

## Enhanced early-latency electromagnetic activity in the left premotor cortex is associated with successful phonetic categorization

Jussi Alho<sup>a,\*</sup>, Marc Sato<sup>b</sup>, Mikko Sams<sup>a</sup>, Jean-Luc Schwartz<sup>b</sup>, Hannu Tiitinen<sup>a</sup>, Iiro P. Jääskeläinen<sup>a,c</sup>

<sup>a</sup> Brain and Mind Laboratory, Department of Biomedical Engineering and Computational Science, Aalto University, School of Science, P.O. Box 12200, FIN-00076 AALTO, Finland

<sup>b</sup> Gipsa-Lab, Département Parole & Cognition, CNRS & Grenoble Université, BP 25, 38040 Grenoble cedex 9, France

<sup>c</sup> Advanced Magnetic Imaging Centre, Aalto University School of Science, P.O. Box 13000, FIN-00076 AALTO, Finland

### ARTICLE INFO

#### Article history:

Received 8 November 2011

Revised 12 January 2012

Accepted 4 February 2012

Available online 14 February 2012

#### Keywords:

Magnetoencephalography

Speech perception

Dorsal stream

Premotor cortex

Auditory

Motor

### ABSTRACT

Sensory-motor interactions between auditory and articulatory representations in the dorsal auditory processing stream are suggested to contribute to speech perception, especially when bottom-up information alone is insufficient for purely auditory perceptual mechanisms to succeed. Here, we hypothesized that the dorsal stream responds more vigorously to auditory syllables when one is engaged in a phonetic identification/repetition task subsequent to perception compared to passive listening, and that this effect is further augmented when the syllables are embedded in noise. To this end, we recorded magnetoencephalography while twenty subjects listened to speech syllables, with and without noise masking, in four conditions: passive perception; overt repetition; covert repetition; and overt imitation. Compared to passive listening, left-hemispheric N100m equivalent current dipole responses were amplified and shifted posteriorly when perception was followed by covert repetition task. Cortically constrained minimum-norm estimates showed amplified left supramarginal and angular gyri responses in the covert repetition condition at ~100 ms from stimulus onset. Longer-latency responses at ~200 ms were amplified in the covert repetition condition in the left angular gyrus and in all three active conditions in the left premotor cortex, with further enhancements when the syllables were embedded in noise. Phonetic categorization accuracy and magnitude of voice pitch change between overt repetition and imitation conditions correlated with left premotor cortex responses at ~100 and ~200 ms, respectively. Together, these results suggest that the dorsal stream involvement in speech perception is dependent on perceptual task demands and that phonetic categorization performance is influenced by the left premotor cortex.

© 2012 Elsevier Inc. All rights reserved.

### Introduction

Understanding how listeners process acoustic speech signals to recover phonetic information, considering the high variability of the speech signal, is one of the long-standing problems in the neurosciences. It is well known that there is no one-to-one correspondence between perceived phonemes and their acoustic manifestations due to coarticulation (e.g., the phoneme /d/ is acoustically quite different in syllables /di/ and /du/). Given this lack of invariance problem, several alternative theoretical accounts have been proposed to explain efficiency of speech perception, with the main difference between them being whether purely auditory perceptual mechanisms vs. the speech motor system and/or articulatory control processes are implicated to be involved in decoding speech (for reviews, see Diehl et al., 2004; Galantucci et al., 2006; Schwartz et al., 2010, 2008).

Auditory processing models of speech perception are based on the assumption that acoustics of speech itself contain sufficient amount

of invariant properties to recover phonetic information and that speech perception relies on general mechanisms of audition and perceptual learning (for a review, see Diehl et al., 2004). In contrast, the motor theory of speech perception (Liberman et al., 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000) argues that rather than the acoustic properties of speech, the objects of speech perception are the speaker's intended phonetic gestures that are processed by a specialized speech decoder. In these models, speech perception and speech production therefore share a close relationship; that is, the role of the speech motor system is not only to produce speech articulations but also to recover the intended articulatory gestures with the help of the acoustic (and visual; Jaaskelainen, 2010) cues. It has been further suggested that the speech decoder (or vocal-tract synthesizer) uses an analysis-by-synthesis mechanism, based on the constructivist approach of perception (Helmholtz, 1867), to determine the "strongest candidate hypothesis" of the speaker's intended gesture (Liberman and Mattingly, 1985).

Consistent with the motor theory of speech perception (and contradicting the purely auditory approaches), magnetoencephalography (MEG), transcranial magnetic stimulation (TMS), and functional magnetic resonance imaging (fMRI) studies have demonstrated

\* Corresponding author. Fax: +358 947024833.

E-mail address: [jussi.alho@aalto.fi](mailto:jussi.alho@aalto.fi) (J. Alho).

involvement of the speech motor system, that is, the posterior part of the left inferior frontal gyrus (pIFG) and the left ventral premotor cortex (vPMC), during auditory, visual, and audio-visual speech perception (e.g., Nishitani and Hari, 2002; Ojanen et al., 2005; Pekkola et al., 2005; Pulvermuller et al., 2006; Skipper et al., 2005, 2007; Watkins and Paus, 2004; Wilson and Iacoboni, 2006; Wilson et al., 2004). Furthermore, ambiguity of the sensory inputs has been demonstrated to enhance the activation of the speech motor areas, as shown by fMRI studies involving the perception of intact compared to masked (e.g., Binder et al., 2004) and degraded (e.g., Zekveld et al., 2006) speech sounds.

Repeated TMS (rTMS) and intracortical stimulation studies have shown that temporarily disrupting the left pIFG (i.e., Broca's area) can result in impaired ability to discriminate between phonemes (Boatman, 2004; Romero et al., 2006). rTMS stimulation of the left vPMC also resulted both in decreased ability to identify auditory syllables embedded in white noise (Meister et al., 2007) and in slower response times in a phoneme discrimination task (Sato et al., 2009). Further, stimulation of the motor lip or tongue areas specifically impaired the ability to discriminate speech sounds that involve the corresponding speech effector when the sounds were artificially produced phonemes along a continuum (Mottonen and Watkins, 2009) and syllables immersed in white noise (D'Ausilio et al., 2011). In a related study, fMRI and MEG were used to measure responses for correct and incorrect trials of an auditory phonetic identification task, in which stimuli consisted of speech sounds masked with continuous white noise (Callan et al., 2010). The results indicated significantly differential activity in the vPMC and pIFG for correct over incorrect perceptual trials, supporting the hypothesis that articulatory motor processes facilitate perceptual performance.

In order to explain the discrepancies between the acoustic processing and motor theories of speech perception, large-scale brain network models of speech processing have been recently advanced (e.g., Hickok and Poeppel, 2007; Rauschecker and Scott, 2009; Skipper et al., 2007). The models are based on two functionally distinct auditory cortical processing streams: the ventral (“what”) stream and the dorsal (“where/how”) stream (Ahveninen et al., 2006; Arnott et al., 2004; Rauschecker and Tian, 2000). While the ventral stream is presumably involved in acoustic-phonetic decoding, the dorsal stream has been hypothesized to underlie sensorimotor mapping between auditory representations and articulatory motor representations (Hickok and Poeppel, 2000, 2007; Scott, 2005; Scott and Johnsrude, 2003). Sharing the same fundamental idea of the constructivist theories of perception (Helmholtz, 1867) as the motor theory of speech perception, these models postulate that, by the means of successive sensory-to-motor and motor-to-sensory projections, the analysis of incoming speech inputs is mediated by a simulation signal from the speech motor system (i.e., efference copy) that predicts the sensory consequences of the activated articulatory motor commands, thus constraining the phonetic interpretation of the sensory inputs.

For example, Rauschecker and Scott (2009) (elaborated by Rauschecker, 2011) propose that, starting out from the Heschl's gyrus, the model performs forward mapping by hierarchically decoding acoustic speech information in the ventral stream from the primary auditory cortex to the anterior temporal cortex and from there to the inferior regions of frontal cortex. Information is then further mapped in premotor cortex (PMC) onto articulatory representation and finally transmitted to the IPL and pSTG/STS as an efference copy. In the reverse direction, inverse mapping takes place whereby an early “sketch” of the incoming sensory information, coming from the pSTG/pSTS to the IPL, is compared with the predictive efference copy from the vPMC. Furthermore, changes in the IPL related to attention or intention (Andersen and Cui, 2009; Colby and Goldberg, 1999) can have a modulatory effect on the activated motor plans in the PMC,

thus influencing the selection of the motor commands and consequently the content of the efference copy in this model.

