

available at www.sciencedirect.comwww.elsevier.com/locate/brainres**BRAIN
RESEARCH****Research Report****The effect of viewing speech on auditory speech processing is different in the left and right hemispheres**Chris Davis^{a,*}, Daniel Kislyuk^{b,c}, Jeusun Kim^a, Mikko Sams^{b,d}^aMARCS Auditory Laboratories, University of Western Sydney, Australia^bDepartment of Biomedical Engineering and Computational Science, Helsinki University of Technology, Espoo, Finland^cDepartment of Higher Nervous Activity, Saint Petersburg State University, Saint Petersburg, Russia^dLow Temperature Laboratory, Helsinki University of Technology, Espoo, Finland

ARTICLE INFO

Article history:

Accepted 30 April 2008

Available online 11 May 2008

Keywords:

Auditory–visual

Auditory evoked response

Multisensory

Evoked response reduction

Hemispheric differences

ABSTRACT

We used whole-head magnetoencephalography (MEG) to record changes in neuromagnetic N100m responses generated in the left and right auditory cortex as a function of the match between visual and auditory speech signals. Stimuli were auditory-only (AO) and auditory–visual (AV) presentations of /pi/, /ti/ and /vi/. Three types of intensity matched auditory stimuli were used: intact speech (Normal), frequency band filtered speech (Band) and speech-shaped white noise (Noise). The behavioural task was to detect the /vi/ syllables which comprised 12% of stimuli. N100m responses were measured to averaged /pi/ and /ti/ stimuli. Behavioural data showed that identification of the stimuli was faster and more accurate for Normal than for Band stimuli, and for Band than for Noise stimuli. Reaction times were faster for AV than AO stimuli. MEG data showed that in the left hemisphere, N100m to both AO and AV stimuli was largest for the Normal, smaller for Band and smallest for Noise stimuli. In the right hemisphere, Normal and Band AO stimuli elicited N100m responses of quite similar amplitudes, but N100m amplitude to Noise was about half of that. There was a reduction in N100m for the AV compared to the AO conditions. The size of this reduction for each stimulus type was same in the left hemisphere but graded in the right (being largest to the Normal, smaller to the Band and smallest to the Noise stimuli). The N100m decrease for the Normal stimuli was significantly larger in the right than in the left hemisphere. We suggest that the effect of processing visual speech seen in the right hemisphere likely reflects suppression of the auditory response based on AV cues for place of articulation.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Seeing a speaker's articulatory movements improves speech perception, especially so under unfavourable listening conditions in which, e.g., the signal-to-noise ratio is poor (Sumbly and Pollock, 1954). Evidence is accumulating that seeing speech

influences neural processing in the auditory cortex (Campbell et al., 2001; Möttönen et al., 2002; Olson et al., 2002; Sams et al., 1991; Saint-Amour et al., 2007). However, very little is known about specificity of such influences. Viewing speech might increase general sensitivity of the auditory cortex, or it might provide very specific cues to modify some aspects of the

* Corresponding author. Fax: +61 2 9772 6326.

E-mail address: chris.davis@uws.edu.au (C. Davis).

processing involved in matching an auditory token. In this paper we investigate the specificity of the modulation of the auditory cortex signals caused by viewing speech. To measure the neurophysiology of the auditory cortex, we utilized whole-head magnetoencephalography (MEG). As an index of auditory–visual speech processing we examined the size of the modulation in the magnitude of the auditory N100 response (henceforth we will use the term N100m to refer to MEG data and N100 for EEG data).

The N100m is reliably elicited around 100 ms after the onset of an auditory input. The most prominent generator of the N100m is located in the auditory cortex on the lower bank of the lateral sulcus (Lutkenhöner and Steinsträter, 1998; Kuriki et al., 1995; Nakasato et al., 1995; Pantev et al., 1990; Rogers et al., 1990). The amplitude and latency of N100m is sensitive to acoustic stimulus features (Näätänen and Picton, 1987; Näätänen and Winkler, 1999). The prominence, consistency and stability with which the N100m occurs as a consequence of acoustic input make it a useful tool in probing activation of the auditory cortex.

Recently it has been found that N100/N100m magnitude can be significantly reduced by an accompanying visual signal (Klucharev et al., 2003; Besle et al., 2004; van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007). The stimuli in these studies were the auditory and visual signals generated from the same events (e.g., the signals were generated by a person talking). Importantly, the reduction effect appears to be related to whether the visual signal predicts the auditory one.¹ In what follows, we briefly describe the studies that support the view that prediction of the auditory signal is important factor in the generation of the reduction effect and consider how this particular AV effect has been interpreted.

One factor that appears to be important for the AV reduction of the N100m response is that the visual signal enables aspects of the auditory one to be predicted. Support for this proposition comes from two studies in which the visual signal did not reduce the magnitude of the N100m response. In the first, Miki et al. (2004) examined the effect on N100m for processing the vowel /a/ of seeing the talker's articulation. However, in this study, the visual articulation consisted of a still photograph of the talker with their mouth closed (displayed for 800 ms) followed by a photograph of the talker with an open mouth pronouncing the vowel (the vowel sound was presented at the onset of the still photograph of the talker with an open mouth). In this case, the visual speech provided no preparatory movement from which the auditory speech could be predicted and there was no N100m reduction effect.

¹ Ghazanfar et al. (2005) recorded local field potential activity in both core and lateral belt regions of the auditory cortex of rhesus monkeys viewing vocalizing conspecifics and found response suppression. They noted that more enhanced responses occurred for shorter voice-onset times, whereas suppressed responses more often occurred when voice-onset times were longer. Response depression was thus seen as simply being due to the asynchrony between the presentation of the face and voice stimuli. However, in the Miki et al. (2004) study, the still face was presented 800 ms before the onset of the presentation of the auditory vowel and no response suppression was found. This suggests that visual-auditory asynchrony per se is not the only factor involved.

The other study that found no N100 reduction effect was by Stekelenburg and Vroomen (2007). In this study, the authors specifically examined non-speech AV events that either had natural visual preparatory movements or did not. It was predicted that if the reduction in N100 required the visual prediction of auditory onset, then stimuli that provided no visual anticipatory information should likewise produce no AV N100 reduction effect. This was what was found: an effect for stimuli that has visual anticipation; none for stimuli that did not. Based on this result, Stekelenburg and Vroomen (2007) proposed that the prediction provided by the visual input is the crucial property for the suppression of auditory N100 amplitude.

Conversely, the circumstances in which AV presentations have lead to N100m reductions are consistent with the notion that the visual signal allows for the prediction of aspects of the auditory signal. For example, in the experiment of Jääskeläinen et al. (2004), N100m responses were recorded to the auditory presentation of two vowels; these N100m responses were reduced when the vowels were preceded by silent movies of a person articulation them compared to when presented without a preceding stimulus. The study of Martikainen et al. (2005) also provides evidence that prediction is an important factor in response reduction. In this study participants either triggered the (non-speech) stimulus by themselves by a button press, or the stimulus was triggered externally. N100m amplitude was reduced in the former condition compared to the latter.

