen[4], and Mikko Sams[4]

1) Human Brain Research Group, Human Physiology Department, Moscow State University, 119899 Moscow, Russian Federation
2) BM-Science Brain & Mind Technologies Research Centre, P.O. Box 77, FI-02601 Espoo, Finland
3) Cognitive Science / Department of Psychology, University of Helsinki, P.B. 9, 00014 University of Helsinki, Finland
4) Laboratory of Computational Engineering, Helsinki University of Technology, 02015 HUT, Finland

Abstract

Information from different sensory modalities is processed in different cortical regions. However, our daily perception is based on the overall impression resulting from the integration of information from multiple sensory modalities. At present it is not known how the human brain integrates information from different modalities into a unified percept. Using a robust phenomenon known as the McGurk effect it was shown in the present study that audio-visual synthesis takes place within a distributed and dynamic cortical networks with emergent properties. Various cortical sites within these networks interact with each other by means of so-called operational synchrony (Kaplan et al., 1997). The temporal synchronization of cortical operations processing unimodal stimuli at different cortical sites reveals the importance of the temporal features of auditory and visual stimuli for audio-visual speech integration.

Keywords: multisensory integration, crossmodal, audio-visual, synchronization, operations, large-scale networks, MEG.

INTRODUCTION

People usually perceive the external world as a seamless whole. Our perception of the external world depends on the integration of information from different senses (Driver & Spencer, 1998). When and where in the human brain the integration of such multisensory information occurs is not yet known (Giard & Peronnet, 1999). The human brain cannot be considered a passive, stimulus-driven device or a passive transformer (see reviews, Erdi, 2000; Engel, Fries, & Singer, 2001), but rather as an extraordinary integrative organ, which not only perceives but also creates new realities (Nunez, 2000; Erdi, 2000).

The issue concerning perceptual integration within separate sensory systems has been widely investigated both in the visual modality (Singer & Gray, 1995; Treisman, 1996; Zeki, 2001) and in the auditory modality (Loveless et al., 1996, Näätänen & Winkler, 1999). Inputs from different sensory modalities are processed in different cortical regions, but our daily perception is based on the global multisensory percept resulting from the integration of information from various sensory modalities (Driver & Spencer, 1998; Giard & Peronnet, 1999). Indeed, the integration of information from different sensory modalities is clearly beneficial: multimodal events are detected more accurately and faster than unimodal events (Frens, Vanopstal, & Vanderwilligen, 1995; Calvert, 2001). Human speech is a prime example of this.

For example, for individuals with impaired hearing, lip-reading can supplement the auditory signal and enhance its intelligibility (Rosenblum & Saldana, 1996). Visual speech cues are also used by individuals with normal hearing in a noisy environment (MacLeod & Summerfield, 1987) or in recovering a difficult message (Reisberg, McLean, & Goldfield, 1987). One example of audio-visual speech integration is provided by a robust illusion known as the McGurk effect (McGurk & MacDonald, 1976). In this effect normal listeners report *hearing* audio-visually fusion syllables as some combination of the auditory and visual syllables (e.g., auditory /ba/ + visual /ga/ are perceived as /va/) or as a syllable dominated by the visual syllable (e.g., auditory /ba/ + visual /va/ are perceived as /va/). The vast majority of people (but not all) experience the McGurk illusion.

Although audio-visual speech integration is well-established experimentally (Massaro, 1987; Massaro & Cohen, 1996; Rosenblum, Yakel, & Green, 2000), the brain (neural) processes that subserve it remain to be assessed (Giard & Peronnet, 1999). Most research in humans only demonstrates the existence of the phenomenon rather than

reveals the physiological processes underlying it (for a review, see O'Hare, 1991). However, recently, some studies on multisensory integration have focused on the underlying mechanisms. For example, in behavioral studies the main finding has been that reaction times to congruent audio-visual stimuli are typically shorter than to their unimodal counterparts (Miller, 1982; Frens, Vanopstal, & Vanderwilligen, 1995). Incongruent stimuli have the opposite effect, slowing response times (Stein et al., 1989; Lewkowicz, 1996) and producing perceptual anomalies (McGurk & MacDonald, 1976). Some neuro-imaging (Lewis, Beauchamp, & DeYoe, 2000; Macaluso, Frith, & Driver, 2000; Calvert et al., 2001) and electro-magnetic (Sams & Imada, 1997; Giard & Peronnet, 1999; Krause et al., 2001; Möttönen et al, 2002) studies also have shed some light on the neural processes underlying audio-visual integration.

In animal studies, detailed observation of the behavior of multisensory neurons at the single neuron level has been resulted in four "integration rules" (for a review, see Stein & Meredith, 1993). The central rule is that of temporal coincidence, according to which the greatest integration effects are obtained if inputs are temporarily synchronized (Stein, Meredith, & Wallace, 1994). This rule could also be applied to the large-scale cortical level of the human brain. It is possible to plausibly argue that the crossmodal binding in the human brain may be achieved by the synchronized processing of sensory inputs between the unimodal cortical areas (see reviews, Philips & Singer, 1997; Salinas & Sejnowski, 2001), rather than in so-called convergence regions of the cortex. Indeed, extensive analysis of the lesion studies has found that none of the structures known to receive converging input from more than one sensory system has been shown to be specifically crucial for both the development and display of crossmodal performance (Ettlinger & Wilson, 1990; see also Murray, Malkova, & Goulet, 1998). Instead, synchrony generated intrinsically by functional interactions between distant cortical areas might be the mechanism underlying multisensory integration (see reviews, Ettlinger & Wilson, 1990; Engel, Fries, & Singer, 2001).

The idea of large-scale cortical networks, where elementary operations are localized in discrete cortical and subcortical regions, and complex functions involve parallel (or synchronous) processing in a wide-spread network, is a highly promising concept in the modern neural theories of cognition (see reviews, Nunez, 1995, 2000; Bressler & Kelso 2001).

In this respect the methodology of brain operational activity was developed (see reviews, Kaplan & Shishkin, 2000; Fingelkurts & Fingelkurts, 2001). In the framework

of this methodology it is possible to extract from EEG/MEG recordings information about the discrete brain operations and estimate the level of inherent synchrony of these operations appearing simultaneously in different cortical areas. This type of synchronization has been named "*Operational Synchrony*" (Kaplan et al., 1997). At the EEG/MEG level such operations are reflected in the form of quasi-stationary segments in corresponding locations/sites (Kaplan & Shishkin, 2000). It was shown that the segment sequences in different cortical locations are synchronized, forming short-term metastable topological combinations underling mental states (Fingelkurts & Fingelkurts, 2001).

Although there is evidence that crossmodal processing in the cerebral cortex may underlie a phenomenon referred to as "multisensory integration" (Calvert, 2001), none of the known studies have explicitly investigated *cortical functional interactions* during audio-visual speech integration. In order to address this question operational synchrony analysis (Fingelkurts & Fingelkurts, 2001) was conducted to detect those cortical brain areas which actively interact between each other during audio-visual speech perception using both congruent and incongruent (McGurk-type) audio-visual stimuli. If one assumes that some aspects of language processing involves cortical networks, and if one assumes that a cortical network processing some aspect of language can be detected by observing synchronous brain activity in different cortical regions, then one could ask subjects to perform a task that is likely to involve a cortical network. The stimuli which produce the McGurk effect are suitable candidates for these kinds of language processing. In the present work we examined two frequency bands: alpha (7-13Hz) and beta (15-21Hz). These brain oscillations seem to respond to the perception of audio-visual speech information, as was observed in the previous analysis of the same data (Krause et al., 2001). We analyzed the *rapid transition processes* – RTP (which are the markers of boundaries between quasi-stationary segments) for each local MEG location. Also the temporal synchronization of RTPs between different MEG locations was used as a measure of functional interaction between cortical areas. This synchronization corresponds to *operational synchrony* (OS) process (see Method, section Calculation of Operational Synchrony Index). Data was obtained during three experiments: audio-visual, visual and auditory.

We hypothesized that audio-visual speech perception would result in the emergence of a new dynamic large-scale cortical network (involving synchronous operations at different brain oscillatory frequency bands), which would not consist simply of a summation of the separate unimodal audio and visual cortical networks.

METHOD

Subjects

Ten Finnish-speaking volunteer subjects (three females, mean age of all subjects was 28) participated in the audio-visual experiment. All subjects had normal hearing, vision and were right-handed except one female (self-reported). The data from one male subject was excluded from further analysis because of extensive artifacts in the recordings. All of the nine subjects participated in the audio-visual experiment. In the audio-visual experiment congruent (auditory /iti/ + visual /iti/) and incongruent (auditory /ipi/ + visual /iti/) audio-visual stimuli were presented. Seven subjects perceived the incongruent stimuli always as "iti" (indicating that these subjects had the McGurk effect). Two subjects reported "ipi", when incongruent stimuli were presented (indicating that these subjects did not have the McGurk effect). The seven subjects with the McGurk effect also completed the experiment in the visual modality, and three of them completed additionally the experiment in the auditory modality.

Stimuli

Meaningless disyllables (vowel-consonant-vowel) uttered by a female Finnish speaker were recorded in a chamber with a professional video camera. Visual clips (frame rate 25 Hz) and sound files (digitized at 22 050 Hz) were extracted from the digital video for each stimulus utterance ("ipi", "iti", and "ivi"). The duration of the visual utterances was about 900 ms. The duration of acoustic /ipi/ was 588 ms and that of acoustic /iti/ was 581 ms. The mouth opening in the visual /ipi/ and /iti/ stimuli began 230 ms prior to the start of the acoustic utterance, when presented together. The total lengths of acoustic and visual stimulus files were 1600 ms including the periods of silence, or where the face had a closed mouth before and after the utterances. The audio-visual experiment included four stimuli: congruent "ipi" (auditory /ipi/ + visual /ipi/), congruent "iti" (auditory /iti/ + visual /iti/), incongruent "iti" (auditory /ipi/ + visual /iti/) and congruent "ivi" (auditory /ivi/ + visual /ivi/). The visual experiment contained only the visual parts of these stimuli and auditory experiment contained only the auditory parts.

