Publication 3

Sanna Malinen and Riitta Hari. 2010. Comprehension of audiovisual speech: Data-based sorting of independent components of fMRI activity. Espoo, Finland: Aalto University School of Science and Technology. 14 pages. Helsinki University of Technology, Low Temperature Laboratory Publications, Report TKK-KYL-023. ISBN 978-952-60-3170-5.

Comprehension of audiovisual speech: Data-based sorting of independent components of fMRI activity

Sanna Malinen and Riitta Hari

Brain Research Unit, Low Temperature Laboratory,
Advanced Magnetic Imaging Centre,
Aalto University of Science and Technology,
P.O. Box 15100
FI-00076 Aalto, Espoo, Finland

Report TKK-KYL-023

May 2010

Abstract

To find out whether decreased intelligibility of natural audiovisual speech would enhance the involvement of the sensorimotor speech comprehension network, we let ten adults to listen and view a video, in which the intelligibility of continuous audio-visual speech was varied by altering the loudness between well-audible, just audible (–18 dB weaker than loud), and mute. Occasionally the voice and articulations were time-reversed, or the whole stimulus was replaced by tone pips of varying intensity. We combined inter-subject correlation (ISC) and independent component analyses (ICA) of 3-T functional magnetic resonance imaging (fMRI) data in a novel manner by sorting the components by their spatial overlap with a stimulus-specific ISC map and thereby unraveling the most stimulus-related networks. Unlike the general linear model analysis, which revealed only temporal- and parietal-lobe activations, the ISC–ICA approach pinpointed several cortical networks related to audiovisual speech processing. A left-hemisphere-dominant network—comprising the left posterior superior temporal sulcus, inferior frontal gyrus, anterior superior temporal gyrus, premotor cortex, and right inferior parietal lobule—was related to comprehension of natural audiovisual speech, with increased activation during decreased speech loudness. Since the ISC-map provides a stimulus-specific spatial template for sorting the ICs, the analysis requires no predetermined temporal models on stimulus timing. This analysis approach is therefore feasible for fMRI studies where hemodynamic variations related to naturalistic and continuous stimuli are difficult to predict.

1 Introduction

Brain areas related to processing of speech, studied mainly by applying syllables, words, and sentences, are relatively well-known [1]. Instead, brain activation to continuous long-lasting speech [2, 3] with variable signal-to-noise ratio has quite rarely been investigated. In noisy every-day environments speech understanding can be impaired and the listener has to comprehend the message with the aid of visual information from the speaker's articulation movements. In such conditions enhanced effort is required from the subject to tie up the message. Consequently, one would expect increased activation of the whole sensorimotor network supporting speech comprehension.

In the present study we tested this hypothesis by recording the listener's brain activity when the intelligibility of audiovisual speech was impaired by decreasing the loudness of the speaker's voice. The subjects listened and viewed continuous, audiovisual speech during functional magnetic resonance imaging (fMRI).

The conventional fMRI analysis, based on temporal covariates, may suit poorly for this kind of real-life-like experimental setups, since it is difficult to identify the most relevant features in the continuously varying stimulus flow and thereby to predict the complex hemodynamic variations. We therefore analyzed the fMRI data with independent component analysis (ICA), which is a data-driven technique to estimate both spatially independent brain regions (or networks of several areas) and their time courses. Importantly, ICA requires no a priori models of signal temporal behavior. In group ICA, developed by Calhoun and coworkers [4] and applied *e.g.* in our previous study with complex naturalistic stimuli [5], independent components (ICs) are calculated from a concatenated data set of different individuals and later individual ICs and their time courses can be back-reconstructed for statistical inferences.

The estimated independent components do not appear in any specific order, and therefore it is a major challenge to select—among the large set of ICs—those components that are related to a given stimulus or task. Again, in experimental setups with naturalistic and continuously varying stimulation, the brain's hemodynamic variations are not easily predicted and, thus, selection based on temporal covariates may be difficult. In such cases the selection of task-related ICs beyond the primary projection areas may be especially challenging.

So far the most popular quantitative methods to sort the ICs have been the temporal correlates between stimulus features and IC time courses [6–11], Spatial sorting, on the other hand, has been utilized especially in resting-state studies to segregate brain areas of interest with the aid of a brain atlas [12, 13]. However, the sorting based on an atlas requires strong *a priori* assumptions on the brain areas or networks that are known, or assumed, to be involved in the brain processing of the particular phenomenon.