Importantly, although some of these models do not consider the dorsal stream to have a critical role in speech perception under normal listening conditions in adults (Hickok and Poeppel, 2007), they assume that the motor system is critically recruited under adverse conditions in order to resolve phonetic ambiguity. This proposal is indirectly supported by fMRI studies showing stronger activity of the motor system depending on the intelligibility of speech input: during masked or distorted versus intelligible speech (e.g., Binder et al., 2004; Zekveld et al., 2006) or during the auditory identification of non-native versus native phonemes (e.g., Callan et al., 2004; Wilson and Iacoboni, 2006). Previous electroencephalography (EEG) and MEG studies have also used intact and distorted speech sounds to investigate the cortical dynamics underlying speech perception (Davis et al., 2008; Liikkanen et al., 2007; Miettinen et al., 2011, 2010; Shtyrov et al., 1998, 1999).

It is possible to conceptualize speech perception as an interactive neural process that involves reciprocal connections between sensory and motor areas with the connection strengths varying as a function of the perceptual task and external environment (Sato et al., 2009). This is a view supported by the “Perception-for-Action-Control Theory” (PACT; Schwartz et al., 2010), which considers speech units to be neither purely auditory nor purely motor, but rather that they are formed as a result of sensorimotor interaction processes in which the units (or percepts) are shaped both by sensory processing and motor knowledge of speech gestures. In complementing sensory inputs with adequate articulatory information, the role of motor representations is to predict future sensory events and integrate them together for improved perception. Thus, PACT provides a tentative link between sensory and motor representations within the framework of speech perception. In support of the PACT model, it has been argued that given infants' characteristic tendency to pay attention and imitate speech sounds and facial gestures, the neuronal receptive fields in the posterior temporal lobe areas (receiving connections from the auditory, visual, and speech motor areas) are shaped during development to respond to acoustic features that coincide with visual and motor speech cues (Jaaskelainen, 2010).

Here, we investigated the cortical dynamics underlying the perception of speech sounds when the participants are to perform speech motor tasks. Given that the neural processes underlying speech perception occur on a millisecond timescale, an ideal measurement technique for the purpose of this study is provided by MEG, which allows for the non-invasive examination of cortically-induced magnetic fields with a millisecond time resolution and very good spatial localization accuracy for cortical sources (for a review, see Hamalainen et al., 1993). MEG epochs time-locked to auditory stimulation consist of a number of distinct responses, from which one of the most prominent is the so-called N100m that peaks at about 100 ms after sound onset. The N100m has been shown to be sensitive to acoustic attributes of speech signal as well as top-down influences such as efference copies of intended motor actions, therefore constituting an useful index for studying the cortical mechanisms of speech processing (Curio et al., 2000; Kauramaki et al., 2010; Kudo et al., 2004; Makela et al., 2005; Obleser et al., 2003; Poeppel et al., 1997).

The functional status of the longer-latency P200m component, peaking at about 200 ms after sound onset, is less well understood. In the past, it has often been referred to in the context of the N100–P200 complex. Although apparently sharing some processing properties with the N100m (both seem to respond to spectral and temporal cues contained in spoken language; e.g., Whiting et al., 1998), there is increasing evidence suggesting dissociation of the N100 and P200. The cortical source of the auditory P200m component has been shown to be located more anteriorly compared to the N100m (e.g.,

Ross and Tremblay, 2009; Sams et al., 1985). A recent study combining fMRI and MEG suggested that analysis of speech stimulus types and sound location takes place in anterior and posterior auditory areas, respectively (Ahveninen et al., 2006). Given that the P200m source location has been estimated in areas anterior to the primary auditory cortex, these findings yield support to the hypothesis that the P200m reflects the acoustic features of stimulation (Naatanen and Picton, 1987). The auditory P200 component has also been shown to be enhanced in amplitude during selective attention tasks (Coch et al., 2005) and has been suggested to be related to stimulus identification/discrimination (Crowley and Colrain, 2004; Novak et al., 1992), possibly reflecting a comparison between the incoming auditory stimulus and an internal memory representation.

To test the dependency of the sensorimotor integration processes on the perceptual task during speech perception, we examined how the MEG responses to perception of an auditory syllable are influenced by a subsequent overt/covert repetition and overt imitation of the syllable compared to passive listening. Each of the conditions involved the same set of syllables, and according to the assumption that the recruitment of speech motor areas increases with the ambiguity of the incoming sensory information, the syllables were presented with and without masking noise, with the hypothesis that recruitment of the dorsal stream increases during perception of syllables embedded in noise. We also hypothesized that engagement in a speech motor task subsequent to perception enhances dorsal stream electromagnetic activity. More specifically, we expected the level of sensorimotor interactions to increase as a function of the perceptual task difficulty, from passive listening to repetition to imitation, causing increased activity in the key component regions within the dorsal processing stream (especially in the left PMC).

This hypothesis is derived from a neural model of speech acquisition and production, called the DIVA model (Directions Into Velocities of Articulators), according to which additional involvement of the speech sound map (hypothesized to reside in left vPMC and pIFG) is required for “unusual” production, such as imitation, that necessitates more accurate tracking of phonetic details (e.g., accommodation of F0; Guenther et al., 2006). In sum, for perception in the passive condition, a simple category must be provided; in the repetition condition, translation of the category into a template for production is required; and in the imitation condition, all the necessary fine details for more accurate reproduction need to be produced. The covert repetition is hypothesized to have a similar level of speech sound map involvement as overt repetition; however, since hearing one’s own overt vocalization in between the stimulus presentations potentially has a confounding effect on the responses time-locked to the stimulus presentation, the covert condition serves as a control for the passive condition.

We further envisioned that activation of the speech motor system would have a positive correlation with perceptual performance (i.e., syllable identification accuracy correlating with enhanced motor system responses to speech stimuli when the subjects are engaged in a phonetic identification task). Results in line with these hypotheses would (1) argue against the view that speech perception is determined solely based on auditory perceptual mechanisms without any role of the speech motor system; (2) support the constructivist internal model approaches, such as PACT, which suggest that perceptual task difficulty and sensory input ambiguity are reflected as increased activity in the dorsal stream regions; and (3) support the conception that articulatory motor processes facilitate perceptual performance. We specifically set forth to utilize the temporal resolution offered by MEG and spatial localization accuracy provided by MRI-based cortically-constrained minimum norm estimates (MNEs) of electromagnetic source activity (Lin et al., 2006) to probe at which latencies the possible dorsal stream enhancements and correlations with behavioral task performance take place.

## Materials and methods

### Subjects

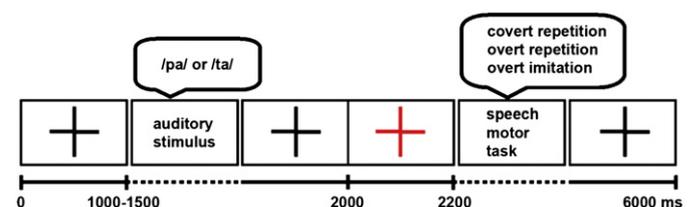
Twenty healthy individuals with informed consent participated in the study (7 females, 2 left-handed, range 21–58 years, mean  $\pm$  SD  $27.4 \pm 8.0$ ). All subjects reported normal hearing ability. 19 of the subjects were native speakers of Finnish, and one was a native speaker of Italian. The experiment was approved by the Ethical Committee of the Helsinki University Central Hospital.

### Stimuli and task

The stimuli that were used in the experiment consisted of intact and noise-embedded /pa/ and /ta/ syllable sounds, articulated by a native male Finnish speaker. Five individual clearly articulated /pa/ and /ta/ tokens were selected, cut 100 ms preceding the detected consonantal burst and 100 ms following it (at zero crossing points), and finally scaled at a fixed intensity. A silence period was added to the end of each token to obtain a total duration of 300 ms. For the noise-embedded stimuli, Gaussian pink-noise masks were generated. The masks had a 5-ms rise decay envelope, were de-emphasized to better match the frequency spectrum of /pa/ and /ta/ syllables (at  $-6$  dB/oct), and were simultaneously presented from the beginning of the consonantal burst to the beginning of the silence period of the stimuli with signal-to-noise ratios (SNRs) of  $+5$  dB,  $0$  dB,  $-5$  dB, and  $-7$  dB.

To define an appropriate noise-embedded set of stimuli for the experiment, hit rates were determined in a separate behavioral experiment ( $N=6$ ) with a forced-choice identification between syllables /pa/ and /ta/ at different signal-to-noise levels. The tests comprised of 50 randomized trials (25 /pa/ and 25 /ta/ syllables) presented via headphones with two sound levels (55 and 65 dB SPL).  $+5$  dB SNR set at 65 dB SPL was chosen as the subjects identified the stimuli at an appropriate level (i.e., 77% correct responses).