Three types of proposals have been advanced for why a reduction in N100 response occurs when the visual input predicts the auditory signal: *auditory facilitation* (Besle et al. 2004, Giard and Peronnet, 1999); *sensory gating* (Stekelenburg and Vroomen, 2007; van Wassenhove et al., 2005) and *motor mediation* (Callan et al., 2003; Skipper et al., 2007).² The auditory facilitation account suggests that N100 is reduced because the visual speech signal pre-activates phonetic units in the auditory cortex (possibly via the Superior Temporal Sulcus) and this results in less energetic demand for subsequent auditory processing (e.g., Besle et al., 2004). We interpret this to mean that visual processing adapts the same phonetically-specific neurons in the auditory cortex that are also activated by the auditory speech. The sensory gating account proposes that visual speech information indicates which aspects of the subsequent auditory signal are redundant (such as information about place of articulation carried by the second and third formants) and that this redundant information is filtered out by deactivating of auditory cortex processing of that information (van Wassenhove et al., 2005). The motor mediation account is similar to the sensory gating one as it also proposes that visual speech selectively deactivates processing in the auditory cortex. In the motor account, it is proposed that observed speech activates articulatory motor programs

² The categorization of approaches can be confusing unless functional interpretations of the AV interaction are separated from electrophysiological ones. For example, Stekelenburg and Vroomen (2007) grouped Besle et al. (2004) with van Wassenhove et al. (2005) as both auditory facilitation proposals whereas we classified the latter as a gating account. The reason for the difference is that Stekelenburg and Vroomen considered functional interpretations of AV interaction; whereas we considered the proposed mechanisms of N100 response reduction.

(possibly via so-called mirror neurons in Broca's area, e.g., Callan et al., 2003; Calvert and Campbell, 2003; Skipper et al., 2007; Watkins and Paus, 2004) and it is these that act to depress subsequent auditory processing via the activation of a forward model (or efference copy, see Callan et al., 2003; Heinks-Maldonado et al., 2006). The current study aimed to decide between the interpretations. Given the broad similarity of the two latter accounts (and that they most likely make very similar predictions), they will be considered together.

Just as changes in the N100m to different auditory signals have been used to investigate auditory processing (e.g., Parviainen et al., 2005), changes in this auditory-visual (AV) reduction effect to different AV signals should provide a window into AV processing. Thus, in order to test between the accounts, the current experiment examined the N100m responses for three AV conditions that manipulated the auditory signal to variously alter the degree of "overlap" between the visual and auditory stimuli. Three different types of auditory stimuli were used in conjunction with visual speech: natural speech stimuli; frequency band reduced stimuli and speech amplitude shaped white noise. The rationale for using the latter two types of stimuli was to alter the match between what might be predicted from the visual speech signal and the subsequent auditory one. That is, compared to natural speech, in the frequency band reduced stimuli (see Experimental procedures for details) temporal and amplitude information in each frequency band is preserved but the spectral detail in each band is absent (Shannon et al., 1995). An even greater mismatch between the visual and auditory signals occurs for the speech amplitude shaped white noise stimuli, since these stimuli retain only the gross temporal properties of speech without spectral content.

Under the auditory facilitation account, the degree of response reduction is a function of the degree to which the visual signal has pre-activated phonetic units. In this regard, the effect is essentially driven by the visual stimulus, i.e., the reduction in N100m is due to savings in processing produced by the visual signal. Since the same visual stimuli were used for each the three AV conditions, this would predict that the amount of AV N100m response depression (relative to the AO N100m response) should be similar for the three types of auditory stimuli. That is, although the magnitude of the AO N100m response might be different for the three types of auditory stimuli (reflecting the stimulus-driven activation of different sized neural populations), the visual facilitation of these activated populations should be similar. On the other hand, in both the gating and motor mediation accounts, the degree of N100m depression is a function of the overlap of processes triggered by the visual speech and those by the auditory signal. This would predict that relatively greater depression would be seen in responses to subsequent natural speech stimuli than to the broad-band noise stimuli, since for natural speech, there would be a better match between those processes triggered by visual speech and the subsequent auditory response (i.e., the noise stimuli would have frequencies outside the expected range and the response to these would be unaffected by depression based on the visual speech).

A final issue regarding the expected pattern of auditory response depression for the three types of auditory stimuli concerns whether there are hemispheric differences in the degree of depression. Although certain aspects of speech processing

Table 1 – Upper panel: mean response times (ms) and standard errors for detecting the /vi/ stimulus as a function of Presentation mode: Auditory Only (AO) and Auditory-Visual (AV)

Stimulus type	Presentation mode: hits (SEM)		AV effect
	AO	AV	
Response time			
Normal	810 (33.3)	734 (35.6)	76
Band	806 (31.9)	756 (31.1)	46
Noise	883 (43.8)	807 (38.3)	50
Hits			
Normal	98.4 (1.0)	98.5 (1.9)	0.1
Band	91.1 (2.7)	95.8 (1.6)	4.6
Noise	87.2 (5.0)	90.8 (3.7)	3.6

Lower panel: mean numbers of hits and standard errors for the /vi/ stimulus as a function of Presentation mode.

appear to be lateralized to the left cerebral hemisphere, the degree to which auditory properties are predicted by the visual ones might vary with respect to the type of processing carried out by the each hemisphere. In terms of processing timescale, it has been suggested that visual speech information primarily provides cues to place of articulation that are distributed over moderately long duration windows and are *trans-segmental* (Grant et al., 2004a,b). For auditory speech, relatively long processing timescales have been proposed for the right hemisphere (Boemio et al., 2005) whereas an advantage for temporal resolution at a shorter timescale has been proposed for the left hemisphere (e.g., Nicholls et al., 1999; Nicholls, 1996; Liégeois-Chauvel et al., 2001; Poeppel, 2003; Schönwiesner et al., 2005; Jamison et al., 2006). Recently, the extent to which the left hemisphere is specialized for processing (rapid) temporal properties of sounds has been questioned (Hickok and Poeppel, 2007). However there is greater consensus that the right hemisphere preferentially processes sounds of a longer duration (and small spectral changes). Given that a greater temporal processing correspondence between visual and auditory speech processing occurs in the right hemisphere and that the reduction in N100m indexes the overlap of visual and auditory speech processing, it is expected that there are hemispheric differences in N100m response depression. Specifically, auditory depression might be greater in the right hemisphere (at least for natural speech).

2. Results

2.1. Behavioural results

On average participants detected 91% of the small faint white dots on the stimulus face, suggesting that they attended the visual stimuli. Mean response times (RTs) to the /vi/ syllables for each auditory stimulus type and in different presentation modes are shown in Table 1 (top panel).³ ANOVA showed a significant main effect of Stimulus Type (mean values:

³ Only the data from participants whose MEG data were analysed were included.

Normal=772 ms; Band=781 ms; Noise=845 ms) on the RTs [$F_{(2,16)}=3.67$, $p<0.05$]. There was also a significant effect of Presentation mode, with AV stimuli eliciting faster responses than AO stimuli [$F_{(1,16)}=15.16$, $p<0.05$]. There was no significant interaction between the Stimulus Type and the Presentation mode [$F_{(2,16)}=0.85$, $p>0.05$].