Stimulus Presentation

The stimulus sequences were presented to the subjects with the "Presentation" software (Neurobehavioral Systems, Inc, 2001). The audio-visual stimuli consisted of frequent (85%) *standard* congruent "ipi" stimuli and infrequent *deviant* congruent (5%) and *deviant* incongruent (5%) "iti" stimuli. The terms "standard" and "deviant" are conventionally used in the mismatch negativity and oddball paradigms to refer to the "frequent" and "infrequent" stimuli respectively (Näätänen & Winkler, 1999). Deviant congruent "ivi" stimuli were presented as targets (5%) which the subjects were instructed silently to count during the registration in order to check that the subjects were consciously attentive to the stimuli. The auditory stimuli were delivered binaurally to the subjects through plastic tubes and earpieces. The intensity of the sound was adjusted to 55 dB above the subject's hearing threshold (defined for the audio-visual stimulus sequence). The visual stimuli were projected into the measurement room through a data projector. The height of the face stimulus was 12 cm and its distance from the subject was 105 cm.

In the unimodal-stimuli audio-only and visual-only experiments, the visual stimuli and the audio stimuli were not presented respectively. However, these experiments were in all other aspects identical to the bimodal-stimuli audio-visual experiment.

Procedure

The audio-visual experiment consisted of 3-4 sessions each lasting between 15 and 20 min. The subjects were instructed to concentrate on the stimuli and silently count the number of "ivi" utterances. After each session the subjects were asked to report the result of their counting. In order to assess how the subjects perceived the incongruent audio-visual (McGurk-type) utterances, a behavioral test was carried out during one of the breaks between the experimental sessions. In this test a sequence consisting of 12 incongruent deviants, 6 congruent deviants, 12 targets and 94 standards was presented. The subjects were instructed to repeat each utterance aloud immediately after identifying what they *heard*. The experimenter wrote down the responses. Seven subjects perceived the incongruent deviants always as "iti" (demonstrating that these subjects had the McGurk effect). Two subjects always reported "ipi", when incongruent deviants were presented (demonstrating that these subjects did not have the McGurk effect).

The visual and auditory experiment consisted of two sessions each lasting 15-20 min. The task was to count silently the "ivi" utterances and to report the result of counting after each session. There was always an interval of at least one week between the audio-visual, visual and auditory experiments.

MEG recording

The magnetoencephalogram (MEG) was recorded continuously in a magnetically shielded room with a 306-channel whole-head device in the Low Temperature Laboratory at the Helsinki University of Technology (Neuromag Vectorview, Helsinki, Finland). The sensor elements of the device comprise two orthogonal planar gradiometers and one magnetometer.

Before each experiment the positions of four marker coils placed on the scalp were determined in relation to three anatomical landmark points (the nasion and both preauricular points) using an Isotrak 3D-digitizer. Measuring the corresponding magnetic fields current through the coils determined the coil locations in the magnetometer coordinate system. The position of the head was measured at the beginning of each session. The data was digitized at 300 Hz. The passband filter of the MEG recordings was 0.06-100 Hz. About 100 responses of the subjects to each deviant stimulus and about 2000 responses to standard stimuli were collected. Epochs containing large-amplitude artifacts on MEG or EOG channels were automatically rejected. Also, the presence of an adequate MEG signal was determined by checking visually the raw signal on the computer screen.

Data Analysis

In all the experiments (audio-visual, auditory and visual), the MEG data was divided into data-segments (the duration was 840 ms); post-standard, post-deviant-congruent or post-deviant-incongruent data intervals with respect to the types of stimulus presentations. Thereafter, the data-segments for each stimulus type were "glued" together. The full data-stream was given simultaneously to three different virtual extraction units (see below).

In the present study we examined the post-stimulus MEG data (still face, no sound), which is assumed not to be influenced by any artifact of the stimulus-events themselves (Figure 1).
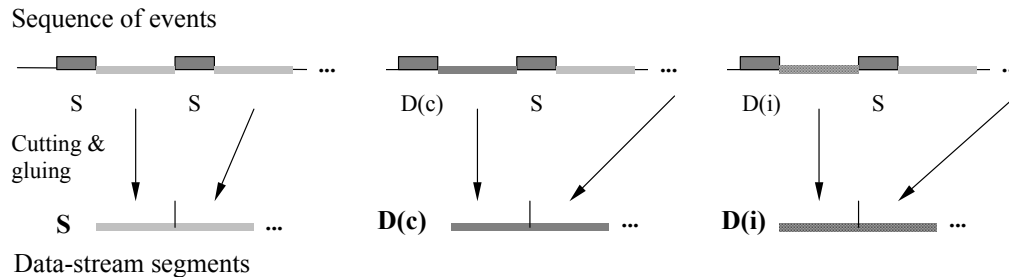
Sequence of events

Data-stream segments

Figure 1. The scheme of the data processing. Extractions of the correspondent post-stimulus intervals (still face, no sound) were done separately for each subject, each MEG location (gradiometer 1). S – standard stimuli, D(c) – deviant-congruent stimuli, D(i) – deviant-incongruent stimuli.

The output of this procedure was a sequence of concatenated data. In order to eliminate any possible short-term non-stationarities in the neighborhood of the connection point, the data of these areas was smoothed. According to modeling calculations, a number of ±3 data points around the connection point (Dt=25 ms) were chosen to symmetrically average the data in these areas.

Thus, the full MEG data-streams were split into three distinct segments: S for standard stimuli, D(c) for deviant congruent audio-visual stimuli, and D(i) for deviant incongruent audio-visual stimuli (Figure 1). In the auditory and visual experiments, only data from the standard (S) and deviant (D) segments was present.

Due to the technical requirements of the tools used later to process the data, 20 MEG locations which correspond to the International 10-20 System of EEG electrode placement ($F_{7/8}$, F_z, $F_{3/4}$, $T_{3/4}$, $C_{5/6}$, C_z, $C_{3/4}$, $T_{5/6}$, P_z, $P_{3/4}$, O_z, $O_{1/2}$) were analyzed with a converted sampling rate of 128 Hz.

Prior to the non-parametric adaptive segmentation procedure, each MEG data sequence (corresponding to different stimulus conditions: S, D(c), and D(i)) was bandpass filtered in the alpha (7-13 Hz) and beta (15-21 Hz) frequency ranges after which the amplitudes of the samples were squared. These frequency bands were chosen because the previous study of the same data showed that brain oscillations at alpha and beta frequency bands seem to respond to the perception of audio-visual speech information (Krause et al., 2001). The filtering procedure was done systematically for

one and the same first gradiometer ($\partial Bz/\partial x$) from each MEG sensor. This gradiometer was chosen because its MEG signal was systematically the biggest in all analyzed sensors.

Nonparametric Adaptive Level Segmentation of MEG-recordings

It has been suggested that an observed piecewise stationary process like an MEG or EEG can be considered as being "glued" from several segments of random stationary processes with different probabilistic characteristics (Kaplan & Shishkin, 2000). The transitions from one segment to another mark the moment in time when the activity in the neuronal network switches. Within the framework of this methodology quasi-stationary segments in an MEG or EEG signal reflect discrete brain operations (Fingelkurts & Fingelkurts, 2001). Thus, the aim of the task was to divide the MEG-signal into stationary segments by estimating such intrinsic points of "gluing". These instants within short-time window, when the MEG amplitude significantly changed, were identified as **rapid transition processes** (RTP) (Kaplan et al., 1997) and these RTPs thus mark the boundaries between quasi-stationary segments.

In order to estimate these RTPs, comparisons were made between the ongoing MEG amplitude absolute values averaged in the test window (6 points=39 ms) and the MEG amplitude absolute values averaged in the level window (120 points=930 ms). These values revealed the optimal means for identifying segments from the signal (according to a previous study). The use of short-time windows was motivated by the need to track non-stationary transient cortical processes on a sub-second timescale. The method ("SECTION" software, Moscow State University) is based on the automatic selection of level-conditions in accordance with a given level of probability of "false alerts" and carrying out simultaneous screening of multi-channel MEG. If the absolute maximum of the averaged amplitude values in the test window is less or equal to the averaged amplitude values in the level window, then the hypothesis of MEG homogeneity is accepted. Otherwise, if the absolute maximum of the averaged amplitude values in the test window exceeds the averaged amplitude values in the level window, according to the threshold of the false alerts (the Student criteria, $p<0.05$ with coefficient 0.3), its time instant becomes the preliminary estimate of a RTP. Also another condition must be fulfilled in order to eliminate the "false alerts" associated with possible anomalous peaks

in the amplitude. The five points of the digitized MEG following this preliminary RTP must have a statistically significant difference between averaged amplitude values in the test and the level windows (the Student criteria, $p<0.05$ with coefficient 0.1). If these two criteria are met, then the preliminary RTP are assumed as actual. Then each of the windows shifts on one data-point from the actual RTP and the procedure is repeated. With this technique, the sequence of RTPs with statistically proven ($p<0.05$, Student t-test) time coordinates has been determined for each MEG location. The details of methodology and theoretical concepts are described elsewhere (Kaplan & Shishkin, 2000; Fingelkurts & Fingelkurts, 2001).