In the MEG experiment, the stimuli were presented to the participants in four different conditions: passive perception; passive perception and subsequent overt repetition; passive perception and subsequent covert repetition; and passive perception and subsequent overt imitation. In each condition, 300 trials consisting of (1) a randomly varying 1000–1500-ms prestimulus baseline for perception, (2) pseudorandom auditory stimulus presentation, (3) a prestimulus baseline for production (up to 2000 ms from the beginning of the trial), and (4) a visual signal to articulate (fixation cross turning from black to red and back to black; 2000–2200 ms) were presented (see Fig. 1). The total duration of the trial was 6 s, with interstimulus interval (ISI) varying between 5.5 and 6.5 s, and the interval between the onset of the auditory stimulus and the subsequent visual signal to articulate varying between 500 and 1000 ms. The auditory stimuli were presented via a panel speaker located 280 cm in front of the subject with an approximate 65-dB sound level. The fixation cross was presented on a screen situated 170 cm in front of the subject (visual angle  $0.3^\circ \times 0.3^\circ$ ). The auditory stimuli as well as the visual



**Fig. 1.** Experimental procedure. The stimuli were pseudorandomly presented with a random 1000 to 1500 ms prestimulus baseline, followed by a prestimulus baseline for production (up to 2000 ms from the beginning of the trial), and a go-signal for the motor task (black fixation cross turning briefly to red between 2000 and 2200 ms). The interstimulus interval therefore varied between 5500 and 6500 ms.

signaling were delivered using Presentation software (v10.1, Neurobehavioral systems).

In the three active conditions (covert repetition, overt repetition, and overt imitation), the task of the subjects was to perceive the phoneme first, then wait for the visual signal, and perform the speech production task. In the overt repetition condition, the subjects were to identify between /pa/ and /ta/ and overtly repeat the syllable. The task in the covert repetition condition was otherwise similar, except to take place covertly without any articulatory movements or sound production. In the imitation condition, the subjects were both asked to imitate the voice of stimulus and repeat exactly what they heard. The covert repetition condition, being a silent motor condition, was included in the experiment as a control given that hearing one's own overt vocalization in between the stimulus presentations is likely to have a confounding effect on the responses time-locked to the stimulus presentation compared to the silent passive condition. Thus, for reliability, the covert repetition condition constituted a standard for comparison with the passive condition and, similarly, the imitation for the overt repetition condition.

#### Data recording

The MEG data were acquired with a whole-head 306-channel neuromagnetometer (VectorView, Elekta-Neuromag, Finland) with 102 sensor elements in a helmet array (two orthogonal planar gradiometers and one magnetometer in each element). The device was situated in a magnetically shielded room, covered with three layers of  $\mu$ -metal and aluminum to attenuate any effects of outside magnetic fields, with additional active noise-cancellation system implemented. In addition to the MEG channels, an electro-oculogram (EOG) channel with electrodes attached below and on the outer canthus of the left eye was recorded to detect eye blinks and movements. The MEG signals were band-pass filtered at 0.03–200 Hz and digitized at a sampling frequency of 2000 Hz.

Before each MEG recording session, the locations of four head-position indicator coils attached to the scalp were recorded with respect to three anatomical landmark points (nasion and two preauricular points) using a 3-D digitizer (Isotrak, Polhemus, Colchester, VT). Also, around 30 additional scalp surface points were digitized to facilitate both the determination of the individual spherical head model used in ECD source modeling as well as the MRI-MEG coregistration used in the distributed source modeling (minimum-norm estimates; MNEs).

Responses time-locked to the onset of the auditory syllable presentation were offline-averaged separately for each eight trial types (four conditions with two stimulus types), with epochs exceeding 3000 fT/cm amplitude in the MEG channels or 150  $\mu$ V in the EOG channel rejected from the average. The epoch time limits were set to 200 ms preceding the stimulus onset and 500 ms following it. The averaged signals were low-pass filtered at 40 Hz and all amplitudes measured with respect to a 200-ms prestimulus baseline. No analyses time-locked to the go-signal for production were performed.

For subsequent identification of the participants' repetitions, microphone recordings with 22.05 kHz sampling rate together with electromyographic (EMG) channels with electrodes placed on three specific articulators (sternohyoid, orbicularis oris superior, and masseter) were recorded. The EMG was further used to control for the presence of any covert articulations that might have occurred after the perception of the syllables (i.e., before the onset of the signaled motor task). For verifying that the participants performed the imitation task as required, the pitch (F0) across the response tokens was measured from the microphone recordings using an acoustic periodicity detection algorithm on the basis of an accurate autocorrelation method implemented in Praat software (Boersma, 1993, 2001).

#### Data analysis

The sources of the measured magnetic fields were first modeled by fitting a single equivalent current dipole (ECD) at the individually determined peak latencies of the N100m response (recorded about 100 ms from the auditory speech stimulus onset) using a spherical head model and a fixed subset of 34 planar gradiometers over the left and right hemisphere temporal areas (Hamalainen et al., 1993). The N100m magnetic peak amplitudes and latencies for each subject and trial type (condition  $\times$  stimulus type) were determined from the individual source waveforms by defining the maximum dipole moment in the interval 80–140 ms. The data from four subjects were excluded from the ECD analysis due to weak signal-to-noise ratio; the average goodness-of-fit values of the ECDs accepted for analysis was 88.7%. Grand average source waveforms were obtained by averaging the individual source waveforms. Differences between the peak latencies, peak amplitudes, and locations of the ECDs were statistically tested using a three-way repeated-measures analysis of variance (ANOVA) with hemisphere, condition, and stimulus type as within-subjects factors.

Additionally, the source current distribution was estimated by computing MNEs (Hamalainen and Ilmoniemi, 1994) from MRI-constrained MEG data. For this purpose, the cortical surfaces of each subject, reconstructed with Freesurfer software (<http://surfer.nmr.mgh.harvard.edu/>), were decimated to 5000–10000 source locations (or vertices) per hemisphere with a 5-mm spacing between the closest vertices. The forward solution (i.e., modeling of MEG sensor measurements), with one current dipole at each of the source locations, was calculated using a single compartment boundary-element model (Hamalainen and Sarvas, 1989), with the inner skull surface derived from the structural MRI data. Activity at each source location (i.e., inverse solution) was estimated for each time step of the averaged evoked responses using the anatomically constrained linear estimation approach (Dale et al., 2000), with the help of an inverse operator calculated using the forward solution and a noise covariance matrix estimated from the 200-ms prestimulus baselines of the individual trials. A loose orientation constraint together with depth weighting was applied in the calculation of the inverse operator (Lin et al., 2006). Since individual MRI-images were not available for six subjects, an MRI template with 1-mm resolution was used as a substitute in these subjects (Holmes et al., 1998). The dipole and distributed source modeling were implemented with Elekta Neuromag software (v5.5.18, Elekta Neuromag Oy, Helsinki, Finland) and MNE software (<http://www.nmr.mgh.harvard.edu/martinos/userInfo/data/sofMNE.php>), respectively. All statistical analyses were done with Statistica software (v9, StatSoft Inc., Tulsa, USA).

MNE time courses were extracted from six *a priori* anatomical regions per hemisphere based on a FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>) parcellation (Fischl, 2004). The selected regions were Heschl's gyrus (HG), planum temporale (PT), pars opercularis (POP), premotor cortex (PMC), supramarginal gyrus (SMG), and angular gyrus (AG). Dynamic statistical parametric maps (Dale et al., 2000) were calculated to reveal the cortical regions where the MNE activity exceeded the noise level. This information was used in the identification/selection of the specific regions-of-interest (ROIs) within the *a priori* selected anatomical areas. The identified ROIs were then drawn on the reconstructed inflated cortical surface of each subject. Time courses were computed for each subject, trial type, and ROI by averaging across the source location waveforms falling inside a given region. For statistical analysis, peak source amplitudes at  $\sim$ 100 and  $\sim$ 200 ms were determined from unsigned time courses using the grand average peak latencies of the respective components identified from signed time courses. The unsigned time courses were used in the statistical analysis because of their higher signal-to-noise ratio. The peak amplitudes were entered into a two-way repeated-measures ANOVA with condition and stimulus type as

within-subjects factors. In addition to testing the interactions and main effects of condition and stimulus type, separate *a priori* comparisons were conducted between the passive and covert repetition conditions, and between the overt repetition and imitation conditions. For assessing the possible difference in responses to /pa/ and /ta/, an additional three-way repeated-measures ANOVA with condition, stimulus type, and syllable as within-subjects factors was conducted. Grand average time courses were calculated by averaging the individual MNE time courses. Spearman rank correlation tests were further used to evaluate correlations between the amplitude of each source at ~100 and ~200 ms latencies based on MNE analyses and accuracy at which the noise-embedded syllables were identified in the overt repetition condition, and between the source amplitudes in the imitation condition and “unusualness” of production (as defined by the individual differences in the F0 measured across the response tokens between the overt repetition and imitation conditions).

