The Role of Speech Production System in Audiovisual Speech Perception

Iiro P. Jääskeläinen*

Department of Biomedical Engineering and Computational Science, Aalto University, Espoo, Finland

Abstract: Seeing the articulatory gestures of the speaker significantly enhances speech perception. Findings from recent neuroimaging studies suggest that activation of the speech motor system during lipreading enhance speech perception by tuning, in a top-down fashion, speech-sound processing in the superior aspects of the posterior temporal lobe. Anatomically, the superior-posterior temporal lobe areas receive connections from the auditory, visual, and speech motor cortical areas. Thus, it is possible that neuronal receptive fields are shaped during development to respond to speech-sound features that coincide with visual and motor speech cues, in contrast with the anterior/lateral temporal lobe areas that might process speech sounds predominantly based on acoustic cues. The superior-posterior temporal lobe areas have also been consistently associated with auditory spatial processing. Thus, the involvement of these areas in audiovisual speech perception might partly be explained by the spatial processing requirements when associating sounds, seen articulations, and one's own motor movements. Tentatively, it is possible that the anterior "*what*" and posterior "*where / how*" auditory cortical processing pathways are parts of an interacting network, the instantaneous state of which determines what one ultimately perceives, as potentially reflected in the dynamics of oscillatory activity.

Keywords: Audiovisual speech perception, speech motor theory, functional MRI, magnetoencephalography, electroencephalography.

INTRODUCTION

Speech perception is not limited to hearing, as seeing the gestures and lip forms of a speaker significantly enhance speech perception, especially under noisy conditions [1]. It has been widely assumed that this effect follows the so-called law of inverse effectiveness, that is, that the effects of visual stimuli are greatest when the auditory input is weakest. Recently, however, it was shown that the integration effects are most robust at intermediate signal-to-noise ratios where more than a three-fold improvement in performance was observed relative to the auditory alone condition, possibly due to reliance on the cues provided by the more salient modality when one of the inputs is too severely degraded [2, 3]. Furthermore, it also matters what is being said: movements of the lips and jaw, and the position of the tongue in the mouth, can yield highly accurate information on certain speech sounds, even more accurate than the auditory input itself, but other speech sounds are very difficult to discern from each other based on articulatory gestures alone, resulting in that there is considerable variability in the lip-readability of sentences [4]. Indeed, there are fewer visemes (the basic constituents of visual speech comparable to phonemes) than there are phonemes.

In addition to lipreading or, as it sometimes is referred to, speech-reading, enhancing hearing, a variety of audio-visual illusions have been reported under artificial conditions where the visual stimuli are not congruent with the auditory input. The ventriloquism effect, where the ventriloquist moves the lips of a hand doll while speaking with his own lips closed to create an illusion of the doll speaking, is perhaps the most commonplace of these phenomena. Another quite striking phenomenon is the so-called McGurk effect, where seeing and hearing incongruent phonetic sounds results in an illusory third-category phonetic percept. In the classical study by McGurk and MacDonald [5], seeing /ga/ and hearing /ba/ resulted in the subjects hearing /da/. Interestingly, the converse, visual /ba/ and auditory /ga/, did not result in the subjects perceiving /da/, but rather in a dual percept of /bga/ [5]. Further, the McGurk effect is highly automatic, occurring even when the subjects are informed of the mismatch in the visual and auditory inputs, as well as when a female voice is dubbed with a male face and *vice versa* [6]. These aspects of the McGurk effect suggest that visual stimuli can modulate processing of auditory information at a relatively early stage. Indeed, the three major questions that subsequent neuroimaging studies have attempted to address are 1) the anatomical locations where visual inputs can modulate auditory processing, 2) the latencies at which the effects take place, and 3) the precise mechanisms of interactions through which the visual stimuli can alter phonetic percepts. In the following, cognitive neuroimaging findings pertaining to these questions are reviewed, followed by a synthesis of the possible network of cerebral events that underlies audiovisual speech perception.

SUPERIOR-POSTERIOR TEMPORAL LOBE ACTIVATIONS DURING AUDIOVISUAL STIMULATION

One of the first neuroimaging studies addressing the question of in which cortical areas and at what latency visual information has access to the auditory system was a magnetoencephalography (MEG) study where healthy volunteers were presented with auditory /pa/ syllables

*Address correspondence to this author at the Department of Biomedical Engineering and Computational Science, Aalto University, P.O. Box 12200, FIN-00076 Aalto, Finland; Tel: +358 9 47001; Fax: +358 9 470 24833;
E-amil: iiro.jaaskelainen@tkk.fi