Calculation of Operational Synchrony Index

Thereafter, the synchronization of rapid transition processes (RTP) (index of operational synchrony) was estimated. This procedure ("JUMPSYN"-software, Moscow State University) reveals the functional interrelationships between cortical sites, different from those measured using correlation, coherence and phase analysis (Kaplan & Shishkin, 2000). Each RTP in the reference MEG location (the location with the minimal number of RTPs from any pair of MEG locations) was surrounded by a "window" (from –3 to +4 digitizing points to each side from RTP point) of 55 ms. Any RTP from another (test) location was thought to be coinciding if it fell within this window. The window of 55 ms provides 70-80% of all RTP synchronizations. The estimation of the **index of operational synchrony** (IOS) for pairs of locations was estimated using this procedure. The IOS was computed as follows:

$$\text{IOS} = \text{m}_{\text{windows}} - \text{m}_{\text{residual}}, \text{ where } m_w = 100 * \frac{sn_w}{sl_w}; \ m_r = 100 * \frac{sn_r}{sl_r};$$

sn_w – total number of RTPs in all windows in the test channel;

sl_w – total length of MEG recording (in data-points) inside all windows in the test channel;

sn_r – total number of RTPs outside the windows in the test channel;

sl_r – total length of MEG recording (in data-points) outside the windows in the test channel.

The IOS tends towards zero where there is no synchronization between the RTPs and has positive or negative values where such synchronization exists. Positive values

indicate "active" coupling of RTPs, whereas negative values mark "active" uncoupling of RTPs.

To arrive at a direct estimation of a *5% level of statistical significance* of the IOS ($p<0.05$), numerical modeling was undertaken (500 independent trials). As a result of these tests the stochastic level of RTPs coupling (IOS_{stoh}), and the upper and lower thresholds of IOS_{stoh} significance were calculated. These values represent an estimation of the maximum (by module) possible stochastic rate of RTPs coupling. Thus, only those values of IOS which exceeded the upper (active synchronization) and lower (active unsynchronization) thresholds of IOS_{stoh} have been assumed to be statistically valid ($p<0.05$). The detailed methodology and theoretical conceptions of RTPs synchronization are described elsewhere (Fingelkurts & Fingelkurts, 2001).

In order to reduce data and select the highest values of IOS (those with the strongest functional connections), an analysis threshold for OS estimation equaling two was chosen. With this threshold:

1) only those connections which exceeded the stochastic upper/lower level of IOS_{stoh} remained, 50% of all connections;

2) randomly coinciding RTPs which may have occurred in the places of smoothing were eliminated.

Separate computer maps of the IOS values were built for each subject and for each MEG under different experimental conditions. The problem of multiple comparisons between maps cannot easily be overcome due to the large number of electrode pairs in the OS maps (Rappelsberger & Petsche, 1988). This problem is common for all studies which require multiple comparisons between maps (Weiss & Rappelsberger, 2000; Razoumnikova, 2000). The comparisons should therefore, be considered descriptive rather than confirmatory (Stein et al., 1999). The changes in maps were only considered relevant if the changes appeared consistently in a majority of the trials and subjects (75-100%) under the same experimental conditions.

BEHAVIORAL RESULTS

In the audio-visual experiment deviant congruent (auditory /iti/ + visual /iti/) and incongruent (auditory /ipi/ + visual /iti/) audio-visual stimuli, both of which were perceived as "iti", were presented amongst standard congruent "ipi" (auditory /ipi/ +

visual /ipi/) stimuli. All subjects (*n*=9) identified correctly the congruent deviants, standards and targets (see section Method for explanation). Seven subjects perceived the incongruent deviants always as "iti" (indicating that these subjects had the McGurk effect). Two subjects reported "ipi" when incongruent deviants were presented (indicating that these subjects did not have the McGurk effect). This led us to consider the "McGurk subjects" and the "non-McGurk subjects" separately for all further analyses.

In the visual experiment, only the visual stimuli were presented, whereas in the auditory experiment only the acoustic stimuli were presented. The "McGurk subjects" (*n*=7 for visual condition and *n*=3 for auditory condition) were able to recognize the visual/auditory utterances. None of the "non-McGurk subjects" participated in the visual or auditory experiments.

NEUROMAGNETIC RESULTS

In the present work we examined two frequency bands: alpha (7-13Hz) and beta (15-21Hz). These brain oscillations seem to respond to the perception of audio-visual speech information, as was observed in the previous pilot analysis of the same data (Krause et al., 2001). We analyzed the *rapid transition processes* – RTP (which are the markers of boundaries between quasi-stationary segments) in each local MEG location. Also the synchronization of RTPs between different MEG locations was estimated. This synchronization corresponds to *operational synchrony* (OS) process (Kaplan et al., 1997), which reflects the functional coupling of different brain areas. Data was obtained during three experiments: audio-visual, visual and auditory.

RTPs in the Audio-Visual, Visual and Auditory Experiments

Table 1 summarizes the results for the number of RTPs obtained in the "McGurk subjects" (*n*=7) for all MEG locations (*n*=20) and presents the corresponding data separately for different stimuli. The number of RTPs in both frequency bands (15-21Hz, 7-13Hz) was on average smaller for the audio-visual deviant congruent (AV(c)) (*p*<0.001, Student *t*-test) and audio-visual deviant incongruent (AV(i)) (*p*<0.001, Student

t-test) stimuli than for audio-visual standard (AV(s)) stimuli (see Table 1). Mathematically, the number of RTPs is negatively correlated with the duration of quasi-stationary segments in MEG signal. This means that the duration of quasi-stationary segments in the MEG signal was on average shorter for AV(s) stimuli than for AV(c) and AV(i) stimuli.

Table 1.
Average number of RTPs for all locations (*n*=20) and all "McGurk subjects" (*n*=7) in different conditions

Hz	Condition	p	Condition	p	Condition	p
	AV (standard) x AV (congruent)		AV (standard) x AV (incongruent)		AV (congruent) x AV (incongruent)	
15-21	287.19 -+ 3.91 x 265.89 -+ 6.34	< 0.001	287.19 -+ 3.91 x 262.83 -+ 7.51	< 0.001	265.89 -+ 6.34 x 262.83 -+ 7.51	> 0.05
7-13	240.88 -+ 3.26 x 233.91 -+ 4.72	< 0.001	240.88 -+ 3.26 x 231.43 -+ 5.21	< 0.001	233.91 -+ 4.72 x 231.43 -+ 5.21	> 0.05
	A (standard) x A (deviant)		V (standard) x V (deviant)			
15-21	282.37 -+ 3.69 x 259.65 -+ 6.02	< 0.001	283.33 -+ 6.37 x 257.5 -+ 5.07	< 0.001		
7-13	238.17 -+ 3.23 x 225.23 -+ 6.1	< 0.001	237.86 -+ 3.91 x 226.45 -+ 7.13	< 0.001		
	AV (standard) x A (standard)		AV (standard) x V (standard)			
15-21	287.19 -+ 3.91 x 282.37 -+ 3.69	< 0.001	287.19 -+ 3.91 x 283.33 -+ 6.37	0.01 < 0.05		
7-13	240.88 -+ 3.26 x 238.17 -+ 3.23	0.01 < 0.05	240.88 -+ 3.26 x 237.86 -+ 3.91	0.01 < 0.05		
	AV (congruent) x A (deviant)		AV (incongruent) x V (deviant)			
15-21	265.89 -+ 6.34 x 259.65 -+ 6.02	< 0.01	262.83 -+ 7.51 x 257.5 -+ 5.07	0.01 < 0.05		
7-13	233.91 -+ 4.72 x 225.23 -+ 6.1	< 0.001	231.43 -+ 5.21 x 226.45 -+ 7.13	0.01 < 0.05		

The number of RTPs was also on average smaller for AV(i) stimuli compared with the AV(c) stimuli in both frequency bands, however, the differences didn't reach a significant level (Table 1, upper right part).

Similar dependencies were found in both of the unimodal experiments. From Table 1 one can see that the number of RTPs in both frequency bands (15-21Hz, 7-13Hz) was on average smaller for the auditory deviant (A(d)) (*p*<0.001, Student *t*-test) and visual deviant (V(d)) (*p*<0.001, Student *t*-test) stimuli than for the auditory standard (A(s)) and visual standard (V(s)) stimuli. This means that the duration of quasi-stationary segments in the MEG signal was on average shorter for A(s) and V(s) stimuli than for A(d) and V(d) stimuli. However, the number of RTPs did not differ between the A(d) and V(d) stimuli (Table 1, compare the first and the second columns). Also there were no differences between the A(s) and V(s) stimuli with respect to the number of RTPs.

The lower part of Table 1 indicates the differences between the RTPs observed during the AV and unimodal experiments. The number of RTPs in both frequency bands (15-21Hz, 7-13Hz) was on average smaller for A(s) (*p*<0.001 and 0.01<*p*<0.05, Student *t*-test), V(s) (0.01<*p*<0.05, Student *t*-test), A(d) (*p*<0.001 and *p*<0.01, Student *t*-test) and V(d) (0.01<*p*<0.05, Student *t*-test) stimuli than for AV(s), AV(c) and AV(i) stimuli respectively. This means that the duration of quasi-stationary segments in the MEG

signal was on average shorter in the audio-visual experiment (for all stimulus types) compared to all the unimodal conditions.

Another question concerns the distribution of RTPs between different frequency bands. Table 2 displays the number of RTPs observed in the "McGurk subjects" (*n*=7) for all MEG locations (*n*=20) and presents the corresponding data separately for different stimuli (AV(s), AV(c), AV(i), A(s), A(d), V(s), and V(d)) and frequency bands (15-21Hz, 7-13Hz).