**Results**

*ECD source modeling*

*N100m ECD amplitude and latency*

Fig. 2 shows the N100m ECD amplitudes and latencies for each trial type. For across-subjects averaged sensor-space ERF responses to the intact and noise-embedded stimuli, see Supplementary Figs. S1 and S2, respectively. A statistically significant main effect of condition was observed for the N100m amplitude and latency in the left hemisphere [amplitude:  $F(3,45) = 5.41, p < 0.01$ ; latency:  $F(3,45) = 4.82, p < 0.01$ ]. N100m response amplitudes were significantly increased in the covert repetition compared to the passive condition in the left hemisphere when the stimuli were embedded in noise [ $F(1,15) = 8.41, p < 0.05$ ]. N100m latencies were shorter in the right hemisphere than in the left [ $F(1,15) = 7.34, p < 0.05$ ]. The noise-embedded stimulus type elicited earlier N100m responses in the right hemisphere [ $F(1,15) = 7.20, p < 0.05$ ] and later responses in the imitation compared to the overt repetition condition in the left hemisphere [ $F(1,15) = 6.92, p < 0.05$ ]. Furthermore, a statistically significant stimulus type  $\times$  condition interaction was observed in the left-hemispheric latencies [ $F(3,45) = 3.41, p < 0.05$ ]. This effect

seemed to be caused by significantly earlier responses for the noise-embedded compared to the intact stimulus type specifically in the passive condition [ $F(1,15) = 10.83, p < 0.01$ ] and in the covert repetition condition [ $F(1,15) = 7.22, p < 0.05$ ].

*N100m ECD source location*

Fig. 3 shows the N100m ECD locations on the xy-plane of the head coordinate system (anterior–posterior/medial–lateral dimension). A comparison between the passive and covert repetition conditions showed a statistically significant difference in the location of the N100m ECD response for the noise-embedded stimulus type [ $F(1,15) = 9.88, p < 0.01$ ], with the approximated center of gravity of neural activity being ~3 mm more posterior in the covert repetition condition.

*MNE source modeling*

*Amplification of source amplitudes in the active conditions*

Cortically constrained MNE source amplitude estimates showed significantly enhanced responses at ~100 ms after stimulus onset in the covert repetition compared to the passive condition in the left hemisphere SMG for the intact stimulus type [ $F(1,19) = 4.77, p < 0.05$ ] and in the left hemisphere AG for both stimulus types [intact:  $F(1,19) = 5.11, p < 0.05$ ; noise-embedded:  $F(1,19) = 4.50, p < 0.05$ ; Fig. 4]. Amplification in the covert repetition compared to the passive condition was also observed for the peak source amplitudes at ~200 ms both in the left hemisphere AG for the noise-embedded stimulus type [ $F(1,19) = 9.07, p < 0.01$ ] and in the left hemisphere PMC for both stimulus types [intact:  $F(1,19) = 5.84, p < 0.05$ ; noise-embedded:  $F(1,19) = 16.9, p < 0.001$ ; Fig. 5]. Furthermore, the source amplitudes at ~200 ms in the left hemisphere PMC exhibited a significant main effect of condition [ $F(3,57) = 9.72, p < 0.001$ ], with amplified responses with respect to the passive condition observed also in the overt repetition condition [intact:  $F(1,19) = 6.70, p < 0.05$ ; noise-embedded:  $F(1,19) = 13.2, p < 0.01$ ] and imitation condition [intact:  $F(1,19) = 15.9, p < 0.001$ ; noise-embedded:  $F(1,19) = 34.9, p < 0.001$ ], yet not between the overt repetition and imitation conditions.

*Effects of stimulus type*

The intact stimulus type elicited stronger responses than the noise-embedded one at ~100 ms after stimulus onset in the left hemisphere HG [ $F(1,19) = 37.2, p < 0.001$ ], POP [ $F(1,19) = 11.8, p < 0.01$ ], PT [ $F(1,19) = 5.11, p < 0.05$ ], and SMG [ $F(1,19) = 11.0, p < 0.01$ ]. An inverse effect was observed in the left hemisphere AG [ $F(1,19) = 5.28, p < 0.05$ ] and in the right hemisphere SMG [ $F(1,19) = 11.8, p < 0.01$ ]. Related to the main effect of stimulus type, a significant condition  $\times$  stimulus type interaction was observed for the peak source amplitude at ~100 ms in the left hemisphere PT [ $F(3,57) = 8.20, p < 0.001$ ], caused by stimulus type dependency present only in the covert repetition [ $F(1,19) = 17.7, p < 0.001$ ] and imitation [ $F(1,19) = 14.1, p < 0.001$ ] conditions. Furthermore, significant interactions for hemisphere  $\times$  condition  $\times$  stimulus type [ $F(3,57) = 3.81, p < 0.05$ ] and hemisphere  $\times$  stimulus type [ $F(1,19) = 32.3, p < 0.001$ ] were observed at ~100 ms in the SMG, caused by significantly stronger responses in the left compared to the right SMG elicited only by the intact stimulus type [ $F(1,19) = 8.44, p < 0.01$ ] and present only in the active conditions. Significant interactions for hemisphere  $\times$  stimulus type at ~100 ms after stimulus onset were observed also in the HG [ $F(1,19) = 23.2, p < 0.001$ ] and PMC [ $F(1,19) = 14.9, p < 0.01$ ], with the effect in the HG caused by stronger responses in the left compared to the right hemisphere elicited only by the intact stimulus type [ $F(1,19) = 5.28, p < 0.05$ ] and in the PMC, conversely, caused by stronger responses in the right compared to the left hemisphere elicited only by the noise-embedded stimulus type [ $F(1,19) = 7.85, p < 0.05$ ].