1874-4400/10

together with short video clips of /pa/ vs. /ka/. MEG responses specific to the auditory /pa/ and visual /ka/ combination, which produced percepts of /ta/ or /ka/, were estimated to originate in the superior and posterior aspects of the temporal lobe at a relatively early latency of 150-200 ms from stimulus onset [7]. Subsequent MEG [8, 9] and EEG [10-14] work has confirmed that lipreading modulates auditory-cortical responses to speech sounds as early as ~100 ms from stimulus onset, while visual motion related non-specific suppression has been observed as early as 50 ms from sound onset [14]. At 50 and 100 ms latencies the auditory evoked responses are presumed to be mostly originating from the primary and secondary auditory cortical areas posterior to the primary auditory cortex, respectively (for a review, see [15]). In a recent study, lipreading /ga/ significantly suppressed left-hemisphere MEG responses at ~100 ms latency to the F2 transition contained in /ga/, as compared with lipreading of /ba/ (see Fig. **1**), suggesting that visual speech selectively adapts posterior auditory cortex neural populations tuned to formant transitions [9]. As the formant transitions are the elementary sound-sweep constituents of phonemes, these results suggest that the effects of lipreading can be highly specific and occur at a relatively early level of sound feature processing.

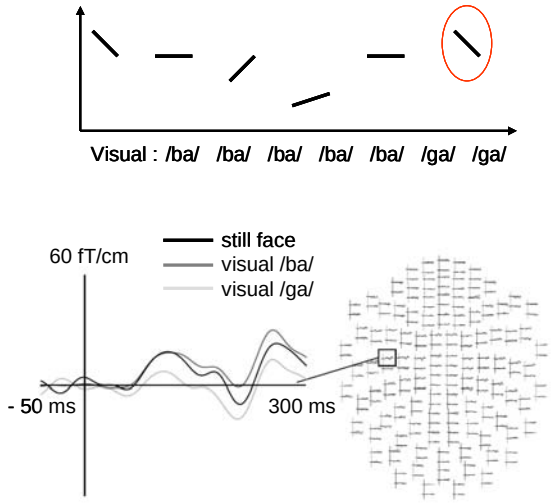

Fig. (1). Lipreading suppresses auditory cortex ~100 ms responses speech sound formant specifically. TOP: Sinusoidal sound sweep analogs of the first formant transition common to /ba/, /ga/, and /da/ sounds, and a continuum of second-formant transitions ranging from the second formant sweep contained in /ba/ to that contained in /ga/ were presented to subjects while they were watching a sequence of short videoclips of a person articulating either /ba/, /ga/, or a still-face control picture. BOTTOM: Comparison of MEG responses to the second-formant sound sweep contained in /ga/ in the still-face baseline, /ba/, and /ga/ lipreading conditions disclosed suppressed responses ~100 ms from sound onset when the subjects were lipeading /ga/ (adapted with permission from [9]).

Functional magnetic resonance imaging (fMRI) studies have suggested that silent lipreading can activate the primary auditory cortex [16, 17] (although see also [18]), and findings of modulation of EEG brain stem responses during lipreading suggest that there might be hierarchically even lower-level effects [19]. Tentatively, it is possible that the primary auditory cortex activations resulted from modulation of the receptive field properties of the auditory cortical neurons thus leading to the ongoing fMRI scanner noise eliciting differential activations during lipreading vs. baseline conditions (see [20]). The primary auditory cortex activations are not, however, as robust as the secondary auditory cortex activations during silent lipreading [16, 18, 21]. Further, given the limited temporal resolution of the hemodynamic responses that fMRI measures, it was not possible to determine whether the primary auditory cortex activations were due to back-projections from hierarchically higher secondary auditory or heteromodal cortical areas, or whether they were due to direct connections that have been described between the visual cortex and primary auditory cortex in monkey [22]. There are findings suggesting that at least some of the visual stimulus influence is due to direct connections from the visual to auditory cortical areas (for a review, see [23]), but it is likely that direct visual-to-auditory cortical connections do not fully explain the influence of lipreading on speech perception. For a recently published summary of the multisensory inputs and outputs of the auditory cortex in the monkey, see [24].

There are several fMRI studies that have mapped cortical areas exhibiting either larger or smaller hemodynamic responses to audiovisual stimuli as compared with the sum of auditory and visual unimodal stimuli in an attempt to identify loci of multimodal sensory integration. It can be surmised that neurons responding to both auditory and visual stimuli are weakly driven when near-threshold auditory and visual stimuli are presented unimodally, whereas bimodal presentation results in the inputs exceeding the threshold, resulting in that bimodal stimuli elicit supra-additive responses compared with responses to unimodal stimuli. On the other hand, when using clearly audible/visible stimuli, unimodal inputs can also drive the multimodal neurons and thus the sum of unimodal responses exceeds the responses elicited by bimodal stimulation. The first of such neuroimaging studies disclosed supra-additive hemodynamic responses to continuous audiovisual speech in the posterior part of the left superior temporal sulcus (pSTS) [25]. The finding of pSTS being involved in audiovisual speech integration has been replicated in subsequent fMRI [26-37], positron emission tomography (PET) [38], and intracranial EEG [39] studies (although see also [40]). Corroborating findings have also been obtained in non-human primate studies, where visual stimuli were found to specifically modulate caudal auditory cortical fields [41]. Further, much like what has been observed in animal studies [42], there appears to be patches of cortex adjacent to each other within the human pSTS responding to auditory, visual, and audiovisual stimulation [31].