The mean percentage of detected /vi/ syllables is shown in Table 1 (bottom panel). Once again, there was a significant main effect of Stimulus Type (Normal=98.5%, Band=93.5%, Noise=89.0%) [$F_{(2,16)}=4.11$, $p<0.05$]. The effect of Presentation Mode was not significant, but there was a weak tendency for more hits for the AV stimuli [$F_{(1,16)}=2.0$, $p=0.18$]. There was no

significant interaction between Stimulus Type and the Presentation Mode [$F_{(2,16)}=1.33$, $p>0.05$].

2.2. MEG results

Fig. 1 shows the grand average N100m responses over the left and right hemispheres for the different Stimulus types and Presentation mode.

Mean amplitudes are shown in Table 2. In the left hemisphere, the Normal stimuli elicited the largest N100m, both for AO and AV conditions. Also for these conditions, the N100m was second largest for the Band stimuli and smallest for the

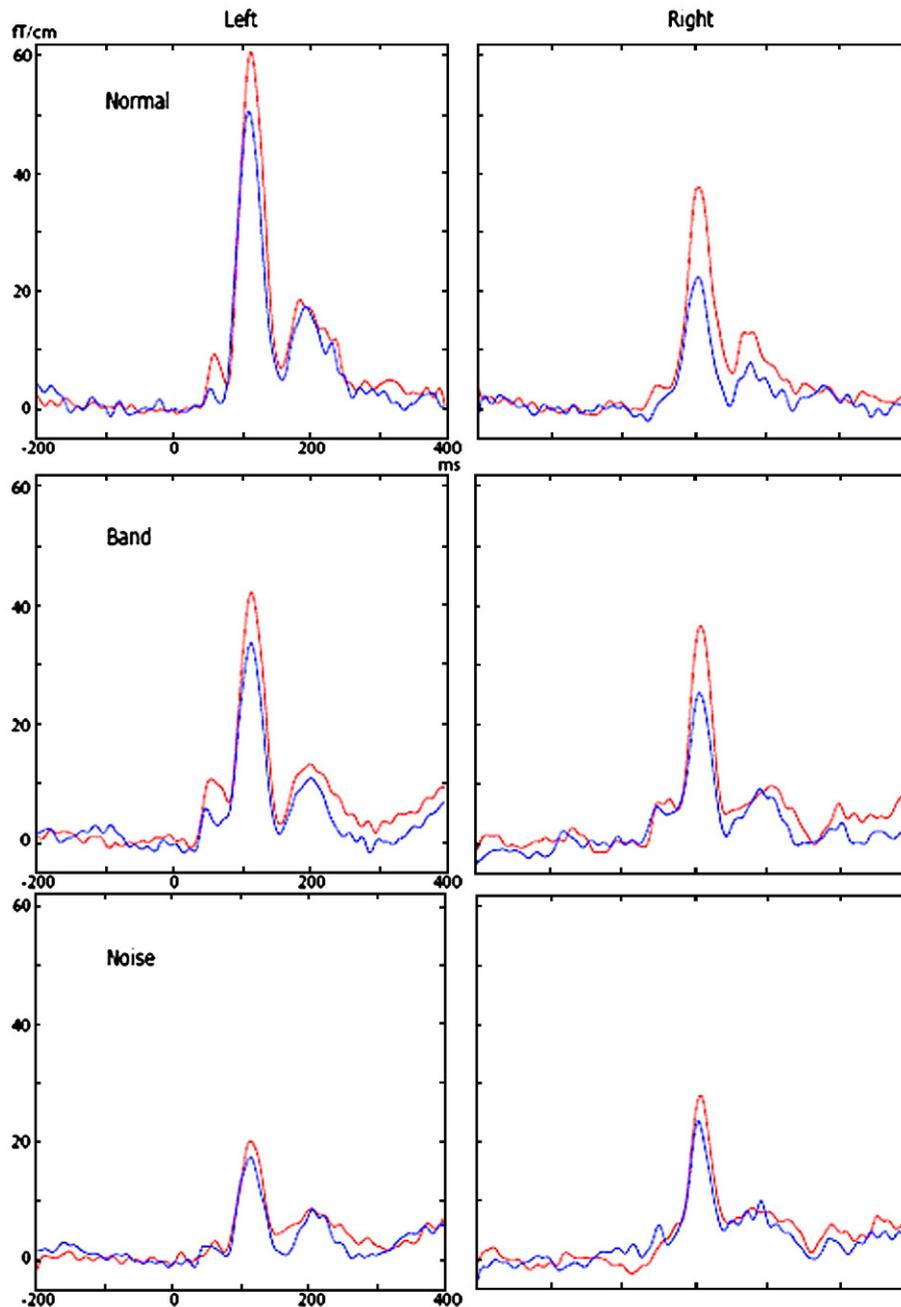


Fig. 1 – Grand average auditory evoked fields from the left and right hemispheres to both the /ti/ and /pi/ stimuli. Responses to AO stimuli are red; those to AV stimuli are blue.

Table 2 – N100m mean amplitudes (fT/cm) for different AO and AV stimuli in the left and right hemispheres

Stimulus type	Left hemisphere		Right hemisphere	
	AO	AV	AO	AV
Normal	57.4 (10.4)	47.5 (10.1)	37.9 (7.2)	22.4 (5.1)
Band	39.2 (5.3)	32.1 (4.0)	34.4 (5.8)	25.1 (4.9)
Noise	19.6 (2.0)	16.8 (4.1)	27.4 (4.6)	22.5 (3.4)

Noise stimuli. In the right hemisphere, Normal and Band AO stimuli elicited N100m responses of quite similar amplitudes, with the N100m amplitude to Noise less but larger than that in the left hemisphere. The amplitude of N100m elicited by Normal and Band stimuli was smaller in the AV than in the AO condition. This reduction effect appeared to be more pronounced in the right than in the left hemisphere.

Results of the ANOVA of N100m mean amplitudes are shown in Table 3. As can be seen in the table, the main effect of the Stimulus Type was significant [$F_{(2,16)}=9.94, p<0.05$], as were the main effect of Presentation mode (AO vs AV) [$F_{(2,16)}=16.37, p<0.05$] and the Presentation mode <0.05 . Post-hoc tests revealed that the difference between the AO and AV conditions for the Normal and Band stimuli was significant [$p<0.05$ for both]. However, this difference was not significant for the Noise stimuli [$p>0.05$]. The Stimulus Type \times Hemisphere interaction was also significant [$F_{(2,16)}=13.51, p<0.01$] and a post-hoc test showed that in the left hemisphere, response to the Normal speech stimuli significantly differed from that to the Band and Noise stimuli ($p<0.01$ for both). In the right hemisphere none of the three possible contrasts was significant.

Table 4 shows the number of the subjects whose N100m was smaller in the AV than in the AO condition. Pooled over different auditory stimulus types, smaller responses to AV stimuli were found in the right hemisphere in 8/9 subjects, and in the left hemisphere in 6/9 of the subjects. The most consistent decrease was observed in the right hemisphere for the Normal speech stimuli (9/9 subjects), and the least consistent in the left hemisphere for the Noise stimuli (5/9 subjects).