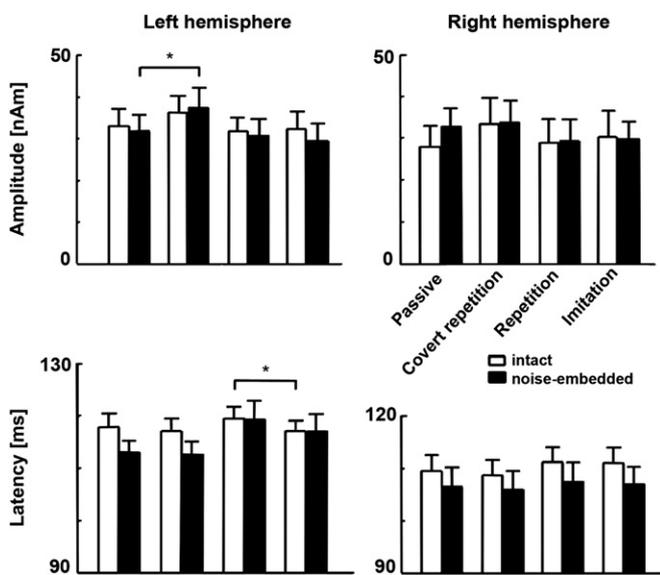
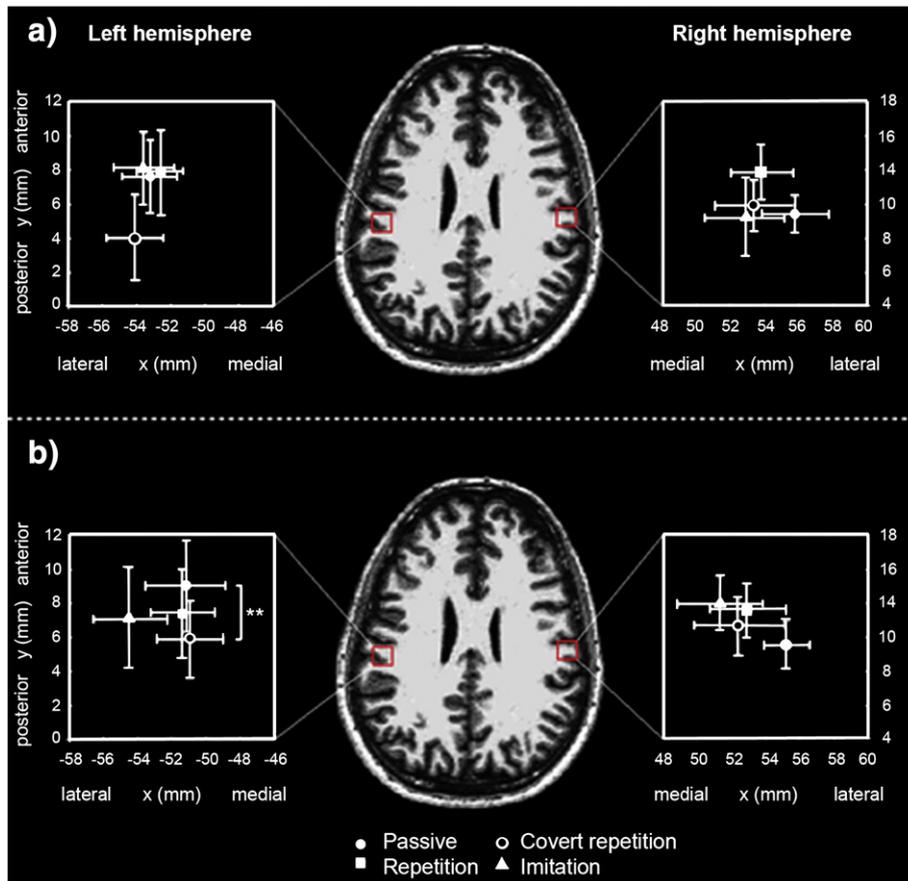
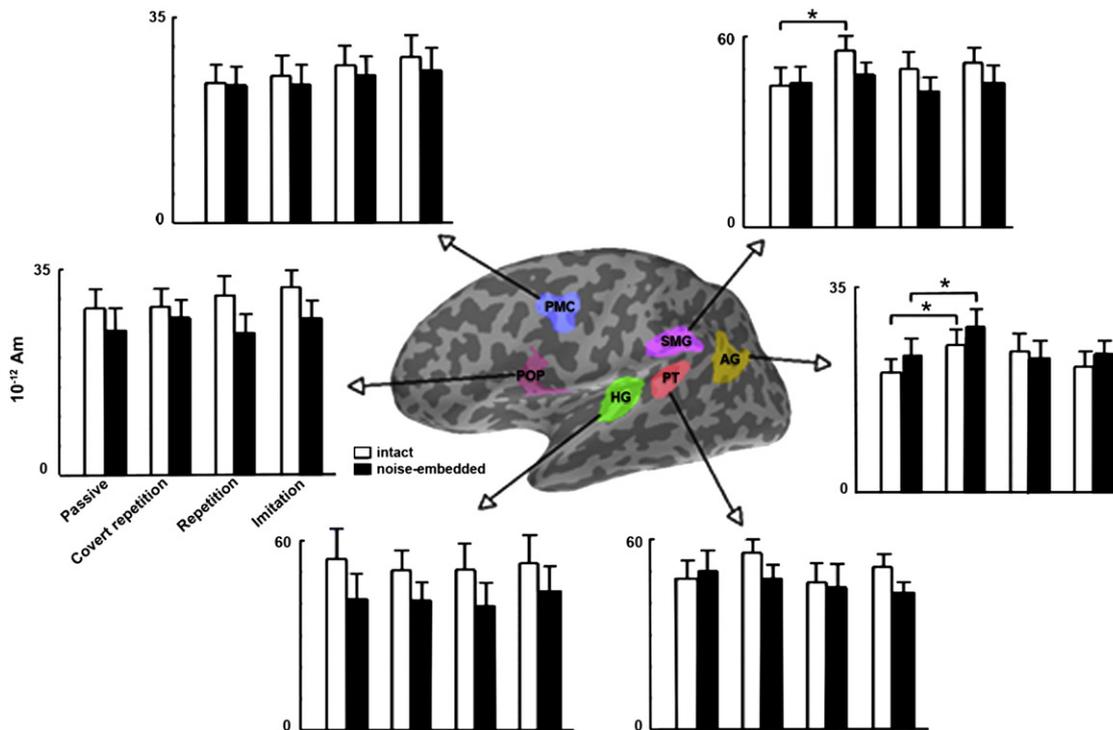


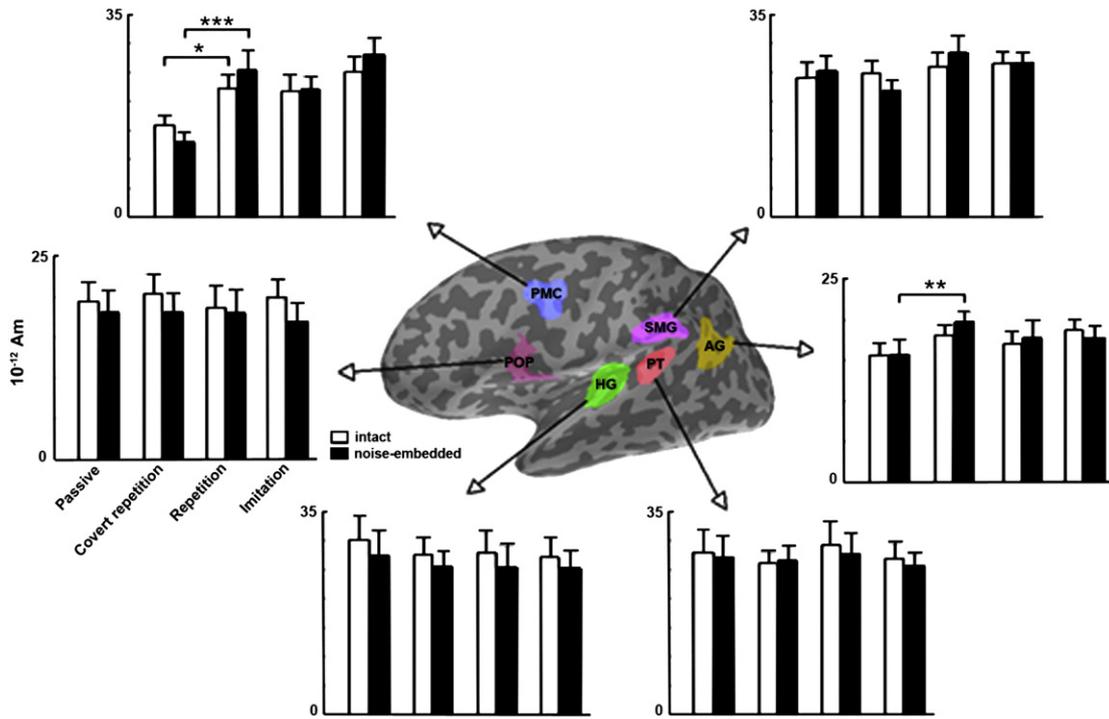
Fig. 2. N100m peak amplitudes and latencies. Mean N100m ECD amplitudes (top half) and latencies (bottom half) for each trial type estimated using a subset of 34 planar gradiometers over the left and right hemisphere temporal areas. Error bars indicate SEM. Asterisks indicate significant differences (\* $p < 0.05$ ).



**Fig. 3.** N100m ECD locations. Mean N100m ECD locations for the (a) intact and (b) noise-embedded stimulus types in the anterior–posterior and medial–lateral dimensions for all conditions estimated using a subset of 34 planar gradiometers over the left and right hemisphere temporal areas. Error bars indicate SEM. Asterisks indicate significant differences (\*\* $p < 0.01$ ).



**Fig. 4.** MNE source amplitudes at -100 ms after stimulus onset. MNE peak source amplitudes at -100 ms after stimulus onset for all trial types extracted from the predefined cortical locations in the left hemisphere. Error bars indicate SEM. Asterisks indicate significant differences (\* $p < 0.05$ ).



**Fig. 5.** MNE source amplitudes at ~200 ms after stimulus onset. MNE peak source amplitudes at ~200 ms after stimulus onset for all trial types extracted from the predefined cortical locations in the left hemisphere. Error bars indicate SEM. Asterisks indicate significant differences (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).

In the AG, stronger responses were elicited both by the intact and the noise-embedded stimulus type in the right compared to the left AG at ~200 ms [intact:  $F(1,19) = 9.81$ ,  $p < 0.01$ ; noise-embedded:  $F(1,19) = 7.43$ ,  $p < 0.05$ ]. In addition, main effect of hemisphere was observed for the responses at ~100 ms in the PMC and at ~200 ms in the AG.