What makes the audiovisual posterior STG/STS activations especially relevant is that they correlate with perception: posterior STS activity was observed when temporally offset auditory and visual stimuli were perceptually fused [32], and in a recent study STG and STS response magnitudes were found to significantly correlate with the perceptual magnitude of the McGurk illusion [43]. A recent study showed, using intracranial recordings in epileptic patients, that the audiovisual integration effects can take place as quickly as 30 ms from sound onset in the secondary auditory cortical areas [44], even though it is

unlikely that such early interaction effects could account for audiovisual integration effects at the phonetic level, as the McGurk illusion has been reported to tolerate a relatively wide degree of asynchrony between the auditory and visual speech stimuli, with fusion percepts arising when stimulus asynchronies ranged from −34 ms auditory leading the visual stimulus to +173 ms auditory lagging the visual stimulus [45]. It is possible this is due to there being, under natural conditions, variable delays from the visual to the auditory stimuli, depending on the type and context of a given articulatory gesture, as well as the physical distance between the speakers.

Another highly interesting question to which neuroimaging studies have attempted to address is which aspects of the seen articulations influence auditory speech processing. When the visual stimuli were wavelet-filtered to specifically retain information on the place of articulation, specific activations were observed in the middle temporal gyrus (MTG), pSTS, and posterior superior temporal gyrus (pSTG) in an fMRI study [30] (see Fig. **2**). Place of articulation cues have also been suggested to underlie the suppression of auditory cortical MEG responses ~100 ms from stimulus onset [46]. In behavioral studies, the onset of opening of the vocal tract in the videoclip was observed to trigger the verbal transformation [47] and McGurk [48] effects. Further, the highly dynamic auditory information contained in the second and higher formants have been specifically implicated in the McGurk effect [48].

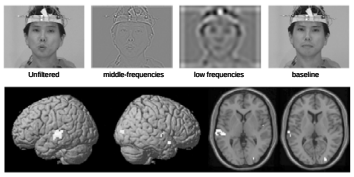

Fig. (2). Seeing the place of articulation enhances activations to speech sounds in the superior-posterior temporal lobe. TOP: The unfiltered articulating face contained all visual speech gesture motion, the spatial midfrequency wavelet band-pass filtered condition maintained the place of articulation information, and the spatial low-frequency wavelet band-pass filtered condition consisted of gross properties of movement of the lips, jaw, and head. BOTTOM: Sites of multi-sensory integration selectively induced by auditory and visual correspondence of place of articulation information, revealed by contrasting the activity during both the middle-frequency and unfiltered conditions with the activations during the low-frequency condition, were localized predominantly to the left middle temporal gyrus, left superior temporal sulcus, and left superior temporal sulcus (adapted with permission from [30]).

The importance of synchronous timing of auditory and visual stimulus presentation has also been documented in non-human primate studies. Multisensory enhancements were observed in the primary and secondary auditory cortical local field potentials when species-specific vocalizations and video-clips of associated articulatory gestures were presented to non-human primates in synchrony, whereas suppression of responses abounded with delayed voice-onset times [49]. A subsequent study disclosed timing-dependent multisensory integration effect specifically in the LFP alpha-frequency band in the monkey STS [50]. Thus, it is possible that the visually salient onset events in seen articulations cause rapid enhancements in the posterior auditory cortical areas, followed by post-stimulus inhibition, which can be even more important for audiovisual integration than the initial excitation [20, 51]. It has also been recently proposed that the articulatory gestures phase-reset the ongoing auditory cortical oscillatory activity, thus enhancing processing of the auditory stimulus especially when listening to continuous speech [52].

IS THE SUPERIOR POSTERIOR TEMPORAL LOBE PART OF A SPEECH MOTOR PROCESSING PATHWAY?

Taken together, these findings suggest that the posterior temporal lobe areas are pivotal for integration of visual information with auditory inputs to facilitate, or in the case of artificially mismatched visual and auditory inputs, distort, speech perception. This is in somewhat of a contrast with findings from neuroimaging studies of speech perception that have consistently shown speech-sound specific activations in areas anterior and lateral to the primary auditory cortex [53-55], while the posterior temporal lobe areas have been shown to be more involved in sound location processing than with processing of speech sounds. Studies directly contrasting selectivity of anterior *vs*. posterior auditory cortex to spatial location of the auditory stimuli vs. species specific vocalizations (in non-human primates) and speech sounds (in humans) showed that the anterior auditory cortex is selective to speech/vocalization sounds and the posterior auditory cortex is selective to spatial locations [56-59]. Contrasting these findings, there are studies that have found evidence for speech specific responses in the posterior temporal areas [60-62]. While studies comparing responses to speech *vs*. non-speech sounds could always be criticized on the grounds that it is possible that physical differences between the sounds could have caused the differential activations, activation of the left pSTS was also observed when subjects perceived physically identical sine-wave speech stimuli as speech *vs*. meaningless noise in a recent fMRI study [62].