Table 2.
Average number of RTPs for all locations (*n*=20) and all "McGurk subjects" (*n*=7) within alpha and beta frequency bands

Condition	Hz	P
	15-21 x 7-13	
AV (standard)	287.19 -+ 3.91 x 240.88 -+ 3.26	< 0.001
AV (congruent)	265.89 -+ 6.34 x 233.91 -+ 4.72	< 0.001
AV (incongruent)	262.83 -+ 7.51 x 231.43 -+ 5.21	< 0.001
A (standard)	282.37 -+ 3.69 x 238.17 -+ 3.23	< 0.001
A (deviant)	259.65 -+ 6.02 x 225.23 -+ 6.1	< 0.001
V (standard)	283.33 -+ 6.37 x 237.86 -+ 3.91	< 0.001
V (deviant)	257.5 -+ 5.07 x 226.45 -+ 7.13	< 0.001

Under all experimental conditions the number of RTPs was on average smaller for the alpha frequency (7-13Hz) band compared to the beta frequency (15-21Hz) band (*p*<0.001, Student *t*-test). This means that the duration of the quasi-stationary segments in the MEG signal was on average shorter in the beta than in the alpha frequency band under all experimental conditions.

Operational Synchrony of Cortical Areas during Audio-Visual, Visual and Auditory Experiments

To get an idea of the overall topographical pattern of the main operational synchrony (OS) differences elicited by the different stimuli, schematic brain maps in the alpha band (7-13Hz) were drawn for the AV(s), AV(c) and the AV(i) stimuli (Figure 2). By way of example, the data is shown for one "McGurk subject". The statistically significant values of OS are plotted as lines connecting the involved MEG locations. Widespread networks

of cortical areas were involved during three stimulus presentations (Figure 2). Similar results were obtained for all "McGurk" and "non-McGurk subjects". In order to assess the *principal* process of operational synchrony (OS), all possible pairs of MEG locations exhibiting statistically proven OS were ranged in accordance with their rate of occurrence within all epochs of analysis for each subject and across all subjects. Then only the most frequently found pairs (not less than 75% of occurrence in all epochs and all subjects) were analyzed further.
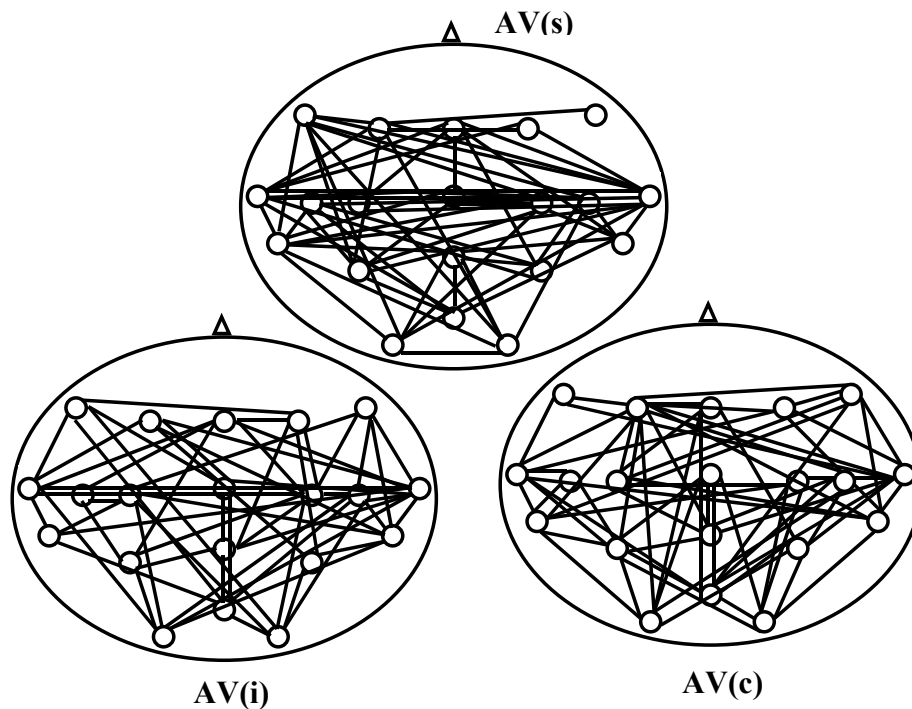

Figure 2. Values of index of operational synchrony (IOS) for audio-visual standard [AV(s)], audio-visual congruent [AV(c)] and audio-visual incongruent [AV(i)] (the McGurk effect) stimuli in the alpha frequency band. The IOS values which exceeded (p < 0.05) the stochastic level of synchronization are mapped onto schematic brain maps as connecting lines between the MEG locations involved.

Interactions During Audio-Visual Experiment (the "McGurk Subjects")

Figure 3 presents the most frequently found brain area connections (indexed by operational synchrony – IOS) in the all "McGurk subjects" for the three stimuli (AV(s), AV(c), AV(i)) in the alpha (7-13Hz) and beta (15-21Hz) frequency bands. The

presentation of these three different stimulus types elicited different cortical networks consisting of operationally synchronized brain areas. The largest networks of OS were found in the beta band for both deviant congruent and deviant incongruent stimuli. Also in the alpha band the richest map of OS was revealed for the incongruent deviant stimuli (Figure 3). In Figure 3 black thin doted line indicates the functional connections which were specific for the deviant audio-visual stimuli (both congruent and incongruent). Black thick solid and doted colors indicate the functional connections specific for the congruent and incongruent stimuli, respectively. Grey color indicates connections which were common for all three stimuli (Figure 3). Most OS connections were found in the left brain hemisphere, and bilaterally in the temporal regions.
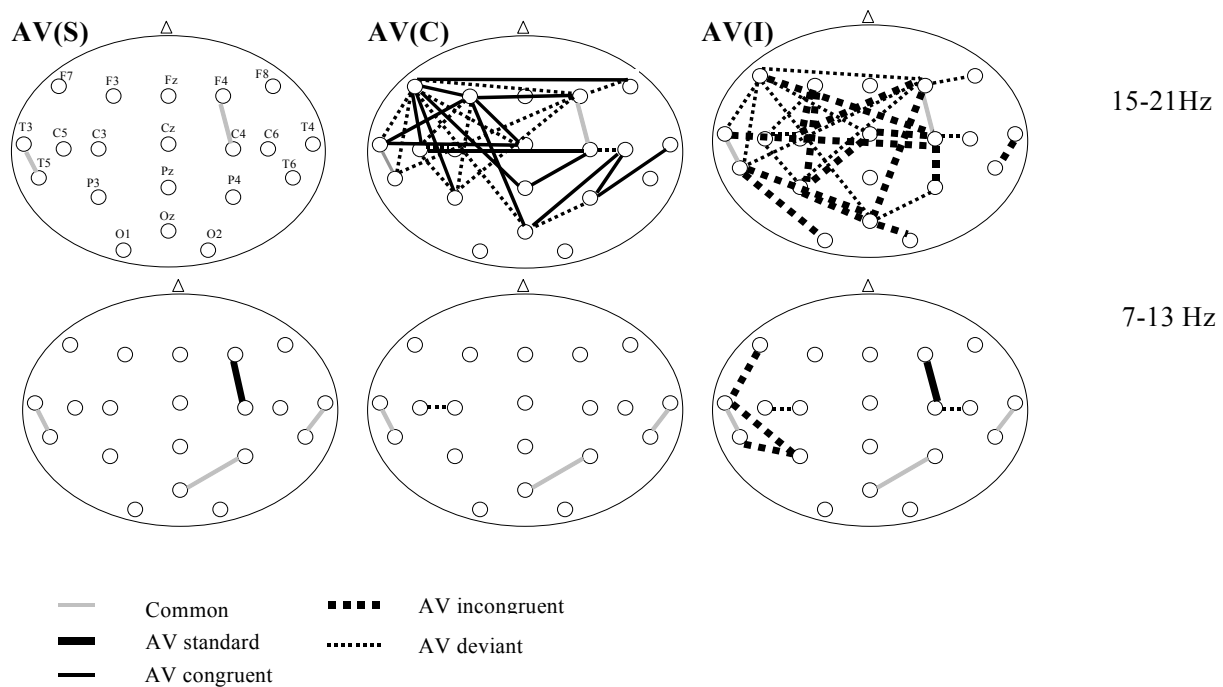

Figure 3. Values of IOS for AV(s), AV(c) and AV(i) (the McGurk effect) stimuli in the alpha and beta frequency bands. The IOS values which occurred more than in 75% of repetitions across all "McGurk subjects" are mapped onto schematic brain maps as connecting lines between the MEG locations involved. On the upper left image the labels of MEG sensors correspondent to EEG locations (see Methods) are shown.

Superimposition of Unimodal Auditory and Visual Maps of Interactions

In order to extract the cortical network reflecting audio-visual integration, the OS maps derived from the auditory and visual experiments were summed. In the framework of the coactivation model (Miller, 1982, 1986), such connections between cortical areas which were not present in both of the unimodal A and V conditions but emerged in the bimodal AV condition were supposed to reflect the integration process.