In what follows we present two measures relating to the decrease in the N100m response due to the presence of visual speech, the absolute decrease and the decrease ratio. The absolute decrease score indicates the overall difference in the size of the active populations between the AO and AV conditions and was determined by taking the algebraic difference between N100 peak amplitude in the AO and AV conditions. The decrease ratio indicates the sensitivity of the activated

Table 3 – Results of a three way ANOVA of N100m mean amplitudes for the factors Stimulus type (1), Presentation mode (2) and Hemisphere (3)

Factors	df effect	df error	F	p-level
1	2	16	9.940	0.002
2	1	8	16.377	0.004
3	1	8	2.125	0.183
1 \times 2	2	16	4.547	0.027
1 \times 3	2	16	13.510	0.000
2 \times 3	1	8	4.069	0.078
1 \times 2 \times 3	2	16	0.335	0.721

Table 4 – Number of subjects (out of nine) that showed smaller N100m amplitude in the AV than in AO condition

Stimulus type	Left hemisphere	Right hemisphere
Normal	6/9	9/9
Band	7/9	7/9
Noise	5/9	8/9

population of neurones to the visual stimuli. This ratio was determined by dividing the absolute decrease score by the AO score and multiplying by 100. Since this measure takes into account the magnitude of the AO response, it provides a more appropriate index with which to evaluate the auditory facilitation and gating/motor mediation accounts.

Table 5 shows the results of the ANOVA conducted with the absolute decrease scores. As can be seen, the main effect of Hemisphere was not quite significant [$F_{(1,8)}=4.07, p=0.08$], but the effect of Stimulus Type was [$F_{(2,16)}=4.55, p<0.05$].

Fig. 2 shows the decreased ratio scores and Table 6 the results of the ANOVA of decrease ratios. There was a significant difference for the Hemisphere factor [$F_{(2,16)}=10.68, p<0.05$], but not for Stimulus Type [$F>1$]. This pattern can be readily interpreted with reference to Fig. 2. As can be seen, in the right hemisphere the magnitude of the decrease ratios were more pronounced for Normal and Band stimuli with the amount for Noise the stimuli less (the values of all stimulus types in the left hemisphere similar to that for the right hemisphere Noise stimuli).

Planned comparison showed that for all auditory stimulus types, the size of the decrease ratio differed significantly from zero [Normal: $t(8)=6.79, p<0.05$], Band: $t(8)=3.27, p<0.05$, Noise: $t(8)=3.26, p<0.05$]. There was a significant difference between the decrease ratios for the Band and Normal Speech stimuli contrasted with the Noise stimuli for the right, but not for the left hemisphere. In the right hemisphere the difference in decrease ratios for the Normal and Noise stimuli was significant ($p<0.05$). These results suggest that for the right hemisphere there was a qualitative difference between the population of neurones activated by the more speech-like stimuli and those by non-speech. That is, the populations activated by the more speech-like stimuli were more prone to the influence of visual speech than the populations activated by noise.

In the left hemisphere, the decrease of the N100m amplitude elicited by Noise stimuli did not differ from zero, due to very large variability of the measure ($t(8)=0.76, p>0.05$). Decrease of the N100m amplitude elicited by Normal stimuli was significant [$t(8)=2.39, p<0.05$]. Decrease of the N100m elicited by the Band stimuli was also significant [$t(8)=3.44,$

Table 5 – Results of the repeated measures ANOVA for the absolute decrease (algebraic difference between N100 peak amplitude in the auditory and audio-visual conditions)

Factors	df effect	df error	F	p-level
1	2	16	4.547	0.027
2	1	8	4.069	0.078
1 \times 2	2	16	0.335	0.721

Factor 1 — Stimulus type; 2 — Hemisphere.

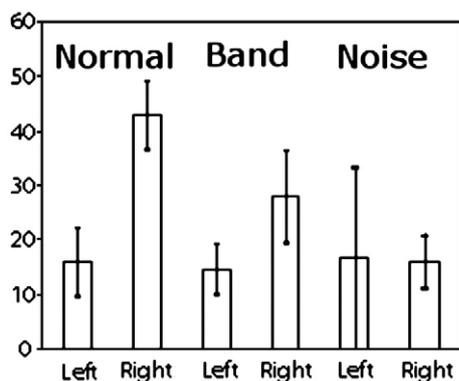


Fig. 2 – Mean decrease of the N100m amplitude due to the processing of the visual speech stimuli $((AO - AV) / AO) * 100$ for the different stimulus types.

$p < 0.05$]. Overall, the mean decrease values were similar for the different auditory stimulus types.

The difference between the decreases in the right and left hemispheres was significant for the Normal stimuli [$t(8) = 4.97$, $p < 0.05$] but not for the Band stimuli [$t = 1.83$, $p > 0.05$] and Noise stimuli [$t(8) = 0.73$, $p > 0.05$]. ANOVA showed that the factor of Stimulus Type was not significant [$F_{(2,16)} = 1.06$, $p > 0.05$], whereas the Hemisphere factor was [$F_{(2,16)} = 10.68$, $p < 0.05$].

3. Discussion

The current study varied the overlap between the visual and auditory speech signals in order to examine two different accounts of the visual speech reduction of the auditory N100m response (the auditory facilitation and the gating/motor mediation accounts). The prediction for the auditory facilitation account was for equal-sized reduction across the three types of auditory stimuli; whereas the prediction for the gating/motor mediation account was for reductions graded by the degree of the overlap between visual and auditory speech processing (e.g., greater reduction for the Normal compared to the Noise stimuli).

The results from the behavioural task demonstrated an effect of Stimulus Type. Response times and identification accuracy varied as a function of the reduction in the speech signal: responses to intact speech being fastest and attracting the lowest number of misses. There was also a visual speech effect, AV stimuli eliciting faster responses than AO stimuli. There was, however, no significant Stimulus Type \times Presentation mode interaction, indicating that the effect of visual speech was similar across stimulus type.

In the MEG data, there were main effects of Stimulus Type and Presentation mode for the decrease ration scores. The pattern in the reduction of N100m with the AV stimulus types varied across the left and right hemispheres. In the left hemisphere, the size of reduction was similar across stimulus type. In the right hemisphere, the reduction was graded, with the largest reduction for the Normal, less reduction for the Band and less still for the Noise stimuli. In our view, the different pattern of reductions across the hemispheres is problematic for the auditory facilitation account (which predicted equal-sized reduction) and more compatible with the gating/

motor mediation account in which reduction can vary according to the overlap between the processes triggered by visual and auditory speech. However, in order to evaluate this account, it is necessary to consider the likely overlap of visual and auditory speech processing. As a first step in doing this, we consider what the data reveal about AO processing with regards to how auditory processing is affected by stimulus type and how this varies in the different hemispheres.

As the auditory signal varied across stimulus type, the AO N100m results can be taken to show how each hemisphere responded to the different auditory stimuli. The overall AO N100m response was generally larger in the left than the right hemisphere, although this pattern was reversed for the Noise stimuli. In the left hemisphere, N100m was largest for the intact speech stimuli and progressively smaller when the speech properties of the stimuli were reduced. This pattern in the size of N100m indicates that the left hemisphere was sensitive to a broad range of speech features present in the signal and when these were reduced (e.g., in the Band and Noise conditions), the size of the N100m was reduced. This finding is consistent with that of a previous MEG study (Parviainen et al., 2005). Parviainen and colleagues examined the N100m response to two synthetic vowels, synthetic consonant-vowel syllables and variously matched non-speech sounds that either contained the first three formant frequencies of the speech sounds or one formant frequency (F2). A larger N100m response was observed for the synthetic speech stimuli in the left hemisphere (but not the right) and it was suggested that the neurons in the left hemisphere are more sensitive to stimuli that formed phonetically relevant arrangements of acoustic features (see also Scott et al. 2000).