*Effects of syllable type*

A significant main effect of syllable was observed for the responses at ~100 ms after stimulus onset in the left hemisphere HG [ $F(1,19) = 15.1$ ,  $p < 0.001$ ] and POP [ $F(1,19) = 9.64$ ,  $p < 0.01$ ], and at ~200 ms in the left hemisphere HG [ $F(1,19) = 5.30$ ,  $p < 0.05$ ], POP [ $F(1,19) = 5.78$ ,  $p < 0.05$ ], and PMC [ $F(1,19) = 8.89$ ,  $p < 0.01$ ], with the syllable /pa/ eliciting stronger responses. In the right hemisphere, similar main effect of syllable was observed for the responses at ~100 ms in the POP [ $F(1,19) = 7.80$ ,  $p < 0.05$ ]. Looking at the responses to the intact and noise-embedded stimulus type separately revealed significant differences between the syllable type responses only for the intact stimulus type in the left POP at ~200 ms [ $F(1,19) = 8.51$ ,  $p < 0.01$ ] and left PMC at ~200 ms [ $F(1,19) = 5.27$ ,  $p < 0.05$ ], and only for the noise-embedded stimulus type in the right POP at ~100 ms [ $F(1,19) = 11.5$ ,  $p < 0.01$ ], left HG at ~100 ms [ $F(1,19) = 14.4$ ,  $p < 0.01$ ], and left HG at ~200 ms [ $F(1,19) = 9.88$ ,  $p < 0.01$ ]. The left PMC responses at ~200 ms showed also an interaction for condition x syllable [ $F(3,57) = 7.64$ ,  $p < 0.001$ ], with significant differences between the passive and covert repetition conditions elicited by the syllable /ta/ [ $F(1,19) = 13.7$ ,  $p < 0.01$ ] and significant differences between the overt repetition and imitation conditions elicited by the syllable /pa/ [ $F(1,19) = 5.60$ ,  $p < 0.05$ ].

*Behavioral results*

Identification accuracy was measured as the ratio of correctly vs. incorrectly identified noise-embedded syllables (/pa/ vs. /ta/) in the overt repetition condition (Mean  $d'$  = 1.29,  $SD = 0.95$ ; Mean percent correct = 70.4% (62.4% for /pa/, 78.0% for /ta/),  $SD = 13.6\%$ ). A Spearman rank correlation test revealed that the amplitude of the left PMC responses at ~100 ms after stimulus onset had a positive

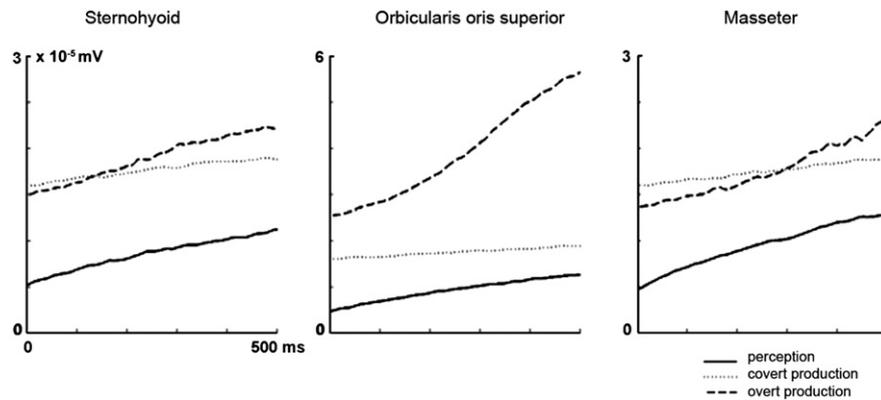
correlation with the accuracy at which the noise-embedded syllables were identified ( $r = 0.45$ ,  $p < 0.05$ ). An additional correlation was calculated between the average amplitude of only the correctly categorized trials and behavioral categorization accuracy. This test revealed positive correlations between source amplitudes and categorization accuracy in the left PMC at ~100 ms ( $r = 0.46$ ,  $p < 0.05$ ), left POP at ~200 ms ( $r = 0.49$ ,  $p < 0.05$ ), and left HG at ~200 ms ( $r = 0.52$ ,  $p < 0.05$ ). The grand-average change in the pitch ( $F_0$ ) across the participants' response tokens from repetition to imitation was 6.4 Hz ( $SD = 10.8$  Hz) towards the  $F_0$  of stimulation. Further, a significant positive correlation was observed between the “unusualness” of production (i.e., the individual difference in the  $F_0$  measured across the response tokens between overt repetition and imitation conditions) and the amplitude of the left PMC responses to the intact stimulus type at ~200 ms after stimulus onset in the imitation condition ( $r = 0.54$ ,  $p < 0.05$ ).

*EMG responses*

Fig. 6 shows grand-averaged absolute value EMG responses of the three recorded articulators (i.e., sternohyoid, orbicularis oris superior, and masseter). As expected, the EMG signals from all three articulators were weaker during the perception compared to the covert and overt production of the syllables.

**Discussion**

In the present study, we analyzed MEG responses to speech syllables during speech motor tasks vs. passive listening to test whether, and at which latencies, the auditory dorsal stream is involved in speech perception. Our results demonstrate that engaging in a phonetic identification/repetition task subsequent to perception enhances electromagnetic responses to syllables within the dorsal processing stream, thus suggesting task-dependency in the sensori-motor interactions during speech perception (see Figs. 2–5). In an initial ECD analysis, this dorsal stream involvement was manifested as enhanced amplitude and significantly more posterior position of the



**Fig. 6.** EMG responses. Grand-averaged absolute value EMG responses of the recorded articulators during perception and production. The signal time-locked to the stimulus onset (perception) is extracted from the covert repetition condition and the signals time-locked to the onset of the signaled motor task (covert/overt production) are from the covert repetition and overt imitation tasks.

left-hemisphere N100m when the perception of the syllables was followed by a subsequent covert repetition compared to passively listening to the syllables (Figs. 2 and 3). The more posterior ECD position suggests that the center-of-gravity of neural activity along the superior temporal plane shifted posteriorly during the speech motor task. A spatially more refined source analysis using cortically constrained MNEs demonstrated that the responses at ~100 ms from the predefined cortical locations of the left-hemisphere SMG and AG were significantly enhanced in the covert repetition compared to the passive condition (Fig. 4). Additionally, the longer-latency responses at ~200 ms were amplified in the covert repetition condition in the left AG (although significantly only for the noise-embedded stimulus type) and in all three active conditions in the left PMC (Fig. 5). Importantly, no such effects were observed in the right-hemisphere regions, which is in accord with the general conception of the dorsal stream being strongly left-hemisphere dominant (e.g., Hickok and Poeppel, 2007).

Evidence for increased dorsal stream involvement as a function of stimulus degradation was also found. First, both the N100m ECD location shift and amplitude increase were greater for the noise-embedded compared to the intact stimulus type between the passive and covert repetition conditions in the left hemisphere (Figs. 2 and 3b). Similarly, the MNEs at ~200 ms showed greater amplification of peak source amplitude for the noise-embedded stimulus type in the covert repetition compared to the passive condition in the left hemisphere AG and PMC (Fig. 5), and the MNEs at ~100 ms in the left AG exhibited a main effect of stimulus type with the noise-embedded stimulus type eliciting greater source amplitudes. Curiously, however, the intact stimulus type elicited greater 100-ms peak source amplitudes in the left hemisphere PT and SMG.

Further, the peak response amplitudes at ~100 ms in the left hemisphere SMG and AG exhibited opposite main effects of stimulus type (i.e., noise-masking enhanced source amplitudes in the left-hemisphere AG yet had a decreasing effect in the adjacent SMG), thus suggesting a functional segregation within the IPL. This observation is in concordance with recent studies showing that perception of spectrally degraded speech was associated with increased activity specifically in the AG rather than SMG (Obleser and Kotz, 2010; Obleser et al., 2007). The finding that the intact stimulus type elicited stronger responses in the left compared to right HG (and SMG) at ~100 ms after stimulus onset is also in line with previous studies (Davis et al., 2008). Interestingly, in the AG, the right-hemispheric responses at ~200 ms to both stimulus types were stronger compared to the left-hemispheric responses. Support for the earlier studies showing that the right auditory cortex would be sensitive to noise-masked speech (e.g., Shtyrov et al., 1998, 1999) were observed only

in the form of shorter N100m peak latencies for the noise-embedded compared to the intact stimulus type. The only right-hemispheric region shown to exhibit stronger responses to the noise-embedded stimulus type was SMG. The observation that the N100m response amplitudes were insensitive to the speech sounds with transient noise masks is in accordance with Miettinen et al. (2011). The current findings together with the findings of earlier studies suggest that the cortical processing of speech sounds is highly dependent on the means with which the distortion is generated; that is, whether it is achieved by internal spectral degradation or by the use of external continuous or transient noise masks. The stronger responses elicited by intact /pa/ compared to /ta/ syllables in the left hemisphere POP and PMC at ~200 ms is in line Binder et al. (2004) and suggest that these regions appear strongly sensitive to the intelligibility of the stimulus.