Figure 4 displays the superimposition of the most frequently found connections (IOS) in the all "McGurk subjects" for the audio-visual *standard* AV(s) stimuli and the algebraic sum of OS connections for unimodal A(s) and V(s) stimuli in the alpha (7-13Hz) and beta (15-21Hz) frequency bands. Although there were some connections which resembled the sum [A(s) + V(s) = AV(s)], the emerging of new and unique connections (for alpha band – black thick doted lines) and the disappearing of some of the connections specific to the unimodal conditions (for beta band – black thick and thin solid lines) indicate that multimodal information processing activates specific networks and cannot be considered a linear sum of the unimodal networks (Figure 4).
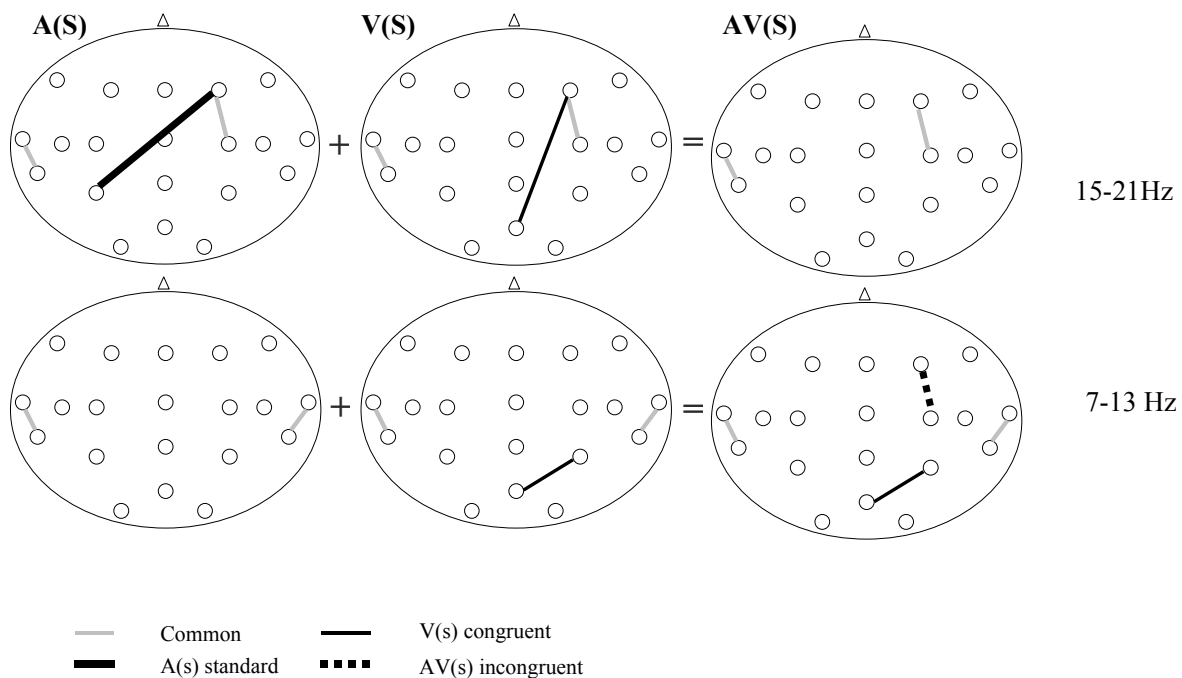

Figure 4. The superimposition of the most frequently found brain sites' connections (the IOS values) across all "McGurk subjects" for AV(s) stimuli and the algebraic sum of operational synchrony connections for unimodal A(s) and

V(s) stimuli in the alpha and beta frequency bands. The IOS values which occurred more than in 75% of repetitions across all "McGurk subjects" are mapped onto schematic brain maps as connecting lines between the MEG locations involved.

The same design of analysis is presented in Figure 5, where the superimposition of the most frequently found brain sites' connections (IOS) across the all "McGurk subjects" for audio-visual *congruent* AV(c) stimuli and the algebraic sum of OS combinations of unimodal A(d) and V(d) stimuli in the alpha (7-13Hz) and beta (15-21Hz) frequency bands are displayed. The cortical network observed during the bimodal AV experiment was a mixture of combinations which reflects the summation [A(d) + V(d) = AV(c)] (Figure 5, black thick and thin solid lines in AV) and the new combinations (Figure 5, black thin doted lines in AV), which emerged only during the bimodal AV experiment and which were not present in both the unimodal A and V experiments. Also there were some cortical connections which were irrelevant to modality – they were revealed in A, V and AV experiments (Figure 5, gray lines). The networks in the beta frequency band were denser compared to the alpha frequency band in all modalities (Figure 5). For the beta frequency band, the unimodal and bimodal effects were widely distributed and mostly confined to the left hemisphere.
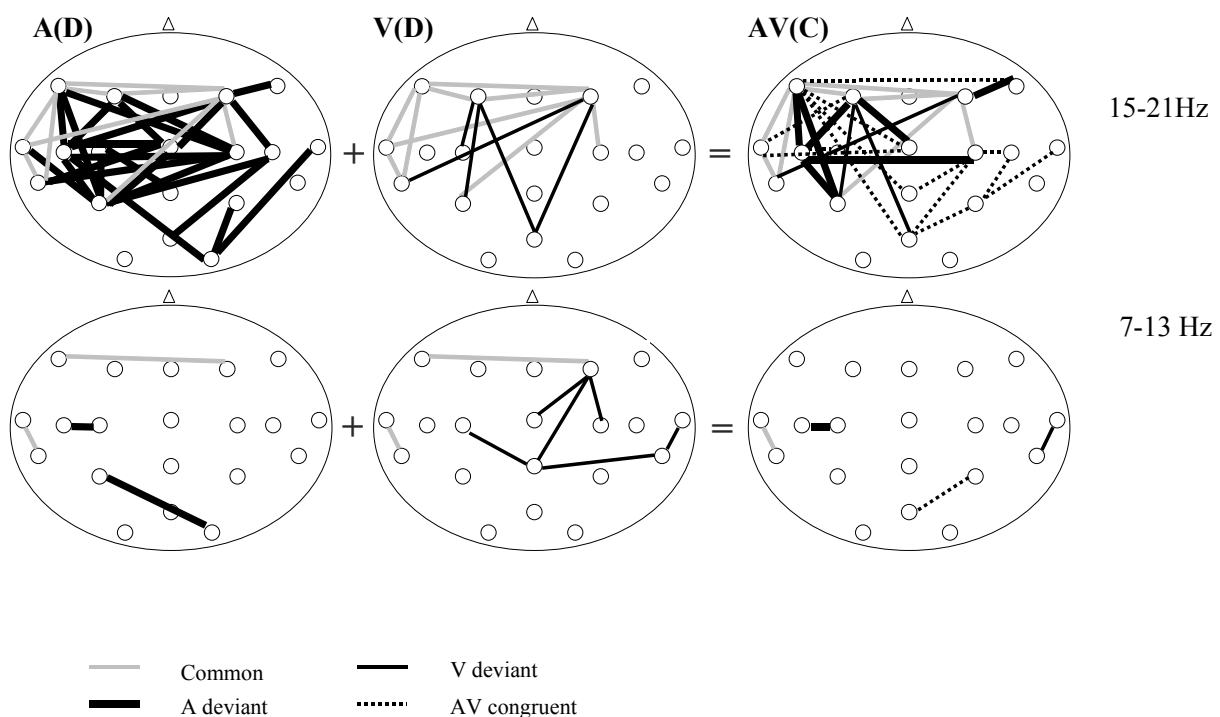
Figure 5. The superimposition of the most frequently found brain sites' connections (the IOS values) across all "McGurk subjects" for AV(c) stimuli and the algebraic sum of operational synchrony connections for unimodal A(d) and V(d) stimuli in the alpha and beta frequency bands. The IOS values which occurred more than in 75% of repetitions across all "McGurk subjects" are mapped onto schematic brain maps as connecting lines between the MEG locations involved.

Audio-visual integration during *incongruent* AV(i) stimuli (the McGurk effect) requires another design for analysis. This analysis can be written as [V(s) – V(d) = AV(s) – AV(i)]. If the AV integration is a simple algebraic summation, then the result of subtraction on the right-hand side of the equation should be equal to the result of subtraction on the left-hand side of the equation. Note that the auditory component in the right-hand side of the equation should have been eliminated because it is the same "ipi" (see Methods) for AV(s) and AV(i) stimuli. Figure 6 displays the result of subtractions on both sides of the equation. The type of lines indicates the modality from which the particular connection comes. Black thin doted lines show exclusive connections, which organized network of cortical areas, specific for audio-visual integration during the McGurk effect. Figure 6 indicates that AV integration network is not the result of a linear sum of the unimodal networks and that it has emergent properties. For beta activity the AV network was more widespread and denser than for alpha activity, although in both frequency bands the dominance of the left hemisphere was revealed (see Figure 6).

Interactions Between Brain Oscillations

A comparison of Figures 3, 4, and 5 reveal that some brain areas connections were the same for the alpha and the beta frequency bands under some experimental conditions. Such connections may indicate that the alpha and beta frequency bands, in these cortical areas, were operationally synchronized between each other. Table 3 summarizes the connections of cortical areas which were present during the same experimental condition simultaneously in alpha (7-13Hz) and beta (15-21Hz) frequency bands. It was observed that these connections involved the left temporal, and frontal and central areas bilaterally. The connection T3-T5 was present in all experimental conditions and modalities. In contrast, connections C5-C3 and C4-C6 were observed only during AV(i) stimuli – the McGurk-type stimuli (see Table 3).

Figure 6. The result of subtractions for both sides of equation [V(s) – V(d) = AV(s) – AV(i)]. The IOS values which occurred more than in 75% of repetitions across all "McGurk subjects" are mapped onto schematic brain maps as connecting lines between the MEG locations involved.