In the right hemisphere, the AO N100m responses were generally smaller (similar in size to the Band condition in the left hemisphere) and the differences between the stimulus types was less marked (similar sized reduction for the Normal and Band stimuli, which were larger than that for the Noise stimuli). This pattern suggests that, compared to the left hemisphere, the right may be less sensitive to particular signal properties (e.g., features signalled by rapid signal change, such as cues for voicing, Trébuchon-Da Fonseca et al., 2005; Sandmann et al., 2007). On the other hand, the right hemisphere might be more sensitive to degraded speech (Liikkanen et al., 2007) in which the slow temporal features of the speech envelope have been relatively emphasised (Abrams et al., 2008), hence the larger N100m in the right hemisphere to the Noise stimuli. The difference between the hemispheres in the N100m response patterns observed across stimulus type is consistent with the proposition of hemispheric asymmetries in information processing in which the left hemisphere codes

Table 6 – Results of the repeated measures ANOVA for the decrease rate $((AO - AV) / AO) * 100$

Factors	Effect	Error	F	p-level
1	2	16	1.062	0.369
2	1	8	10.678	0.011
1 \times 2	2	16	1.340	0.290

Factor 1 — Stimulus type; 2 — Hemisphere.

for memory traces of acoustic templates for speech sounds formed in this hemisphere (Shtyrov et al., 2000) and this may assist in processing rapid changes in auditory stimuli (Hutsler and Galuske, 2003).

To summarize the interpretation of the AO results, it is suggested that the left hemisphere is sensitive to more speech features than the right hemisphere (e.g., those involving rapid changes), so the right hemisphere response was smaller to speech, but supplemented in the Band and Noise conditions by an additional sensitivity to non-speech features. This pattern of N100m for AO presentations serves as the basis for understanding the pattern of response reduction for the AV conditions. In evaluating the proposal that the overlap of AV information was responsible for the pattern of N100m reduction over stimulus type and hemisphere, it is important to note what the relevant information was. Auditory speech provides information about place and manner of articulation and voicing whereas the place cue⁴ is the primary information available from the visual signal (Grant et al., 2004a). Thus it seems likely that it was the combined AV information provided by the place cue that played a role in generating the reduction in the N100m response.

To explain the left and right hemisphere AV reduction results in terms of the overlap between visual and auditory cues to place, it is necessary to specify the nature of the place cues that may have been used. There are several important acoustic features that distinguish place in initial stop consonants, properties of the initial release burst (frequency spectrum; burst amplitude, frication, aspiration) and the dynamic trans-segmental formant transition that follows (e.g., Hazan and Rosen, 1991; Greenberg and Arai, 2004). These two types of property differ in their temporal deployment and this may bias how they are processed in the different hemispheres. For instance, a number of studies (Belin et al., 1998; Boemio et al., 2005; Schönwiesner et al., 2005; Jamison et al., 2006) have suggested that the timescale over which this information processing is integrated in the right hemisphere is longer than the left; given this, then processing of trans-segmental speech information would be more suited to the right than left hemisphere.⁵ Additionally, it has been proposed that the right hemisphere may be more sensitive to spectral information (e.g., Zatorre and Belin, 2001; Hyde et al., 2008). These processing properties suggest that the left hemisphere might be more

sensitive to properties of the initial release burst whereas the right hemisphere might be more sensitive to trans-segmental place cues (syllable-sized; spanning several phonetic segments) that are associated with the dynamic patterns of the trans-segmental formant transitions that bind consonants and vowels (Greenberg and Arai, 2004).⁶ (Greenberg and Arai, 2004).

It was suggested above that hemispheric differences might have led to variation in the processing of auditory properties. There is also evidence to suggest that the processing of visual speech may be different in the two hemispheres (more lateralized to the right hemisphere). For instance, Grant et al. (2004a,b) have proposed that visual speech place of articulation cues are deployed across multiple segments over relatively long duration windows. Moreover Wright et al. (2003) have shown that cortical activation to visual speech occurs primarily in the right hemisphere (posterior STS). Likewise Diesch (1995) presented visual speech stimuli to different hemifields in a McGurk paradigm and argued that the results indicated that the auditory–visual integration process is more lateralized to the right hemisphere.

Taking the above auditory and visual processing characteristics in each hemisphere into account allows a more principled review of the pattern of AV N100m reduction. Consider the pattern of N100m reduction in the left hemisphere (weaker than the right with little change in the reduction from the Normal to Band stimuli). The overall weaker reduction effect could be due to the presumed overall weaker influence of visual speech in the left hemisphere. That the band manipulation did not affect the size of AV reduction in the left hemisphere might be because the left hemisphere responds more to the burst release cue for place, which is less affected by the Band manipulation. That is, although Shannon et al. (1995) have shown that place information is degraded by the band manipulation this disruption is more likely to affect the trans-segmental cue as it is spectral information that is degraded in the Band stimuli. The idea that the right hemisphere responds more to the trans-segmental place cue would also explain why there is a more graded reduction in the right hemisphere as a function of stimulus type (as this auditory cue for place is degraded in the Band and Noise stimuli). In summary then, the pattern of the reduction in the N100m response can be well described by taking into account possible difference in the hemispheric processing of the speech features jointly involved in auditory and visual processing and the effects of the stimulus manipulation.

The general idea in the gating/motor mediation view is that the N100m should be reduced when there is a mismatch between what could be expected on the basis of the visual signal and the auditory signal. Recently, Stekelenburg and Vroomen (2007) used pairs of congruent and incongruent AV speech pairings (auditory /fu/ combined with visual /bi/

⁴ We focus upon place cues here because visual speech typically provides information about place of articulation whereas manner information is often provided by auditory speech. However, it must be pointed out that place and manner are imperfectly related to visibility. Thus although only place cues are mentioned, we acknowledge that visual cues for manner might also be present.

⁵ One proposal that has provided estimates for the preferential time spans of processing in the left and right hemispheres is the Asymmetric Sampling in Time theory (Poeppel, 2003). In this account the auditory cortices preferentially sample at rates tuned to fundamental speech units: integration of auditory signals in the left auditory cortex preferentially occurs on approximately phoneme length segments (i.e., 20–50 ms) whereas the preferentially integration time period for the right auditory cortex is in the order of 100–300 ms (allowing sensitivity to slower acoustic modulations). Recent support for such asymmetric sampling comes from Giraud et al. (2007).