To account for this discrepancy in findings, it has been proposed that the posterior temporal areas would not be restricted to spatial processing, but might constitute a parallel stream processing also certain aspects of speech [63-66]. Specifically, it has been proposed that the dorsal stream would be responsible for mapping speech sounds onto articulatory-based representations, or “doable” articulations and sounds [63-65], thus suggesting that the competing acoustic-perceptual [67, 68] and motor theories [69, 70] of speech perception would both hold true, but manifested in parallel and complementary processing pathways. Given the abundance of findings linking audiovisual speech perception with the posterior speech-motor processing pathway, it is easy to come up with a hypothesis in which the visual information during lipreading would be merged with information/knowledge of the speech motor schemes *via* a mirroring type of process, as well as with the associated speech sounds. Findings from developmental studies, as well as the way that anatomic connections are organized in the brain, yield support for this interpretation.

DEVELOPMENT OF AND CONNECTIVITY SUPPORTING AUDIOVISUAL SPEECH PERCEPTION

Learning to perceive audiovisual speech occurs relatively fast after birth. It has been reported that infants can perceive correspondence between speech sounds and articulatory gestures, and also imitate speech sounds that are presented to them, already at the age of 18 to 20 weeks [71]. Further, at this age infants have also been observed to exhibit the McGurk illusion [72]. It is quite characteristic for infants to pay attention to facial stimuli [73]. As a result, there are plentiful of occasions in which auditory and visual speech cues take place simultaneously. Further, given that it is also characteristic for infants to imitate facial gestures, already as early as at the age of 12-21 days [74], it is likely that there is associated motor system activity occurring more or less simultaneously as the auditory and visual speech cues. Following from the principles of Hebbian learning [75], such a pattern of simultaneously occurring firing across the auditory, visual, and motor systems should result in ensembles of neurons that are sensitive to acoustic features that co-occur with certain salient visual and speech motor cues (although see also [51]). Supporting the role of speech motor system in speech perception learning, subtle deficits in phonetic/word learning have been observed in children suffering from dys/anarthia [76].

The posterior temporal lobe areas constitute a candidate area for convergence of auditory, visual, and motor system inputs, given that it lies at the intersection of auditory and visual cortical areas, and that it receives motor connections *via* the *arcuate fasciculus*. Especially the close proximity to the posterior temporal areas of the visual middle temporal area (MT) that is devoted to visual motion processing is important as it is vital for the processing of moving lips. Further, it is possible that the spatial processing requirements involved in the task of associating sounds, seen articulations, and one's own motor movements play a role in audio-visuo-motor integration, given the prominent role of posterior auditory cortex in spatial processing [56-59]. While the anterior temporal lobe also receives connections from the prefrontal speech motor areas *via* the *uncinate fasciculus*, the anterior temporal lobe areas are further away from sensory visual processing areas such as the area MT. Indeed, connectivity across a wider network of cortical areas is vital for the formation of the posterior temporal cortex speech motor processing qualities. Consistent with this, audiovisual stimuli have been, in addition to posterior temporal lobe, consistently reported to activate the speech motor system and parietal areas (see Fig. **3**). These findings are reviewed below.

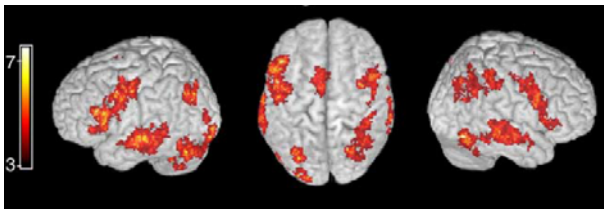

Fig. (3). An example of activation of speech motor areas, including Broca's area, motor cortex, and parietal cortical areas, during audiovisual speech perception in a recent study (adapted with permission from [79]).

THE SPEECH MOTOR SYSTEM IS ACTIVATED DURING AUDIOVISUAL SPEECH PERCEPTION

The brain areas that have been found to be activated during lipreading / audiovisual stimulus processing in addition to the posterior STG/STS include the inferior-lateral prefrontal cortex corresponding to the Broca's area [28, 32-34, 77-81], motor cortex [33, 34, 43, 77, 81], somatosensory cortex [33, 34], posterior parietal cortex [32, 77, 80, 81], claustrum [27], and insular cortex [32, 43]. Taken together these studies indicate that the speech motor system, consisting of Broca's area and the motor cortex and closely linked with the somatosensory and posterior parietal areas, participates in audiovisual speech perception. The specific role and the relative timing of the speech motor system in shaping posterior STG/STS activity during audiovisual speech perception has been a topic of intense study.