Table 3.
The cortical sites' combinations which occur simultaneously in two frequency bands (alpha and beta) during different experimental conditions in the McGurk subjects

Conditions	Combinations	Conditions	Combinations	Conditions	Combinations	Conditions	Combinations	Conditions	Combinations
AV(standard)	T3-T5	AV(congruent)	T3-F7	AV(congruent)	F3-F4	AV(congruent)	F4-C4	AV(incongruent)	C5-C3, C4-C6
AV(congruent)		AV(inconguent)		A(deviant)		AV(inconguent)			
AV(incongruent)		A(deviant)		V(standard)					
A(standard)		V(deviant)		V(deviant)					
A(deviant)									
V(standard)									
V(deviant)									

Comparison of Audio-Visual Interaction Maps in the "McGurk Subjects" and in the "Non-McGurk Subjects"

Figure 7 presents the networks of connections between different MEG locations mapped onto schematic brain maps for the subjects who had the McGurk effect ("McGurk subjects", *n*=7) and those subjects who did not have the McGurk effect ("non-McGurk subjects", *n*=2). Since there were only two "non-McGurk subjects", this data should be treated with care. Although both groups of subjects had common brain sites' connections (Figure 7, gray lines), the majority of connections typical for the "McGurk subjects" were absent in the "non-McGurk subjects" (Figure 7, black thick solid lines). Instead, the "non-McGurk subjects" had unique connections (Figure 7, black thin lines). The main finding was the existence of negative values for the index of operational synchrony (IOS) between some MEG locations in the "non-McGurk subjects" (black thin doted lines in Figure 7). This means that the MEG signals recorded from these locations had systematically unsynchronized segments. Such type of connections was observed in both frequency bands studied.

MG   non-MG   15-21Hz

7-13 Hz

——— Common connections for MG and non-MG subjects

━━━ Disappearing in non-MG subjects

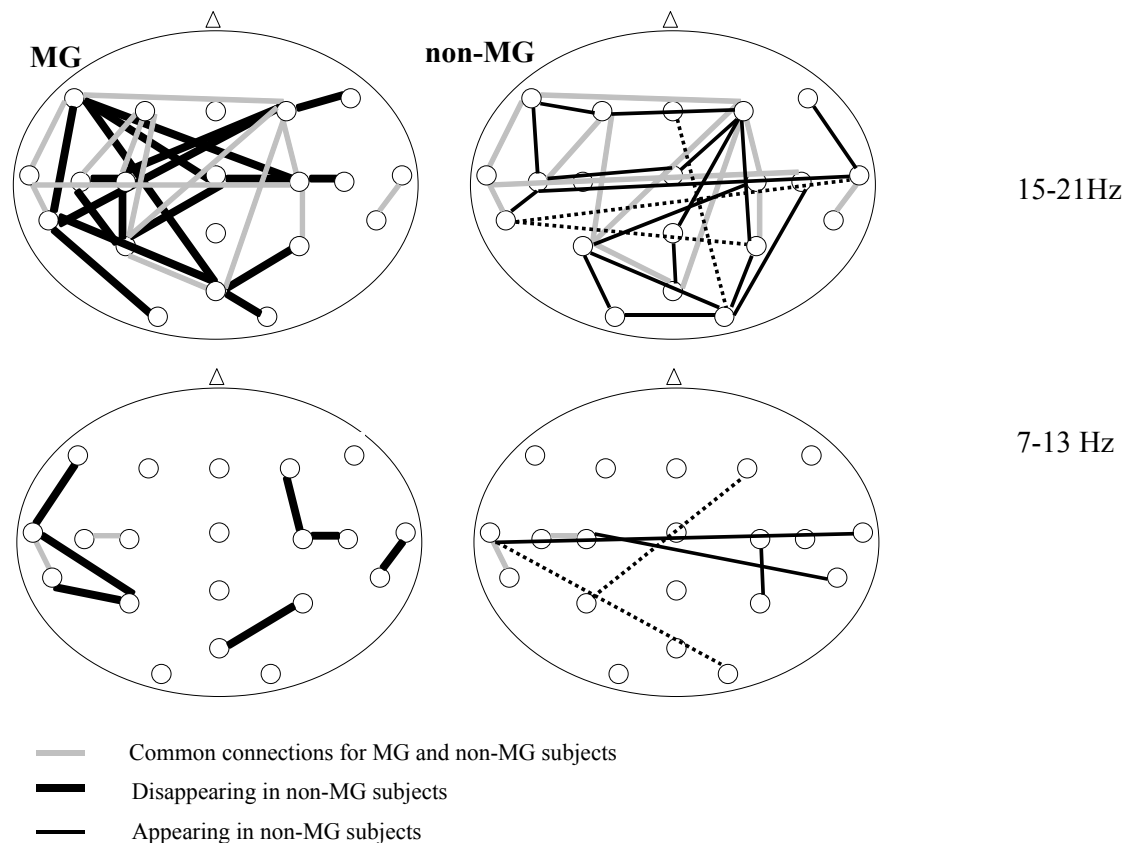——— Appearing in non-MG subjects

Figure 7. The networks of interactions between various brain sites mapped onto schematic brain maps for the subjects who had the McGurk effect (MG) and the subjects who did not have the McGurk effect (non-MG). The IOS values which occurred more than in 75% of repetitions across all subjects are mapped onto schematic brain maps as connecting lines between the MEG locations involved.

DISCUSSION

Dynamic Network of Cortical Interactions

In the present study we observed the existence of widespread networks of active functional interactions between various cortical brain sites involved in audio-visual speech information integration (Figure 3). It should be remembered that the changes in operational synchrony maps here were only considered relevant if these changes appeared consistently in a majority of the trials (not less than 75% of occurrence in all trials and all subjects) under the experimental conditions being analyzed. This permits us to overcome the common problem of multiple comparisons between maps which exists due to the large number of electrode pairs in the maps (Rappelsberger & Petsche, 1988). However, such comparisons between maps should be considered descriptive rather than confirmatory (Stein et al., 1999) which is common for studies with multiple comparisons between maps (Weiss & Rappelsberger, 2000; Razoumnikova, 2000).

The components of networks observed in the present study seemed to be different depending on the nature of the information that had been combined (vowel-consonant-vowel disyllables), the particular combination of modalities (auditory and visual) and the stimulus type (standard and deviant stimuli). The main cortical sites which functionally interacted with each other during the AV integration in the present study roughly included zones overlaying the superior temporal sulcus (STS), inferior parietal sulcus (IPS), parieto-preoccipital cortex (occipital for incongruent AV condition), central and motor cortices, posterior cortex and frontal regions including premotor and prefrontal cortices (Figure 2). These cortical regions are in congruence with the brain areas considered crucial for crossmodal integration (Fries, 1984; see review, Calvert, 2001; and also Dogil et al., 2002). It has been assumed that the STS plays an important role in audio-visual speech integration whereas the IPS specializes in the synthesis of crossmodal coordinate cues and attention (Calvert, 2001). The involvement of frontal

regions as indexed by the process of operational synchrony during AV integration seemed somewhat unusual, but there is evidence that areas within these regions may also be involved in audio-visual information processing (audio-visual temporal synchrony-asynchrony detection) (Bushara, Grafman, & Hallett, 2001). Anterior brain areas have been found to become activated also during speech perception, visual judgments (Dogil et al., 2002), working memory (Petrides, 1994) and involved in integrating newly acquired crossmodal associations (Calvert, 2001). The motor areas probably processed kinematic operations, important for visual speech perception. It was shown that kinematic primitives are crucially important for AV integration in the McGurk effect (Rosenblum & Saldana, 1996).

Probably the so-called transmodal cortical areas already explored in other works (Calvert, 2001) and large-scale networks found in the present study are the parts of the same system, where transmodal areas act as critical gateways for binding information from multiple brain areas into distributed but integrated multimodal representations (Mesulam, 1998). It is important to stress that the transmodal areas referred to above are not necessarily centers where unified percept resides but rather are critical gateways for accessing the relevant distributed information (Mesulam, 1994).

Interaction Between Brain Oscillations during Audio-Visual Integration

Both hemispheres were involved in the process of audio-visual speech integration, with the left hemisphere exhibiting more interconnections than the right hemisphere (in both frequency bands) (Figure 3). In the beta frequency band, the network of cortical operational synchrony interactions was denser then in the alpha frequency band. The reason for such a strong interconnected net of cortical sites in the beta band during AV speech perception was most probably due to the processing of the kinematic properties of the moving biological face as visual speech information (Rosenblum & Saldana, 1996). It is supposed that particularly the visual speech information is of primary importance for AV speech integration (Sams et al., 1991; Rosenblum, Yakel, & Green, 2000; Möttönen et al., 2002). The previously mentioned kinematic properties of a moving biological face are coded as motor functions with which beta brain oscillations have been usually associated (Hari & Salmelin, 1997; Pfurtscheller et al., 1998).

Some cortical sites synchronized their operations simultaneously in both the alpha and beta frequency bands (see Table 3). This may mean that the temporal structure (segmental structure) of the MEG signals within the alpha and beta frequency bands in these sites was approximately the same. If so, then it may be the case that cortical sites involved may synchronize their operations also between different frequency bands. The possibility of operational synchronization between brain oscillations at different frequencies has been demonstrated previously for the first time (Kaplan et al., 1998; Fingelkurts, 1998). The present data reflects the modern view of interfrequency consistency as one principle of brain integrative functioning (Nunez, 1995; see also the review, Fingelkurts & Fingelkurts, 2001). According to this theory, brain information processing takes place at multiple timescales and is mediated by binding between various frequencies (see the review, Kaplan, 1998; Nunez, 2000). This allows rapid information processing simultaneously on both a local and global scale (Ingber, 1995; Nunez, 2000; Fingelkurts & Fingelkurts, 2001).