⁶ Note that we are not suggesting that division of hemispheric separation in the processing of these cues is complete; indeed, there is evidence that formant transition cues per se that occur over a period of 50 ms are processed in the left hemisphere and for some stimuli these can result in reduction of the N100m (Jääskeläinen et al., 2008). What we are suggesting is similar to Poeppel (2001) who proposed that processing of the initial formant transitions occurs in the left hemisphere whereas the processing of CV syllables occurs in the right hemisphere.

auditory /bi/ combined with visual /fu/) and it appeared that AV congruity of these pairs did not affect the amount of reduction in the N100m response. This finding suggests that the measure of the reduction in N100m is a rather coarse one (as /fu/ and /bi/ the items used, differ on numerous dimensions). However, before this could be concluded, the results of Stekelenburg and Vroomen would need to be broken down by left and right hemisphere sources, as the current results have shown that the left hemisphere was insensitive to differences in stimulus type, so a whole-head measure would mask any potential differences.

In conclusion, the results of the current study showed that whereas the size of the N100m was greater for AO stimuli in the left hemisphere, its reduction for AV stimulus was larger in the right. Furthermore, whereas this reduction was constant in the left hemisphere, it was graded in the right (occurring most for the stimuli that had spectral structure). Following Parviainen et al. (2005) and Jamison et al. (2006) it was suggested that the greater N100m amplitude in the left hemisphere for AO normal speech might be due to this hemisphere's sensitivity to phonetically relevant arrangements of acoustic features. On the other hand, it was proposed that the greater reduction in the right hemisphere for AV normal speech reflects this hemisphere's sensitivity to AV cues for place of articulation such that the auditory response was suppressed due to the match between the expected and received auditory speech signal.

4. Experimental procedures

4.1. Subjects

Fourteen subjects (mean age 22; 6 females) participated in the experiment. All subjects reported normal hearing and normal or corrected vision. All but one subject were native Finnish speakers. Informed consent was obtained from all subjects.

4.2. Stimuli

Two auditory and audio-visual syllables (/pi/, /ti/) were selected for the MEG analysis on the basis that they had similar amplitude envelopes (see Fig. 3). A third syllable /vi/ was selected as a target in the behavioural experiment as its visual

articulation was easy to distinguish. A colour digital video of a Finnish female producing these syllables was recorded (PAL format, 25 frames per second, sound sampling frequency 48,000 Hz). Auditory stimuli were presented binaurally through plastic tubes and earpieces at a sound intensity that was adjusted to be at a comfortable listening level (65–75 dB SPL).

Visual stimuli were back-projected onto a screen directly in front of the subject. The onset of preparatory visual articulation preceded the auditory stimulus onset by 20 ms for /pi/, 100 ms for /ti/ and 100 ms for /vi/. The view was frontal, including head and upper part of shoulders. Clips were digitized and presented as sequences of images. Images were cropped to remove facial features from and above the eyes and filtered to lessen the contribution of high spatial frequency edges. The contribution of the dark regions of the nostrils was contrast reduced (see Fig. 4).

Three versions of each auditory stimulus were constructed, 1) an unadjusted Normal version, 2) a Band version in which temporal and amplitude cues were preserved in four spectral bands (50–800 Hz, 800–1500 Hz, 1500–2500 Hz, 2500–4000 Hz) by using the amplitude envelope to modulate each band but the spectral detail within each band was replaced by white noise, i.e., white noise filtered with the same bandwidth as each frequency band (see Shannon et al., 1995), and 3) a Noise version in which the overall amplitude envelope was filled with white noise (see Schroeder, 1968). The peak stimulus intensities were then normalized. Two versions of each stimulus were constructed: An Auditory Only version (AO) and an Auditory–Visual version (AV). In the AO version, the subject was presented with a still frame from the video in which the talker's mouth was closed (top row Fig. 4).

4.3. Data recording and processing

The neuromagnetic responses were recorded in a shielded room with 306-channel whole-head MEG recording device (Neuromag, Vectorview) in the Low Temperature Lab of Helsinki University of Technology. MEG was recorded with a passband 0.03–200 Hz, digitized at 600 Hz and online averaged from 200 ms before till 400 ms after the onset of the auditory token. Eye movements were monitored with an electrode placed above and below right eye canthus and near the corners of left and right eyes. Trials containing large amplitude MEG or EOG signals (>3000 fT/cm and >150 μ V respectively) were not included in the

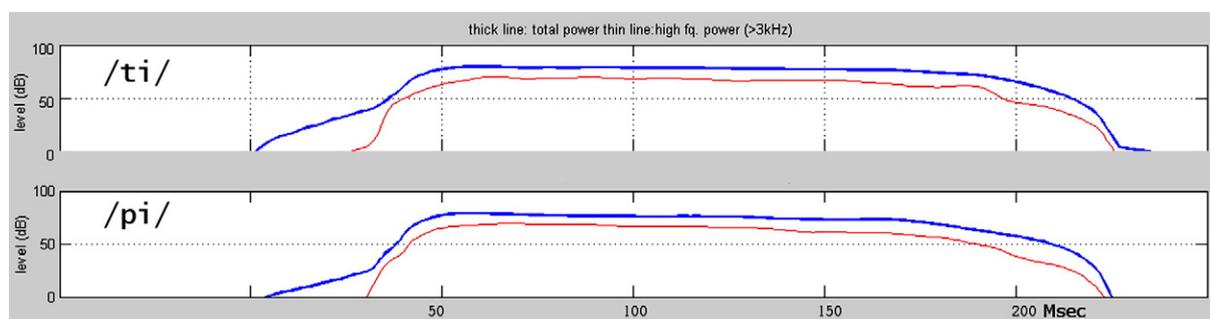


Fig. 3 – Power envelopes for the Normal /ti/ and /pi/ stimuli. The thicker upper curve represents total power and the thinner inner curve high frequency power (>3 kHz). Only the /ti/ and /pi/ stimuli (Normal, Band and Noise) were used in the MEG analysis.

averages. 120–190 responses were averaged for each stimulus category.

Averaged evoked responses were low-pass filtered at 45 Hz. For each subject, six gradiometer pairs showing maximal N100m responses were selected over each hemisphere. An example of the selection of channels for one subject (Normal stimuli, Auditory only presentation) is shown in Fig. 5. A vector sum was calculated for each six gradiometer pair, and these vector sums were averaged. After these operations, we had two spatially averaged responses over each hemisphere (Fig. 5). These averages were baselined from –100 to 0 ms, relative to the auditory stimulus onset.

N100m amplitude peaks were defined visually. Because this was not possible for all different stimuli in the data of five subjects (including three females), their data were not analysed. For the remaining nine subjects, N100m amplitudes were measured over 23-ms interval around the peak of N100m. These values were subjected to three-way ANOVA with factors Stimulus Type (Normal/Band/Noise), Presentation mode (AO/AV), Hemisphere (Left/Right). To compare the effect of processing visual speech on the N100m responses elicited by different types of auditory stimuli, this effect was

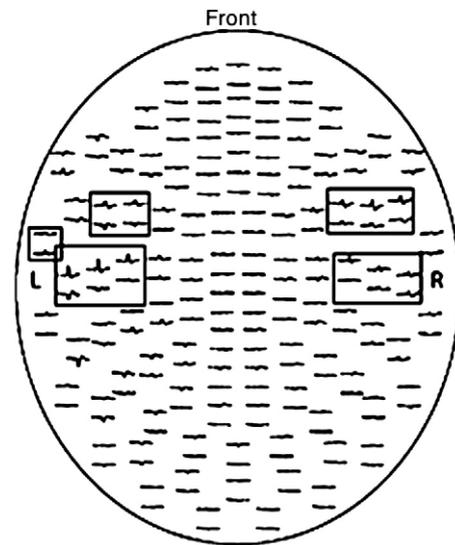


Fig. 5 – An example of the channels selected for one subject (Normal stimuli, Auditory Only presentation).