There are MEG findings suggesting that the speech motor area activations occur during silent lipreading at a significantly longer latency than the posterior STG/STS activations [77], thus raising the question of whether the speech motor area activity is secondary to the perceptual process itself—for instance, one could speculate whether the activation of the speech motor system was due to the subjects subvocally rehearsing the stimuli that were presented to them, something that is, however, not supported by findings of a lack of facial EMG activity in experimental designs eliciting robust speech motor system activations during audiovisual stimulation [79]. Further, in trans-cranial magnetic stimulation (TMS) studies, it has been shown that reducing the excitability of motor cortex disrupts phonetic categorization [82], and that this effect was recently shown to be articulator-specific [83]. It is also possible that the use of still rather than dynamic face stimuli [77] affected the MEG results, as this has been subsequently shown to potentially affect how audiovisual speech is processed [78]. Further, the MEG inverse estimates are always subject to certain degree of localization uncertainty and between-source cross-talk given the ill-posed nature of the electromagnetic inverse problem [84].

In a recent fMRI study, motor cortical and posterior parietal activations were observed to correlate, in addition to STG/STS, with the perceptual magnitude of the McGurk illusion [43], suggesting that the speech motor system does play a crucial role in how visual stimuli modulate speech perception. These findings corroborate and extend earlier findings of parietal and inferior frontal activations when temporally offset auditory and visual stimuli are perceptually fused [32]. Perhaps one of the most compelling findings speaking for the crucial role of speech production areas in audiovisual speech perception comes from a recent fMRI study (see Fig. **4**), where activity patterns in frontal areas, resulting from the illusory /ta/ percept that was produced by simultaneously presented auditory /pa/ and visual /ka/, were more similar to the activity patterns evoked by audiovisually congruent /ta/ than they were to patterns evoked by congruent audiovisual /pa/ or /ka/ [34]. In contrast, the activity elicited by auditory /pa/ combined with visual /ka/ initially resembled in posterior auditory and visual cortical areas the activity evoked by congruent audiovisual /pa/ and /ka/ stimuli, respectively [34]. At a longer latency, the activity patterns in the posterior temporal and visual cortical

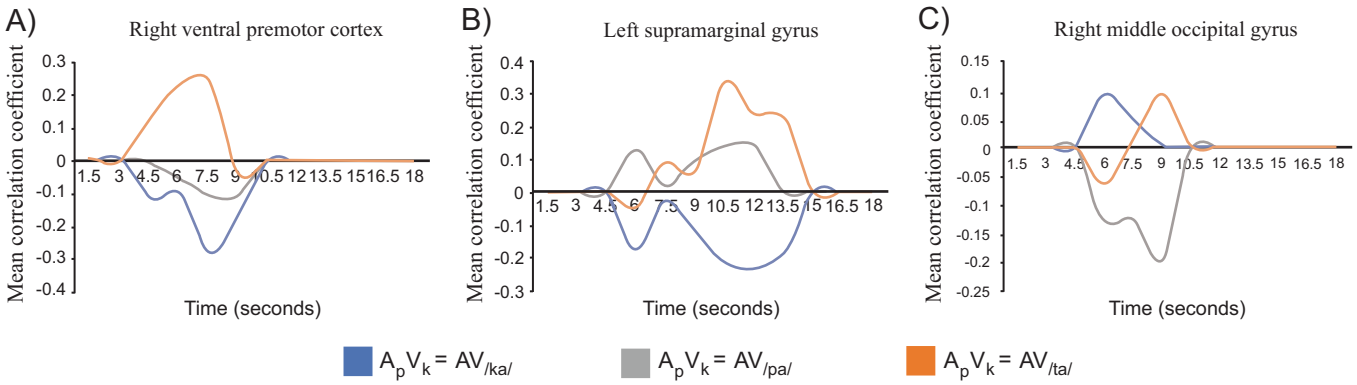

Fig. (4). Correlations as a function of time of the distributions of activations elicited by incongruent audiovisual syllable (auditory /pa/ and a visual /ka/) with the distributions of activation to congruent audiovisual /pa/ (gray), /ka/ (blue), /ta/ (orange) in **A**) ventral premotor cortical areas, **B**) left supramarginal gyrus, and (**C**) visual cortical areas. Note that the activity patterns in premotor areas correlated to those elicited by the /ta/ at a shorter latency than in the temporo-parietal and visual cortical areas, suggesting that that there was an efference copy from the speech motor system that shapes phonetic perception at the sensory-cortical level (adapted with permission from [34]).

areas became to resemble the activity pattern elicited by audiovisual /ta/, suggesting that there was an efference copy from the speech motor system that shaped phonetic perception at the sensory-cortical level [34]. The finding of fast access of visual information to prefrontal cortex, followed by a feedback to sensory areas has also been suggested by combined fMRI and MEG studies of object recognition [85]. Further, supporting this line of thinking, recent MEG findings were interpreted as suggesting that visual inputs are converted to motor-linguistic representations prior to their fusion with auditory information [14].