Emergent Properties of Integrative Cortical Network

We also observed that the distributed cortical networks involved in audio-visual speech integration had emergent properties, rather than being a simple sum of the networks present during unimodal stimulation (see Figure 4, 5 and 6). This finding is keeping with recent studies (Giard & Peronnet, 1999; Calvert et al., 2001; for the review, see Calvert, 2001), suggesting that multisensory integration is a process which not only facilitates detection of the multisensory stimuli by amplification of the unimodal sensory signals, but combines these signals to form a *new*, multimodal representational percept (O'Hare, 1991). This new multimodal percept categorization is consistent with the theory of *emergence*, where the complexity of the system makes possible types of phenomena which could not be generated by the components alone or summed together (Kim, 1992).

Although in a number of studies sets of specific brain areas have been found to be involved in AV information integration (Callan et al., 2001; Calvert et al., 2001; Dogil et al., 2002), in the current study, probably for the first time, emphasis was put on the detection of functional connections (so called cross-talk) between different cortical sites. It should be stressed that to reveal the set of brain areas activated during AV information processing is not sufficient to prove whether the activated areas are actually responsible

for multisensory information integration (see the review, Calvert, 2001). We propose that the apparent synthesis of information from different modalities may be achieved through the process of operational synchrony between modality-specific and non-specific cortical areas. The main principle lies in the moment-by-moment metastable *synchronization of the on-going changes* of brain activity between different cortical areas of the large-scale networks (Kaplan & Shishkin, 2000; Bressler & Kelso, 2001; Fingelkurts & Fingelkurts, 2001). These changes (rapid transition processes) have been determined as the triggering moments of the discrete operations processed in various cortical sites (Kaplan et al., 1997; Fingelkurts & Fingelkurts, 2001).

Such an interpretation is consistent with a large-scale functional organization based on selectivity distributed processing (Mesulam, 1994) and stresses that when the new percept is built, it is not tied to a fixed, domain-specific format of representation, but is instead often amodal. The conscious knowledge of such a percept can also be represented in the form of a cross-system code (Fodor, 1983) which provides any conscious content with an abstract value.

The Duration of Cortical Operations during Audio-Visual Integration

We also observed that the duration of quasi-stationary segments (operations) occurring in the different cortical sites under the different experimental conditions varied (as estimated from Table 1). For all deviant stimuli (irrelevant to modality), the duration of brain operations was significantly longer than for the standard stimuli, and for audio-visual stimuli the duration of brain operations was significantly shorter than for unimodal stimuli (irrelevant to modality). Also, brain operations tended to be of a longer duration in response to the presentation of incongruent audio-visual stimuli than to the presentation of congruent audio-visual stimuli (as estimated from Table 1). Such observations might be associated with behavioral results on reaction times (RT). Shorter RT has been reported for standard stimuli than for deviant stimuli (Jääskeläinen et al., 1996; Escera et al., 1998), also longer for incongruent than on congruent audio-visual stimuli (Green & Kuhl, 1989; Green & Gerdeman, 1995), and much shorter on crossmodal stimuli than on unimodal stimuli (Miller, 1982; Frens et al., 1995; Giard & Peronnet, 1999).

Such parallels on a neurophysiological and behavioral level permit us to assume that both measures indicate the speed of the cognitive operations. Probably the decrease in the cognitive speed during deviant audio-visual stimuli reflects the involvement of working memory processes needed to estimate the difference between an incoming deviant stimulus and the sensory memory trace representing the previous standard stimuli. This supposition also agrees with the observation in the present study that the frontal cortical areas are involved during deviant stimuli presentation (see Figure 2). The frontal sites are exactly those areas that have always been associated with working memory (Klingberg, 1998). The fact that the speed of cognitive processes is shorter for multimodal stimuli than for either unimodal stimuli can be explanation by the *coactivation model* (Miller, 1982, 1986) where the parallel processing of unimodal channels interacts somewhere in the processing system and forms a multimodal percept (Giard & Peronnet, 1999).

Integrative Networks in the "McGurk Subjects" and "Non-McGurk Subjects"

The subjects who did not have the McGurk effect had different integrative networks of functional interactions between cortical sites compared to the subjects who had the McGurk effect. Since the vast majority of people have the McGurk effect (Rosenblum & Saldana, 1996) it is hard to find subjects who can't experience it. In the current study, we had only two of such the subjects ("non-McGurk subjects"). Although the data obtained from the two "non-McGurk subjects" must be assumed to be tentative, networks of cortical sites' interactions distinct from the "McGurk subjects" were found (see Figure 7). In the "non-McGurk subjects" negative values of operational synchrony were observed. This means that these cortical sites *actively* unsynchronized their operations (for conceptual review see Fingelkurts & Fingelkurts, 2001).

These observations can be also incorporated with behavioral findings. The McGurk effect has effectively been used as a tool to investigate audio-visual speech information integration (Rosenblum, Yakel, & Green, 2000), suggesting that various features from each modality can be integrated to produce a separate fused percept (Massaro, 1987; Sekiyama & Tohkura, 1991). This effect is quite striking: even if completely aware of the nature of the audio-visual stimuli, the subjects still report *hearing* a clear syllable which is *unconsciously* influenced by what they *see* (Repp et al., 1983; Rosenblum, Yakel, & Green, 2000). Within this framework the "non-McGurk subjects" could not integrate

information coming from different modalities: i.e., they perceived the auditory and visual stimuli independently and they were *conscious* about the mismatch between them. The existence of unsynchronized (negative) interactions between some cortical sites in the "non-McGurk subjects" in our study may be in agreement with this. Uncoupled interactions between cortical sites revealed in the experiment, involved long connections between anterior and posterior brain areas and also between left and right hemisphere temporal sides (Figure 7). This negative operational synchrony observed in the "non-McGurk subjects" may be the mechanism responsible for the absence of integrated audio-visual percept in the "non-McGurk subjects". The lack of interactions between the anterior and posterior parts of the brain, and also the left and right hemispheres, which cause the destruction of a unified percept, has been also shown during the anesthesia studies using coherence analysis (John et al., 2001).

Concluding Remarks

Taken together, the results from the current study (1) support a model of crossmodal integration within distributed and dynamic cortical networks (Calvert et al., 2001; see also the review, Calvert, 2001) with emergent properties. However the novelty of the data obtained from the current study is in revealing the functional connections (cross-talk) between various cortical sites as indexed by *operational synchrony*. It should be mentioned that the term "network" is often used in a very broad meaning. For example, in neuroimaging this term is usually used inappropriately to indicate a collection of activated areas (McIntosh, Fitzpatrick, & Friston, 2001). A central feature in the organization of the large-scale network is the absence of one-to-one correspondences among anatomical sites, information processing and complex functional act (Mesulam, 1990). We used the term "network" as a computational model (or conceptual one) that relates cognitive processes or subprocesses (Friston & Price, 2001).

(2) Our findings on interfrequency (alpha and beta) consistency in terms of operational synchrony during audio-visual speech perception suggest the involvement of brain oscillations at these frequencies in audio-visual speech integration. And also stress the importance of operational synchronization between different brain oscillations for integrative brain activity (Nunez, 1995; Fingelkurts & Fingelkurts, 2001).

(3) The obtained results provide physiological evidence for a coactivation model (Miller, 1982, 1986) for successful audio-visual speech integration ("McGurk subjects"). However, subjects without the McGurk effect possibly process information from both modalities independently due to active process of unsynchronized brain operations.

 (4) The temporal synchronization of operations processed unimodal stimuli in different cortical sites suggests the importance of time-varying features for audio-visual speech integration. Temporal characteristics of visual and auditory primitives can act as input for speech integration (for discussion see, Green & Gerderman, 1995; Rosenblum & Saldana, 1996).

(5) The present study concentrated on audio-visual crossmodal speech integration but similar principles may also be applied for other modalities integration.

Acknowledgements

REFERENCES

Bressler, S.L., & Kelso, J.A.S. (2001). Cortical coordination dynamics and cognition. *Trends in Cognitive Sciences,* **5**, 26-36.

Bushara, K.O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *The Journal of Neuroscience,* **21**, 300-304.

Callan, D.E., Callan, A.M., Kroos, C., & Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: A single-sweep EEG study. *Brain Research. Cognitive Brain Research,* **10**, 349-353.

Calvert, G.A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex,* **11**, 1110-1123.

Calvert, G.A., Hansen, P.C., Iversen, S.D., & Brammer, M.J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage,* **14**, 427-438.

Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., Riecker, A., & Wildgruber, D. (2002). The speaking brain: a tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. *Journal of Neurolinguistics,* **15**, 59-90.

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences,* **2**, 254-262.

Engel, A.K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience,* **2**, 704-716.

Erdi, P. (2000). On the 'Dynamic Brain' Metaphor. *Brain and Mind,* **1**, 119-145.

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience,* **10**, 590-604.

Ettlinger, G., & Wilson, W.A. (1990). Cross-modal performance: behavioral processes, phylogenetic considerations and neural mechanisms. *Behavioral Brain research,* **40**, 169-192.

Fingelkurts, An.A. (1998). *Time-spatial organization of human EEG segment's structure.* Ph.D. Dissertation, Moscow State Univ., Moscow, p. 415 (in Russian).

Fingelkurts, An.A., & Fingelkurts, Al.A. (2001). Operational architectonics of the human brain biopotential field: Towards solving the mind-brain problem. *Brain and Mind,* **2(3)**, 261-296.

Fodor, J. (1983). *The modularity of mind.* Cambridge: MIT Press.

Frens, M.A., Vanopstal, A.J., & Vanderwilligen, R.F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception and Psychophysics,* **57**, 802-816.

Fries, W. (1984). Cortical projections to the superior colliculus in the macaque monkey: A retrograde study using horseradish peroxidase. *The Journal of Comparative Neurology,* **230**, 55-76.

Friston, K.J., & Price, C.J. (2001). Dynamic representations and generative models of brain function. *Brain Research Bulletin,* **54**, 275-285.