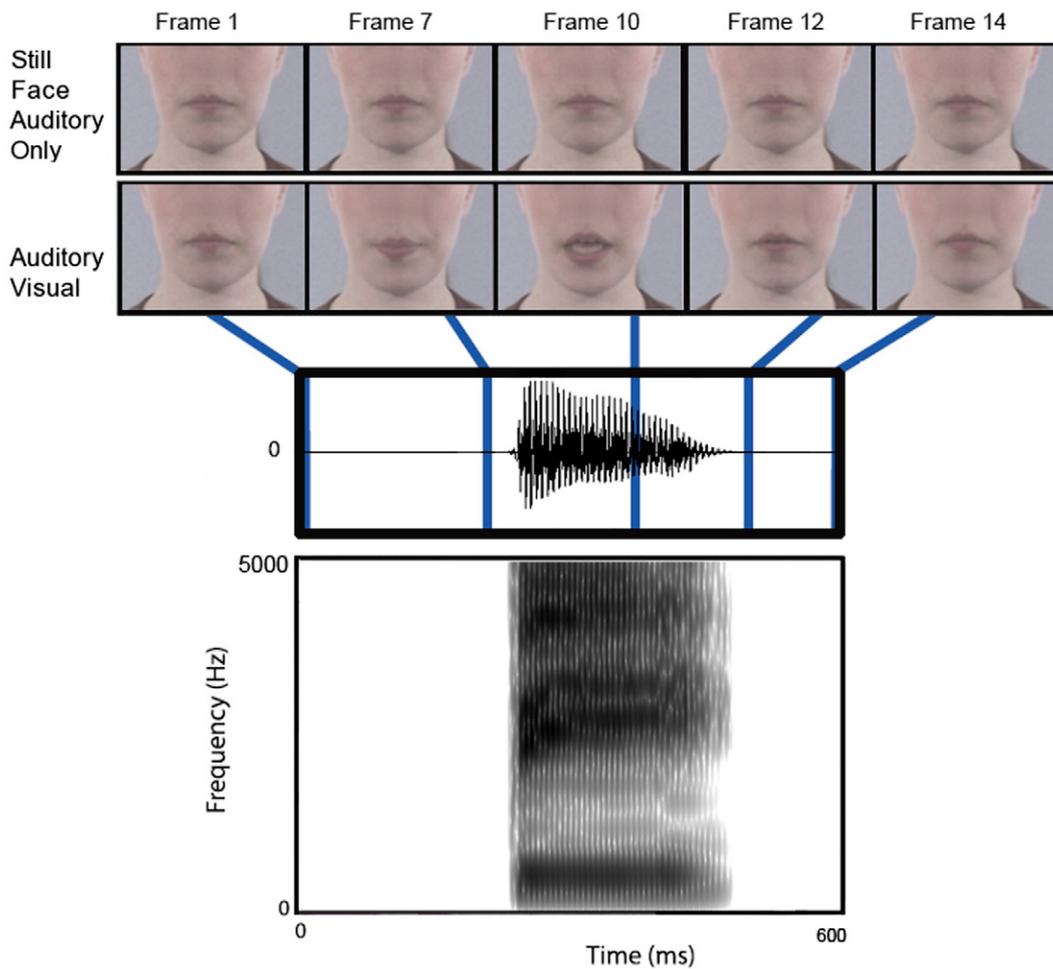


Fig. 4 – An example of one the stimuli (Normal /pi/). The top row shows the same initial frame displayed in AO condition. The next row shows five frames from throughout the video (from the first to last frame). The third row shows the corresponding acoustic speech waveform and the bottom row the corresponding spectrogram. The blue lines indicate the position of the corresponding video frames and the acoustic speech signal.

calculated using the following formula: $((AO - AV) / AO) * 100$. Obtained values were then subjected to ANOVA.

4.4. Procedure

Subjects were studied in a magnetically shielded room. Cortical activation elicited by the stimuli was measured using a whole-head 306-channel neuromagnetometer. The subject's head position with respect to the sensors was located by measuring the magnetic field produced by small currents delivered to three coils attached to the scalp. The location of the coils with respect to the nasion and periauricular points was determined by a three dimensional digitizer. The subjects' task was to detect the /vi/ syllables, which occurred approximately once in every eight trials and formed 12% of trials. Subjects signalled detection by lifting their right hand index finger. These trials were not used in MEG analysis. Different types of auditory stimuli were presented in the same block. The Noise stimuli were always presented first, and then the Band stimuli and Normal stimuli were the last (this order of presenting the reduced speech version first was used in order to minimize the effects of stimulus repetition). Inter-trial intervals varied pseudo-randomly from 355 to 1006 ms. Within each block, AO and AV stimuli were presented in a random order. Stimuli were presented using the Presentation display software. Stimuli were presented until more than 120 valid MEG trials were collected for each AO and AV stimuli. The experiment paused after 10–20 stimuli and continued when the subject lifted their finger. To make sure that subjects paid attention to the visual face stimulus, they were instructed to monitor for the presence of an infrequent small and faint white dot that appeared briefly on the face (just below the nose). The detection of the white dot was signalled by a finger lift. The white dot appeared in approximately 3% of the stimuli. These trials were not included in MEG averages.

Acknowledgments

The authors wish to acknowledge support from Australia Research Council Grant number DP DP0666857 and the Academy of Finland (grant numbers 2103470, 2103938).