CONCLUDING REMARKS

Visual speech cues, lip forms, the position of the jaw, and the position of the tongue in mouth, significantly affect how information through the auditory system is processed in the brain, with congruent audiovisual stimuli improving speech perception, and incongruent stimuli in certain cases resulting in distorted percepts. It appears that visual speech cues, especially cues on the place of articulation, affect auditory processing. Visual-auditory interactions take place even at the level of the primary auditory cortex and/or auditory brain stem, which might play a role in how auditory stimuli are filtered to enhance speech perception. The vast majority of human neuroimaging studies have, however, implicated the posterior STG/STS as the area where articulatory gestures speech-specifically modulate auditory processing, even at surprisingly short latencies of a few tens to one hundred milliseconds. Given the hypothesis that there are two parallel speech processing pathways, an anterior pathway presumably devoted to processing speech based on acoustic cues, and a posterior pathway where neuronal receptive fields are shaped by simultaneous visual, auditory, and speech motor cues, it remains an open question where in the brain speech percepts ultimately emerge. From lesion studies it is well known that patients with damage to the speech motor system can retain their speech comprehension abilities, while patients with lesions of the temporal cortex often experience severe problems comprehending speech. Thus, it is unlikely that the speech motor system is where the conscious speech percepts form, but rather it is possible that the speech motor system activations caused by lipreading modulate the posterior temporal cortex activity patterns. Tentatively, it is possible that the anterior and posterior processing pathways are parts of an interacting network [86], the instantaneous state of which determines what one ultimately perceives, as potentially reflected in the dynamics of oscillatory activity [52, 87].

ACKNOWLEDGEMENT

Financially supported by the Academy of Finland.

REFERENCES

[1] Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. J Acoust Soc Am 1954; 26: 212-15.

[2] Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb Cortex 2007; 17: 1147-53.

[3] Ma WJ, Zhou X, Ross LA, Foxe JJ, Parra LC. Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. PLoS One 2009; 4: e4638.

[4] MacLeod A, Summerfield AQ. Quantifying the contribution of vision to speech perception in noise. Br J Audiol 1987; 21: 131-41.

[5] McGurk H, MacDonald J. Hearing lips and seeing voices. Nature 1976; 264: 746-48.

[6] Green KP, Kuhl PK, Meltzoff AN, Stevens EB. Integrating speech information across talkers, gender, and sensory modality: female faces and male voices in the McGurk effect. Percept Psychophys 1991; 50: 524-36.

[7] Sams M, Aulanko R, Hamalainen M, *et al*. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neurosci Lett 1991; 127: 141-45.

[8] Jaaskelainen IP, Ojanen V, Ahveninen J, *et al*. Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. Neuroreport 2004; 15: 2741-44.

[9] Jaaskelainen IP, Kauramäki J, Tujunen J, Sams M. Formant transition-specific adaptation by lipreading of left auditory cortex N1m. Neuroreport 2008; 19: 93-7.

[10] Klucharev V, Mottonen R, Sams M. Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. Brain Res Cogn Brain Res 2003; 18: 65-75.

[11] Besle J, Fort A, Delpuech C, Giard MH. Bimodal speech: early suppressive visual effects in human auditory cortex. Eur J Neurosci 2004; 20: 2225-34.

[12] van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci USA 2005; 102: 1181-86.

[13] Kislyuk DS, Möttönen R, Sams M. Visual processing affects the neural basis of auditory discrimination. J Cogn Neurosci 2008; 20: 2175-84.

[14] Hertrich I, Mathiak K, Lutzenberger W, Ackermann H. Time course of early audiovisual interactions during speech and nonspeech central auditory processing: a magnetoencephalography study. J Cogn Neurosci 2009; 21: 259-74.

[15] Hari R. The neuromagnetic method in the study of the human auditory cortex. In: Grandori F, Hoke M, Romani G, Eds. Auditory evoked magnetic fields and potentials. Advances in audiology. Basel: Karger 1990; pp. 222-82.

[16] Calvert GA, Bullmore ET, Brammer MJ, *et al*. Activation of auditory cortexd during silent lipreading. Science 1997; 276: 593-96.

[17] Pekkola J, Ojanen V, Autti T, *et al*. Primary auditory cortex activation by visual speech: an fMRI study at 3T. Neuroreport 2005; 16: 125-28.

[18] Bernstein LE, Auer ETJ, Moore JK, Ponton CW, Don M, Singh M. Visual speech perception without primary auditory cortex activation. Neuroreport 2002; 13: 311-15.