The behavioral results demonstrated that the amplitude of the left premotor responses at ~100 ms after stimulus onset correlated with the accuracy at which the noise-embedded syllables were identified. This is in concordance with the results of a recent study where correlation was reported between premotor cortex activity and perceptual performance (Callan et al., 2010). As a novel finding, the present results suggest that the premotor cortex is associated with phonetic categorization at a relatively early latency of ~100 ms from stimulus onset. While a latency this short might sound counterintuitive, the first auditory cortex responses to sounds do occur as early as at ~25 ms (Raij et al., 2010) and thus there is time for the auditory inputs to reach prefrontal cortical areas during the first 100 ms. Further, the present findings could be interpreted in the context of recent findings in the visual system, where it was observed that the dorsal stream is responsible for performing a quick initial analysis of visual features resulting in activation of frontal lobe object representations that then feed back to inferior temporal lobe object areas (Bar et al., 2006; Jääskeläinen et al., 2011).

The results from the additional analysis where the average amplitudes of correctly categorized trials were correlated with the behavioral categorization accuracy revealed significant positive correlations, in addition to the effects within premotor cortex at ~100 ms, also between syllable identification accuracy and response amplitudes in the left POP and HG at ~200 ms after stimulus onset, thus tentatively suggesting that the left POP and HG also participate in successful phonetic categorization. Further, the significant positive correlation between the amplitude of the left PMC responses in the imitation condition and the “unusualness” of production (the individual difference in the FO measured across the response tokens between overt repetition and imitation conditions) could explain the difference in the source amplitudes between the two conditions (see Fig. 5), and supports our hypothesis that additional involvement of

the premotor cortex speech sound map is required for imitation that necessitates accurate tracking of articulatory details.

Considering these results in light of a previous study showing that rTMS-stimulation of the left-hemisphere vPMC resulted in slower response times in a phoneme discrimination task with no effect in (less demanding) phoneme identification and syllable discrimination tasks (Sato et al., 2009), it seems that the motor system is recruited to facilitate processing when the perceptual task requires increased working memory or attentional demands. In other words, the difficulty of the perceptual task seems to be a determinant of the motor-induced modulation of speech perception (concordantly with the constructivist approach of speech perception). In addition, supported by another rTMS study showing that stimulation of the left-hemisphere vPMC resulted in decreased ability to identify auditory syllables embedded in white noise (Meister et al., 2007), the observed effects of stimulus type suggest that the role of the motor system (especially that of the PMC) can be further refined to encompass amplification of the auditory signal when resolving of stimulus ambiguity, such as due to noise, is required. However, whether the motor system (or the dorsal stream) constitutes a critical component of speech perception under normal listening conditions remains to be investigated in future studies with more naturalistic listening conditions.

Since differential activation was observed between the passive and active conditions in the left hemisphere PMC during the perception of speech sounds (as reflected in the 200-ms responses, Fig. 5), the results lend support to the constructivist (internal model) approaches of speech perception (Hickok and Poeppel, 2007; Rauschecker, 2011; Rauschecker and Scott, 2009; Schwartz et al., 2010) and, at the same time, contradict those that claim that the extent of the engagement of the speech motor system is dependent not on the task but rather directly on the amount of information in the sensory stimulation (e.g., Fowler, 1996). Indeed, the constructivist approaches assert that the activated motor representations constrain the interpretation of the incoming sensory inputs through a hypothesis testing based on internally generated candidate articulatory categorizations, and that the extent of this process is dependent on the difficulty of the perceptual task.

Regarding the internal model approaches, the observed effects at ~100 ms in the left-hemisphere IPL (SMG and AG) could reflect attention-related changes (e.g., Colby and Goldberg, 1999) and/or an early inverse mapping whereby a quick “sketch” of the incoming sensory information is sent from the pSTG/pSTS to the IPL (Rauschecker and Scott, 2009). Similarly, it can be speculated that the later effect at ~200 ms in the left-hemisphere PMC reflects mapping of incoming auditory signals onto articulatory motor representations. Notably, it is also possible that these effects can be accounted for by a top-down attentional mechanism operating through the inverse and forward mapping within the computational framework of the dorsal stream (Hickok et al., 2011).

The results from the simultaneous EMG recordings (Fig. 6) do not only indicate that the participants were performing the covert repetition task as requested, but also speak against an alternative interpretation of the dorsal stream activity simply reflecting quick covert articulation of the perceived syllables prior to the signaled motor task. Indeed, the EMG activity after the stimulus onset was considerably weaker compared to that during the overt production task, which in turn was relatively similar to the activity during the overt production tasks. Further, it is unlikely that any covert articulation could occur as quickly as in only 100 or 200 ms from syllable onset, the latencies at which the present findings were observed to take place.

In conclusion, our results argue for a close relationship between speech perception and production systems by demonstrating that activity in the speech motor areas is modulated by both perceptual task demands and speech intelligibility. Indeed, activity in the left hemisphere PMC and IPL (SMG and AG) increased when a perception of

a syllable was followed by a subsequent phoneme identification and production task, with further activity enhancements in the PMC and AG when the syllables were embedded in noise. Moreover, the behavioral results showed that early-latency left PMC activation is associated with syllable identification accuracy. These results have significant implications in that they, at the same time, lend support for the activity observed in the speech production system during speech perception being causally related to perceptual performance as well as provide millisecond-scale temporal information on the latencies at which this motor contribution takes place. The behavioral results also showed that the “unusualness” of production correlated with longer-latency left PMC activation, supporting our hypothesis that additional involvement of the premotor cortex speech sound map is required during perception which is followed by imitative repetition (necessitating more accurate tracking of articulatory details). Finally, considering the behavioral results and the differential activation of the dorsal stream regions between the passive and active conditions, the results (1) argue for the existence of a motor component to speech perception; (2) support the constructivist theory of speech perception, and particularly the internal model approaches, which suggest that perceptual task difficulty is reflected as increased activity in the speech motor system; and (3) suggest that the left premotor cortex influences perceptual performance already at an early latency of ~100 ms.

Supplementary materials related to this article can be found online at [doi:10.1016/j.neuroimage.2012.02.011](https://doi.org/10.1016/j.neuroimage.2012.02.011).

## Acknowledgments

This study was financially supported by the Academy of Finland (projects 135009, 130412, and 138145), and by research grants from CNRS (Centre National de la Recherche Scientifique) and ANR (Agence Nationale de la Recherche, ANR SPIM and MULTISTAP) to M.S. and J.-L.S.

## References

- Ahveninen, J., Jaaskelainen, I.P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., Levanen, S., Lin, F.H., Sams, M., Shinn-Cunningham, B.G., Witzel, T., Belliveau, J.W., 2006. Task-modulated “what” and “where” pathways in human auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14608–14613.
- Andersen, R.A., Cui, H., 2009. Intention, action planning, and decision making in parietal-frontal circuits. *Neuron* 63, 568–583.
- Arnott, S.R., Binns, M.A., Grady, C.L., Alain, C., 2004. Assessing the auditory dual-pathway model in humans. *NeuroImage* 22, 401–408.
- Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmid, A.M., Dale, A.M., Hamalainen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., Halgren, E., 2006. Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci.* 103, 449–454.
- Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of sensory and decision processes in auditory object identification. *Nat. Neurosci.* 7, 295–301.
- Boatman, D., 2004. Cortical bases of speech perception: evidence from functional lesion studies. *Cognition* 92, 47–65.
- Boersma, P., 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proc. Inst. Phon. Sci.* 17, 97–110.
- Boersma, P., 2001. Praat, a system for doing phonetics by computer. *Glott Int.* 5, 341–345.
- Callan, D., Callan, A., Gamez, M., Sato, M.A., Kawato, M., 2010. Premotor cortex mediates perceptual performance. *NeuroImage* 51, 844–858.
- Callan, D.E., Jones, J.A., Callan, A.M., Akahane-Yamada, R., 2004. Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage* 22, 1182–1194.
- Coch, D., Sanders, L.D., Neville, H.J., 2005. An event-related potential study of selective auditory attention in children and adults. *J. Cogn. Neurosci.* 17, 605–622.
- Colby, C.L., Goldberg, M.E., 1999. Space and attention in parietal cortex. *Annu. Rev. Neurosci.* 22, 319–349.
- Crowley, K.E., Colrain, I.M., 2004. A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin. Neurophysiol.* 115, 732–744.
- Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., Hari, R., 2000. Speaking modifies voice-evoked activity in the human auditory cortex. *Hum. Brain Mapp.* 9, 183–191.
- Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26, 55–67.