REFERENCES

- Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2008. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J. Neurosci.* 28, 3958–3965.
- Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.C., Samson, Y., 1998. Lateralization of speech and auditory temporal processing. *J. Cogn. Neurosci.* 10, 536–540.
- Besle, J., Fort, A., Delpuech, C., Giard, M.-H., 2004. Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.* 20, 2225–2234.
- Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* 8, 389–395.
- Callan, D.E., Jones, J.A., Munhall, K., Callan, A.M., Kroos, C., Vatikiotis-Bateson, E., 2003. Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport* 14, 2213–2218.
- Calvert, G.A., Campbell, R., 2003. Reading speech from still and moving faces: the neural substrates of visible speech. *J. Cogn. Neurosci.* 15, 57–70.
- Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., Brammer, M.J., David, A.S., 2001. Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Cogn. Brain Res.* 12, 233–243.
- Diesch, E., 1995. Left and right hemifield advantages of fusions and combinations in audiovisual speech perception. *Q. J. Exp. Psychol.* 48A, 320–333.
- Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., Logothetis, N.K., 2005. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J. Neurosci.* 25, 5004–5012.
- Giard, M.H., Peronnet, F., 1999. Auditory–visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J. Cogn. Neurosci.* 11, 473–490.
- Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T.E., Frackowiak, R.S.J., Lauff, H., 2007. Endogenous Cortical Rhythms Determine Cerebral Specialization for Speech Perception and Production. *Neuron* 56, 1127–1134.
- Grant, K.W., Greenberg, S., Poeppel, D., van Wassenhove, V., 2004a. Effects of spectro-temporal asynchrony in auditory and auditory–visual speech processing. *Semin. hear.* 25, 241–255.
- Grant, K.W., van Wassenhove, V., Poeppel, D., 2004b. Detection of auditory (cross-spectral) and auditory–visual (cross-modal) synchrony. *Speech commun.* 44, 43–53.
- Greenberg, S., Arai, T., 2004. What are the essential cues for understanding spoken language? *IEICE trans. inf. syst.* 87, 1059–1070.
- Hazan, V., Rosen, S., 1991. Individual variability in the perception of cues to place contrasts in initial stops. *Percept. Psychophys.* 49, 187–200.
- Heinks-Maldonado, T.H., Nagarajan, S.S., Houde, J.F., 2006. Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport* 17, 1375–1379.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Hutsler, J., Galuske, R.A.W., 2003. Hemispheric asymmetries in cerebral cortical networks. *Trends Neurosci.* 26, 429–435.
- Hyde, K., Peretz, I., Zatorre, R., 2008. Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* 46, 632–639.
- Jääskeläinen, I.P., Ojanen, V., Ahveninen, J., Auranen, T., Levänen, S., Möttönen, R., Tarnanen, I., Sams, M., 2004. Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. *Neuroreport* 15, 2741–2744.
- Jääskeläinen, I., Kauramäki, J., Tujunen, J., Sams, M., 2008. Formant transition-specific adaptation by lipreading of left auditory cortex N1m. *Neuroreport* 19, 93–97.
- Jamison, H.L., Watkins, K.E., Bishop, D.V.M., Matthews, P.M., 2006. Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb. Cortex* 16, 1266–1275.
- Klucharev, V., Möttönen, R., Sams, M., 2003. Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cogn. Brain Res.* 18, 65–75.
- Kuriki, S., Okita, Y., Hirata, Y., 1995. Source analysis of magnetic field responses from the human auditory cortex elicited by short speech sounds. *Exp. Brain Res.* 104, 144–152.
- Liégeois-Chauvel, C., Giraud, K., Badier, J.M., Marquis, P., Chauvel, P., 2001. Intracerebral evoked potentials in pitch perception reveal a functional asymmetry of human auditory cortex. *Ann. N.Y. Acad. Sci.* 930, 117–132.
- Liikkanen, L.A., Tiitinen, H., Alku, P., Leino, S., Yrttiaho, S., May, P.J.C., 2007. The right-hemispheric auditory cortex in humans is sensitive to degraded speech sounds. *Neuroreport* 18, 601–605.

- Lutkenhöner, B., Steinsträter, O., 1998. High-precision neuromagnetic study of the functional organization of the human auditory cortex. *Audiol. Neurotol.* 3, 191–213.
- Martikainen, M.H., Kaneko, K., Hari, R., 2005. Suppressed Responses to Self-triggered Sounds in the Human Auditory Cortex. *Cereb. Cortex* 15, 299–302.
- Miki, K., Watanabe, S., Kakigi, R., 2004. Interaction between auditory and visual stimulus relating to the vowel sounds in the auditory cortex in humans: A magnetoencephalographic study. *Neurosci. Lett.* 357, 199–202.
- Möttönen, R., Krause, C.M., Tiippana, K., Sams, M., 2002. Processing of changes in visual speech in the human auditory cortex. *Brain Res. Cogn. Brain Res.* 13, 417–425.
- Näätänen, R., Picton, T.W., 1987. The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology* 24, 375–425.
- Näätänen, R., Winkler, I., 1999. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.* 125, 826–859.
- Nakasato, N., Fujita, S., Seki, K., Kawamura, T., Matani, A., Tamura, I., Fujiwara, S., Yoshimoto, T., 1995. Electroencephalogr. Clin. Neurophysiol. 94, 183–190.
- Nicholls, M.E.R., 1996. Temporal Processing Asymmetries between the Cerebral Hemispheres: Evidence and Implications. *Laterality* 1, 97–137.
- Nicholls, M.E.R., Schier, M., Stough, C.K.K., Box, A., 1999. Psychophysical and electrophysiologic support for a left hemisphere temporal processing advantage. *Neuropsychiatr. Neuropsychol. Behav. Neurol.* 12, 11–16.
- Olson, I.R., Gatenby, J.C., Gore, J.C., 2002. A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cogn. Brain Res.* 14, 129–138.
- Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., Elbert, T., 1990. Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalogr. Clin. Neurophysiol.* 94, 26–40.
- Parviainen, T., Helenius, P., Salmelin, R., 2005. Cortical differentiation of speech and nonspeech sounds at 100 ms: Implications for dyslexia. *Cereb. Cortex* 15, 1054–1063.
- Poeppl, D., 2001. Pure word deafness and the bilateral processing of the speech code. *Cogn. Sci.* 25, 679–693.
- Poeppl, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* 41, 245–255.
- Rogers, R.L., Papanicolaou, A.C., Baumann, S.B., Saydjari, C., Eisenberg, H.M., 1990. Neuromagnetic evidence of a dynamic excitation pattern generating the N100 auditory response. *Electroencephalogr. Clin. Neurophysiol.* 77, 237–240.
- Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., Foxe, J.J., 2007. Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia* 45, 587–597.
- Sams, M., Aulanko, R., Hamalainen, 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. *Neurosci. Lett.* 127, 141–145.
- Sandmann, P., Eichele, T., Specht, K., Jäncke, L., Rimol, L.M., Nordby, H., Hugdahl, K., 2007. Hemispheric asymmetries in the processing of temporal acoustic cues in consonant-vowel syllables. *Restor. Neurol. Neurosci.* 25, 227–240.
- Schönwiesner, M., Rübsamen, R., Yves von Cramon, D., 2005. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur. J. Neurosci.* 22, 1521–1528.
- Schroeder, M.R., 1968. Reference signal for signal quality studies. *J. Acoust. Soc. Am.* 44, 1735–1736.
- Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J.S., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406.
- Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. *Science* 270, 303–304.
- Skipper, J.I., van Wassenhove, V., Nusbaum, H.C., Small, S.L., 2007. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb. Cortex* 17, 2387–2399.
- Stekelenburg, J.J., Vroomen, J., 2007. Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.* 19, 1–10.
- Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R., Näätänen, R., 2000. Discrimination of speech and complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage* 12, 657–663.
- Sumby, W.H., Pollock, I., 1954. Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215.
- Trébuchon-Da Fonseca, A., Giraud, K., Badier, J.-M., Chauvel, P., Liégeois-Chauvel, C., 2005. Hemispheric lateralization of voice onset time (VOT) comparison between depth and scalp EEG recordings. *NeuroImage* 27, 1–14.
- van Wassenhove, V., Grant, K.W., Poeppel, D., 2005. Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1181–1186.
- Watkins, K., Paus, T., 2004. Modulation of motor excitability during speech perception: The role of Broca's Area. *J. Cogn. Neurosci.* 16, 978–987.
- Wright, T.M., Pelphrey, K.A., Allison, T., McKeown, M.J., McCarthy, G., 2003. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb. Cortex* 13, 1034–1043.
- Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953.