[19] Musacchia G, Sams M, Nicol T, Kraus N. Seeing speech affects acoustic information processing in the human brainstem. Exp Brain Res 2006; 168: 1-10.

[20] Jaaskelainen IP, Ahveninen J, Belliveau JW, Raij T, Sams M. Short-term plasticity in auditory cognition. Trends Neurosci 2007; 30: 653-61.

[21] Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Sams M. Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. Hum Brain Mapp 2006; 27: 471-7.

[22] Falchier A, Clavagnier S, Barone P, Kennedy H. Anatomical evidence of multimodal integration in primate striatal cortex. J Neurosci 2002; 22: 5749-59.

[23] Foxe JJ, Schroeder CE. The case for feedforward multisensory convergence during early cortical processing. Neuroreport 2005; 16: 419-23.

[24] Smiley J, Falchier A. Multisensory connections of monkey auditory cerebral cortex. Hear Res 2009; 258(1-2): 37-46.

[25] Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 2000; 10: 649-57.

[26] Campbell R, MacSweeney M, Surguladze S, *et al*. 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). Brain Res Cogn Brain Res 2001; 12: 233-43.

[27] Olson IR, Gatenby JC, Gore JC. A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Brain Res Cogn Brain Res 2002; 14: 129-38.

[28] Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E. Neural processes underlying perceptual enhancement by visual speech gestures. Neuroreport 2003; 14: 2213-18.

[29] Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb Cortex 2003; 13: 1034-43.

[30] Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E. Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. J Cogn Neurosci 2004; 16: 805-16.

[31] Beauchamp MS, Arqall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci 2004; 7: 1190-92.

[32] Miller LM, D'Esposito M. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci 2005; 25: 5884-93.

[33] Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: motor cortical activation during speech perception. Neuroimage 2005; 25: 76-89.

[34] Skipper JI, van Wassenhove V, Nusbaum HC, Small SL. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. Cereb Cortex 2007; 17: 2387-99.

[35] Jaaskelainen IP, Koskentalo K, Balk MH, *et al*. Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing. Open Neuroimaging J 2008; 2: 14-9.

[36] Murase M, Saito DN, Kochiyama T, *et al*. Cross-modal integration during vowel identification in audiovisual speech: a functional magnetic resonance imaging study. Neurosci Lett 2008; 434: 71-76.

[37] Hocking J, Price CJ. Dissociating verbal and nonverbal audiovisual object processing. Brain Lang 2009; 108: 89-96.

[38] Malacuso E, George N, Dolan R, Spence C, Driver J. Spatial and temporal factors during processing of audiovisual speech: a PET study. Neuroimage 2004; 21: 725-32.

[39] Reale RA, Calvert GA, Thesen T, *et al*. Auditory-visual processing represented in the human superior temporal gyrus. Neuroscience 2007; 145: 162-84.

[40] Hocking J, Price CJ. The role of the posterior superior temporal sulcus in audiovisual processing. Cereb Cortex 2008; 18: 2439-49.

[41] Kayser C, Petkov CI, Augath M, Logothetis NK. Functional imaging reveals visual modulation of specific fields in auditory cortex. J Neurosci 2007; 27: 1824-35.

[42] Wallace MT, Ramachandran R, Stein BE. A revised view of sensory cortical parcellation. Proc Natl Acad Sci USA 2004; 101: 2167-72.

[43] McKenna BM, Lin F-H, Raij T, Jaaskelainen IP, Stufflebeam S. Primary and multisensory cortical activity is correlated with audiovisual percepts. Hum Brain Mapp 2010; 31:526-38.

[44] Besle J, Fischer C, Bidet-Caulet A, Lecaignard F, Bertrand O, Giard MH. Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. J Neurosci 2008; 28: 14301-10.

[45] van Wassenhove V, Grant KW, Poeppel D. Temporal window of integration in auditory-visual speech perception. Neuropsychologia 2007; 45: 598-607.

[46] Davis C, Kislyuk DS, Kim J, Sams M. The effect of viewing speech on auditory speech processing is different in the left and right hemispheres. Brain Res 2008; 1242: 151-61.

[47] Sato M, Basirat A, Schwartz JL. Visual contribution to the multistable perception of speech. Percept Psychophys 2007; 69: 1360-72.

[48] Green KP, Norrix LW. Acoustic cues to place of articulation and the McGurk effect: the role of release bursts, aspiration, and formant transitions. J Speech Lang Hear Res 1997; 40: 646-65.

[49] Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J Neurosci 2005; 25: 5004-12.

[50] Chandrasekaran C, Ghazanfar AA. Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. J Neurophysiol 2009; 101: 773-88.

[51] Friedel P, van Hemmen JL. Inhibition, not excitation, is the key to multimodal sensory integration. Biol Cybern 2008; 98: 597-618.

[52] Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A. Neuronal oscillations and visual amplification of speech. Trends Cogn Sci 2008; 12: 106-13.