- D'Ausilio, A., Bufalari, I., Salmas, P., Fadiga, L., 2011. The role of the motor system in discriminating normal and degraded speech sounds. *Cortex* doi:10.1016/j.cortex.2011.05.017.
- Davis, C., Kislyuk, D., Kim, J., Sams, M., 2008. The effect of viewing speech on auditory speech processing is different in the left and right hemispheres. *Brain Res.* 1242, 151–161.
- Diehl, R.L., Lotto, A.J., Holt, L.L., 2004. Speech perception. *Annu. Rev. Psychol.* 55, 149–179.
- Fischl, B., 2004. Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22.
- Fowler, C.A., 1996. Listeners do hear sounds, not tongues. *J. Acoust. Soc. Am.* 99, 1730–1741.
- Galantucci, B., Fowler, C.A., Turvey, M.T., 2006. The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377.
- Guenther, F.H., Ghosh, S.S., Tourville, J.A., 2006. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.* 96, 280–301.
- Hamalainen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42.
- Hamalainen, M.S., Sarvas, J., 1989. Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. *IEEE Trans. Biomed. Eng.* 36, 165–171.
- Hamalainen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* 65, 413.
- Helmholtz, H.V., 1867. *Handbuch der physiologischen Optik*. Leopold Voss, Leipzig.
- Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Hickok, G., Houde, J., Rong, F., 2011. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422.
- Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., Evans, A.C., 1998. Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr.* 22, 324–333.
- Jaaskelainen, I.P., 2010. The role of speech production system in audiovisual speech perception. *Open Neuroimaging J.* 4, 30–36.
- Jääskeläinen, I.P., Ahveninen, J., Andermann, M.L., Belliveau, J.W., Raji, T., Sams, M., 2011. Short-term plasticity as a neural mechanism supporting memory and attentional functions. *Brain Res.* 1422, 66–81.
- Kauramäki, J., Jaaskelainen, I.P., Hari, R., Mottonen, R., Rauschecker, J.P., Sams, M., 2010. Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. *J. Neurosci.* 30, 1314–1321.
- Kudo, N., Nakagome, K., Kasai, K., Araki, T., Fukuda, M., Kato, N., Iwanami, A., 2004. Effects of corollary discharge on event-related potentials during selective attention task in healthy men and women. *Neurosci. Res.* 48, 59–64.
- Liberman, A.M., Mattingly, I.G., 1985. The motor theory of speech perception revised. *Cognition* 21, 1–36.
- Liberman, A.M., Whalen, D.H., 2000. On the relation of speech to language. *Trends Cogn. Sci.* 4, 187–196.
- Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. *Psychol. Rev.* 74, 431–461.
- Liikkanen, L.A., Tiitinen, H., Alku, P., Leino, S., Yrttiaho, S., May, P.J., 2007. The right-hemispheric auditory cortex in humans is sensitive to degraded speech sounds. *Neuroreport* 18, 601–605.
- Lin, F.H., Belliveau, J.W., Dale, A.M., Hamalainen, M.S., 2006. Distributed current estimates using cortical orientation constraints. *Hum. Brain Mapp.* 27, 1–13.
- Makela, A.M., Alku, P., May, P.J., Makinen, V., Tiitinen, H., 2005. Left-hemispheric brain activity reflects formant transitions in speech sounds. *Neuroreport* 16, 549–553.
- Meister, I.G., Wilson, S.M., Deblieck, C., Wu, A.D., Iacoboni, M., 2007. The essential role of premotor cortex in speech perception. *Curr. Biol.* 17, 1692–1696.
- Miettinen, I., Tiitinen, H., Alku, P., May, P.J., 2010. Sensitivity of the human auditory cortex to acoustic degradation of speech and non-speech sounds. *BMC Neurosci.* 11, 24.
- Miettinen, I., Alku, P., Salminen, N., May, P.J., Tiitinen, H., 2011. Responsiveness of the human auditory cortex to degraded speech sounds: reduction of amplitude resolution vs. additive noise. *Brain Res.* 1367, 298–309.
- Mottonen, R., Watkins, K.E., 2009. Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819–9825.
- Naatanen, R., Picton, T., 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.
- Nishitani, N., Hari, R., 2002. Viewing lip forms: cortical dynamics. *Neuron* 36, 1211–1220.
- Novak, G., Ritter, W., Vaughan Jr., H.G., 1992. Mismatch detection and the latency of temporal judgements. *Psychophysiology* 29, 398–411.
- Obleser, J., Kotz, S.A., 2010. Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb. Cortex* 20, 633–640.
- Obleser, J., Elbert, T., Lahiri, A., Eulitz, C., 2003. Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Brain Res. Cogn. Brain Res.* 15, 207–213.
- Obleser, J., Wise, R.J., Alex Dresner, M., Scott, S.K., 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. *J. Neurosci.* 27, 2283–2289.
- Ojanen, V., Mottonen, R., Pekkola, J., Jaaskelainen, I.P., Joensuu, R., Autti, T., Sams, M., 2005. Processing of audiovisual speech in Broca's area. *NeuroImage* 25, 333–338.
- Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I.P., Mottonen, R., Tarkiainen, A., Sams, M., 2005. Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport* 16, 125–128.
- Poeppel, D., Phillips, C., Yellin, E., Rowley, H.A., Roberts, T.P., Marantz, A., 1997. Processing of vowels in supratemporal auditory cortex. *Neurosci. Lett.* 221, 145–148.
- Pulvermüller, F., Huss, M., Kherif, F., del Prado, Moscoso, Martin, F., Hauk, O., Shtyrov, Y., 2006. Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U. S. A.* 103, 7865–7870.
- Raji, T., Ahveninen, J., Lin, F.H., Witzel, T., Jaaskelainen, I.P., Letham, B., Israeli, E., Sahyoun, C., Vasios, C., Stufflebeam, S., Hamalainen, M., Belliveau, J.W., 2010. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *Eur. J. Neurosci.* 31, 1772–1782.
- Rauschecker, J.P., 2011. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear. Res.* 271, 16–25.
- Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.
- Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11800–11806.
- Romero, L., Walsh, V., Papagno, C., 2006. The neural correlates of phonological short-term memory: a repetitive transcranial magnetic stimulation study. *J. Cogn. Neurosci.* 18, 1147–1155.
- Ross, B., Tremblay, K., 2009. Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hear. Res.* 248, 48–59.
- Sams, M., Hamalainen, M., Antervo, A., Kaukoranta, E., Reinikainen, K., Hari, R., 1985. Cerebral neuromagnetic responses evoked by short auditory stimuli. *Electroencephalogr. Clin. Neurophysiol.* 61, 254–266.
- Sato, M., Tremblay, P., Gracco, V.L., 2009. A mediating role of the premotor cortex in phoneme segmentation. *Brain Lang.* 111, 1–7.
- Schwartz, J.-L., Sato, M., Fadiga, L., 2008. The common language of speech perception and action: a neurocognitive perspective. *Rev. Fr. Linguist. Appl.* 13, 9–22.
- Schwartz, J.-L., Basirat, A., Ménard, L., Sato, M., 2010. The perception-for-action-control theory (PACT): a perceptuo-motor theory of speech perception. *J. Neurolinguist.* doi:10.1016/j.neuroling.2009.12.004.
- Scott, S.K., 2005. Auditory processing—speech, space and auditory objects. *Curr. Opin. Neurobiol.* 15, 197–201.
- Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26, 100–107.
- Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R.J., Naatanen, R., 1998. Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: magnetic mismatch negativity study. *Neurosci. Lett.* 251, 141–144.
- Shtyrov, Y., Kujala, T., Ilmoniemi, R.J., Naatanen, R., 1999. Noise affects speech-signal processing differently in the cerebral hemispheres. *Neuroreport* 10, 2189–2192.
- Skipper, J.I., Nusbaum, H.C., Small, S.L., 2005. Listening to talking faces: motor cortical activation during speech perception. *NeuroImage* 25, 76–89.
- 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.
- Watkins, K., Paus, T., 2004. Modulation of motor excitability during speech perception: the role of Broca's area. *J. Cogn. Neurosci.* 16, 978–987.
- Whiting, K.A., Martin, B.A., Stapells, D.R., 1998. The effects of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da. *Ear Hear.* 19, 218–231.
- Wilson, S.M., Iacoboni, M., 2006. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *NeuroImage* 33, 316–325.
- Wilson, S.M., Saygin, A.P., Sereno, M.I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702.
- Zekveld, A.A., Heslenfeld, D.J., Festen, J.M., Schoonhoven, R., 2006. Top-down and bottom-up processes in speech comprehension. *NeuroImage* 32, 1826–1836.