Giard, M.H., & Peronnet, F. (1999). Audio-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive neuroscience,* **11**, 473-490.

Green, K.P., & Gerdeman, A. (1995). Cross-modal discrepancies in coarticulation and the integration of speech information: The McGurk effect with mismatched vowels. *Journal of Experimental Psychology. Human Perception & Performance,* **21**, 1409-1426.

Green, K.P., & Kuhl, P.K. (1989). The role of visual information in the processing of place and manner features in speech perception. *Perception and Psychophysics,* **45**, 34-42.

Haken, H. (1999). What can synergetics contribute to the understanding of brain functioning? In C. Uhl (Ed.), *Analysis of neurophysiological brain functioning*, (pp. 7-40), Berlin: Springer-Verlag.

Hämäläinen, M.S., & Ilmoniemi, R. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical and Biology Engineering and Computing,* **32**, 35-42.

Hari, R., & Salmelin, R. (1997). Human cortical oscillations: A neuromagnetic view through the skull. *Trends in Neuroscience,* **20**, 44-49.

Ingber, L. (1995). Statistical mechanics of multiple scales of neocortical interactions. In P.L. Nunez (Ed.), *Neocortical dynamics and human EEG rhythms* (pp. 628-681), New York: Oxford University Press.

Jääskeläinen, I., Alho, K., Escera, K., Winkler, I., Sillanaukee, P., & Näätänen, R. (1996). Effects of ethanol and auditory distraction on forced choice reaction time. *Alcohol,* **13**, 153-156.

John, E.R. (2001). A Field theory of consciousness. *Consciousness and Cognition,* **10**, 184-213.

John, E.R., Prichep, L.S., Kox, W., Valdés-Sosa, P., Bosch-Bayard, J., Aubert, E., Tom, M., diMichele, F., & Gugino, L.D. (2001). Invariant Reversible QEEG Effects of Anesthetics. *Consciousness and Cognition,* **10**, 165-183.

Kaplan, A. Ya. (1998). Nonstationary EEG: Methodological and experimental analysis. *Uspehi Physiologicheskih Nayk (Success in Physiological Sciences),* **29(3)**, 35-55 (in Russian).

Kaplan, A.Ya., & Shishkin, S.L. (2000). Application of the change-point analysis to the investigation of the brain's electrical activity. In B.E. Brodsky, & B.S. Darkhovsky

(Eds.), *Non-parametric statistical diagnosis. Problems and methods* (pp. 333-388), Dordrecht: Kluwer Acad. Publ.

Kaplan, A.Ya., Fingelkurts, Al.A., Fingelkurts, An.A., & Darkhovsky, B.S. (1997). Topological mapping of sharp reorganization synchrony in multichannel EEG. *American Journal of Electroneurodiagnostic Technology,* **37**, 265-275.

Kaplan, A.Ya., Fingelkurts, An.A., Fingelkurts, Al.A., & Ivashko, R.M. (1998). Temporal consistency of phasic changes in the EEG basic frequency components. *Zhurnal Vysshei Nervnoi Deiatelnosty Im. I.P. Pavlova (Journal of Higher Nervous Activity),* **48**, 816-826 (in Russian).

Kim, J. (1992). "Downward causation" in emergentism and nonreductive physicalism. In A. Beckermann, H. Flohr & J. Kim (Eds.), *Emergence of reduction? Essays on the prospects of nonreductive physicalism* (pp. 119-138), Berlin: de Gruyter.

Klingberg, T. (1998). Concurrent performance of two working memory tasks: potential mechanisms of interference. *Cerebral Cortex,* **8**, 593-601.

Krause, C.M., Möttönen, R., Jensen, O., Lampinen, J., & Sams, M. (2001). Brain oscillations during audiovisual speech perception – a MEG study. *Paper was presented at the Human Brain Mapping Conference, June 10-14, England.*

Lewis, J.W., Beauchamp, M.S., & DeYoe, E.A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex,* **10**, 873-888.

Lewkowics, D.J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin,* **126**, 281-308.

Lewkowicz, D.J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology. Human Perception and Performance,* **22**, 1094-1106.

Loveless, N., Levänen, S., Jousmäki, V., Sams, M., & Hari, R. (1996). Temporal integration in auditory sensory memory: Neuromagnetic evidence. *Evoked Potential,* **100**, 220-228.

Macaluso, E., Frith, C., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science,* **253**, 85-87.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology,* **21**, 131-141.

Massaro, D.W. (1987). Speech perception by ear and eye. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 53-83). Hillsdale, NJ: Erlbaum.

Massaro, D.W., & Cohen, M.M. (1996). Perceiving speech from inverted faces. *Perception and Psychophysics,* **58**, 1047-1065.

McGurk, H., & MacDonald, J.W. (1976). Hearing lips and seeing voices. *Nature,* **264**, 746-748.

McIntosh, A.R., Fitzpatrick, S.M., & Friston, K.J. (2001). On the marriage of cognition and neuroscience. *Neuroimage,* **14**, 1231-1237.

Mesulam, M-M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annual Neurology,* **28**, 597-613.

Mesulam, M-M. (1994). Neurocognitive networks and selectivity distributed processing. *Review of Neurology (Paris),* **150**, 564-569.

Mesulam, M-M. (1998). From sensation to cognition. *Brain 121*, 1013-1052.

Miller, J.O. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology,* **14**, 247-279.

Miller, J.O. (1986). Time course of coactivation in bimodal divided attention. *Perception and Psychophysics,* **40**, 331-343.

Möttönen, R., Krause, C.M., Tiipana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Brain Research, Cognitive Brain Research,* **13**, 417-425.

Murray, E.A., Malkova, I., & Goulet, S. (1998). In A.D. Milner (Ed.), *Comparative neuropsychology* (pp. 51-69), Oxford: Oxford University Press.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin,* **125**, 826-859.

Neurobihavioral Systems, Inc., Presentation, http://www.neurobehavioralsystems.com, 2001.

Numminen, J., Ahlfors, S., Ilmoniemi, R., Montonen, J., & Nenonen, J. (1995). Transformation of multichannel megnetocardiographic signals to standard grid form. *IEEE Transactions on Biomedical Engineering,* **40**, 72-77.

Nunez, P.L. (1995). *Neocortical dynamics and human EEG rhythms*. New York: Oxford University Press.

Nunez, P.L. (2000). Toward a quantitative description of large-scale neocortical dynamic function and EEG. *Behavioral and Brain Sciences,* **23**, 371-437.

O'Hare, J.J. (1991). Perceptual integration. *Journal of the Washington Academy of Sciences,* **81**, 44-59.

Petrides, M. (1994). Frontal lobes and behavior. *Current Opinion in Neurobiology, ***4**, 207-211.

Pfurtscheller, G., Neuper, C., Schlogl, A., & Lugger, K. (1998). Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Transactions on Rehabilitation Engineering, ***6**, 316-25.

Phillips, W.A., & Singer, W. (1997). In search of common foundations for cortical computation. *Behavioral and Brain Sciences, ***20**, 657-683.

Rappelsberger, P., & Petsche, H. (1988). Probability mapping: power and coherence analysis of cognitive processes. *Brain Topography, ***1**, 46-54.

Razoumnikova, O.M. (2000). Functional organization of different brain areas during convergent and divergent thinking: an EEG investigation. *Cognitive Brain Research, ***10**, 11-18.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip reading.* Hillsdale, NJ: Erlbaum.

Repp, B.H., Manuel, S.Y., Liberman, A.M., & Studdert-Kennedy, M. (1983). Exploring the "McGurk effect". *Paper Presented at the 24*[th] *annual meeting of the Psychonomic Society in San Diego.*

Rosenblum, L.D., & Saldana, H.M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology. Human Perception and Performance, ***22**, 318-331.

Rosenblum, L.D., Yakel, D.A., & Green, K.P. (2000). Face and mouth inversion affects on visual and audiovisual speech perception. *Journal of Experimental Psychology. Human Perception and Performance, ***26**, 806-819.

Salinas, E., & Sejnowsli, T.J. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience, ***2**, 539-550.

Sams, M., & Imada, T. (1997). Integration of auditory and visual information in the human brain: Neuromagnetic evidence. *Society for Neuroscience. Abstracts, ***23**. 1305.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O.V., Lu, S.T., & Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, ***127**, 141-145.

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America,* **90**, 1797-1805.

Seltzer, B., & Pandya, D.N. (1980). Converging visual and somatic sensory cortical input to the intraparietal sulcus of the rhesus monkey. *Brain Research,* **192**, 339-351.

Singer, W., & Gray, C. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience,* **18**, 555-586.

Stein, A.V., Rappelsberger, P., Sarnthein, J., & Petsche, H. (1999). Synchronization between temporal and parietal cortex during multimodal object processing in man. *Cerebral Cortex,* **9**, 137-150.

Stein, B.E., & Meredith, M.A. (1993). *The merging of the senses.* Cambridge, MA: MIT Press.

Stein, B.E., Meredith, M.A., & Wallace, M.T. (1994). Development and neural basis of multisensory integration. In D.J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 81-105). Hillsdale, NJ: Erlbaum.

Stein, B.E., Meredith, M.A., Huneycutt, W.S., & McDade, L. (1989). Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience,* **1**, 12-24.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology 6*, 171-178.

Weiss, S. & Rappelsberger, P. (2000). Long-range EEG synchronization during word encoding correlates with successful memory performance. *Cognitive Brain Research,* **9**, 299-312.

Zeki, S. (2001). Localization and globalisation in conscious vision. *Annual Review of Neuroscience,* **24**, 57-86.