[53] Binder JR, Frost JA, Hammeke TA, *et al*. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 2000; 10: 512-28.

[54] Obleser J, Boecker H, Drzezga A, *et al*. Vowel sound extraction in anterior superior temporal cortex. Hum Brain Mapp 2006; 27: 562-71.

[55] Leff AP, Iverson P, Schofield TM, *et al*. Vowel-specific mismatch responses in the anterior superior temporal gyrus: an fMRI study. Cortex 2009; 45: 517-26.

[56] Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. Science 2001; 292: 290-93.

[57] Alain C, Arnott SR, Hevenor S, Graham S, Grady CL. "What" and "where" in the human auditory system. Proc Natl Acad Sci USA 2001; 98: 12301-06.

[58] Ahveninen J, Jaaskelainen IP, Raij T, *et al*. Task-modulated "what" and "where" pathways in human auditory cortex. Proc Natl Acad Sci USA 2006; 103: 14608-13.

[59] Lomber SG, Malhotra S. Double dissociation of 'what' and 'where' processing in auditory cortex. Nat Neurosci 2008; 11: 609-16.

[60] Naatanen R, Lehtokoski A, Lennes M, *et al*. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 1997; 385: 432-34.

[61] Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. Nature 2000; 403: 309-12.

[62] Möttönen R, Galvert GA, Jaaskelainen IP, *et al*. Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. Neuroimage 2006; 30: 563-9.

[63] Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. Trends Neurosci 2003; 26: 100-7.

[64] Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 2004; 92: 67-99.
[65] Warren JE, Wise RJ, Warren JD. Sounds do-able: auditory-motor transformations and the posterior temporal plane. Trends Neurosci 2005; 28: 636-43.
[66] Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 2009; 12: 718-24.
[67] Diehl RL, Kluender KR. On the objects of speech perception. Ecol Psychol 1989; 1: 121-44.
[68] Diehl RL, Lotto AJ, Holt LL. Speech perception. Ann Rev Psychol 2004; 55: 149-79.
[69] Liberman AM, Whalen DH. On the relation of speech to language. Trends Cogn Sci 2000; 3: 254-64.
[70] Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychol Rev 1967; 74: 431-61.
[71] Kuhl PK, Meltzoff AN. The bimodal perception of speech in infancy. Science 1982; 218: 1138-41.
[72] Rosenblum LD, Schmuckler MA, Johnson JA. The McGurk effect in infants. Perception Psychophys 1997; 59: 347-57.
[73] Goren CC, Sarty M, Wu PY. Visual following and pattern discrimination of face-like stimuli by newborn infants. Pediatrics 1975; 56: 544-9.
[74] Meltzoff AN, Moore MK. Imitation of facial and manual gestures by human neonates. Science 1977; 198: 75-8.
[75] Hebb DO. The organization of behavior. New York: Wiley 1949.
[76] Bishop DV, Brown BB, Robson J. The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. J Speech Hear Res 1990; 33: 210-19.
[77] Nishitani N, Hari R. Viewing lip forms: cortical dynamics. Neuron 2002; 36: 1211-20.
[78] Calvert GA, Campbell R. Reading speech from still and moving faces: the neural substrates of visible speech. J Cogn Neurosci 2003; 15: 57-70.
[79] Ojanen V, Mottonen R, Pekkola J, *et al*. Processing of audiovisual speech in Broca's area. Neuroimage 2005; 25: 333-8.
[80] Bernstein LE, Auer ETJ, Wagner M, Ponton CW. Spatiotemporal dynamics of audiovisual speech processing. Neuroimage 2008; 39: 423-35.
[81] Okada K, Hickok G. Two cortical mechanisms support the integration of visual and auditory speech: a hypothesis and preliminary data. Neurosci Lett 2009; 452: 219-23.
[82] Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. The essential role of premotor cortex in speech perception. Curr Biol 2007; 17: 1692-6.
[83] Möttönen R, Watkins KE. Motor representations of articulators contribute to categorical perception of speech sounds. J Neurosci 2009; 29: 9819-25.
[84] Liu AK, Dale AM, Belliveau JW. Monte Carlo simulation studies of EEG and MEG localization accuracy. Hum Brain Mapp 2002; 16: 47-62.
[85] Bar M, Kassam KS, Ghuman AS, *et al*. Top-down facilitation of visual recognition. Proc Natl Acad Sci USA 2006; 103: 449-54.
[86] Fingelkurts AA, Fingelkurts AA, Krause CM, Möttönen R, Sams M. Cortical operational synchrony during audio-visual speech integration. Brain Lang 2003; 85: 297-312.
[87] Fingelkurts AA, Fingelkurts AA, Krause CM. Composition of brain oscillations and their functions in the maintenance of auditory, visual and audio-visual speech percepts: an exploratory study. Cogn Processes 2007; 8: 183-99.

Received: August 24, 2009 Revised: September 28, 2009 Accepted: September 30, 2009