We here introduce a new approach for sorting the ICs on the basis of an inter-subject correlation (ISC) map that we propose to provide a functional template for sorting the ICs according to their stimulus-dependence during processing of audiovisual speech of different levels of intelligibility. ISC map is calculated by correlating fMRI signals voxel-by-voxel such that the signal at a certain brain area of one individual is correlated with a signal of a corresponding location another individual who had been subjected to the same experimental condition. The motivation for the ISC approach emerges from findings that the voxel-by-voxel correlation analysis effectively describes the similarity and temporal synchrony between subjects [14]; moreover, it has been shown to reveal the extrinsic, stimulus-related cerebral networks [15, 16]. Since the spatial pattern of the ISC map should be unique for each stimulus series, it can be used as a stimulus-specific functional template that absolves the ordering of stimulus related-ICs of any temporal criteria or user-defined anatomical regions of interest.

Previously, ISC has shown superior performance compared with temporal-covariate-based fMRI analysis during narrative speech comprehension in normal hearing conditions [17]. The assumption about temporal synchrony across individuals is also supported by the

observed high inter-subject correlations of IC time courses in certain brain regions during natural stimulation [18].

An ISC map can, however, cover spatially extensive brain areas. ICA, on the other hand, reveals maximally independent brain networks with related time courses. Therefore, our combination of ICA with ISC map identified *(i)* the brain areas that behave temporally in a similar manner across individuals, and *(ii)* the independent sub-networks within the spatially more extensive correlation map. We computed the ISC maps to first reveal all brain areas that respond to the whole audiovisual stimulation, and then used the combined ISC–ICA to select ICs for more detailed analysis of brain networks that are related to comprehension of the audiovisual speech. Results were also compared with more traditional general-linear-model (GLM) based analysis.

2 Materials and methods

2.1 Subjects

Ten healthy adults (3 females, 7 males; mean age 26.7 ± 4.3 years, range 21–33 years) participated in the experiment after their written informed consent. The study had prior approval by the Ethics Committee of Helsinki and Uusimaa Hospital District. Eight subjects were right-handed, with laterality quotient of 87.7 ± 17.7 (mean ± SD) in the Edinburgh handedness inventory [19], and two subjects were left-handed (–56.1 ± 12.2).

2.2 Stimuli

During fMRI scanning, the subjects were presented with audiovisual stimuli comprising a video of a male actor (whose upper body was visible on a black background). The actor was continuously speaking, describing features of either human faces, hands, or built environment. The sound track was modified in 16–18-s blocks: The voice was either soft, that is just audible during the noisy echo-planar imaging (EPI) scanning, loud (18 dB louder than the soft stimuli), or mute. During soft normal speech, the subjects were able to hear the speech, but understood the message only with effort (see section “Subject reports on stimulus loudness” below).

These three levels of loudness were counterbalanced in the stimulation sequence. Two times within the sequence, the normal speech was time-reversed, or the face was absent and the voice was replaced by tone pips. The pitch of the 0.1-s tone pips varied randomly between 250, 500, 1000, 2000, and 4000 Hz, and the loudness of the sounds varied to the same extent as the speech. The total stimulation sequence lasted for 8 min 30 s, and it was presented twice to every subject in successive imaging runs. The subjects’ task was to listen to and view continuous speech, trying to understand the message.

Stimuli were delivered using Presentation® software (version 0.81, http://www.neurobehavioralsystems.com). Stereo sound was presented binaurally with UNIDES ADU2a audio stimulators (Unides Design, Helsinki, Finland) and videos were projected (projector Vista X3 REV Q, Christie Digital Systems, Canada, Inc.) via a mirror to a transparent screen placed behind the subject. The frame rate of the videos 25 frames/s and image resolution 726 × 576 pixels.

2.3 Imaging

MR images were collected with Signa VH/i 3.0 T MRI scanner (General Electric, Milwaukee, WI, USA). Functional images were acquired using twice the gradient EPI sequence with time to repeat (TR) 3 s, time to echo (TE) 32 ms, 44 oblique axial slices, matrix 64 × 64, voxel size $3 \times 3 \times 3$ mm^3, field of view (FOV) 20 cm, flip angle 90°, 198 volumes.

Structural images were scanned with 3-D T1 spoiled gradient imaging, matrix 256 × 256, TR 9 ms, TE 1.9 ms, flip angle 15°, preparation time 300 ms, FOV 26 cm, slice thickness 1.4 mm, and number of excitations 2.

2.4 Preprocessing of fMRI data

The first 170 volumes from both series were selected for further analysis. Data were preprocessed using SPM2 software (http://www.fil.ion.ucl.ac.uk/spm/), including realignment, normalization with skull stripping (http://imaging.mrc-cbu.cam.ac.uk/imaging/NormalizeSkullStripped) into MNI space and smoothing with a 6-mm (full width half maximum) Gaussian filter.

Individual structural scans were segmented using FreeSurfer software (http://surfer.nmr.mgh.harvard.edu/) and the resulting gray matter images were transformed into the MNI space. These gray-matter masks were later used for calculation of inter-subject correlations.

2.5 ICA and the sorting of ICs using ISC map

Altogether 41 ICs were extracted from the preprocessed data using group-ICA toolbox GIFT (http://icatb.sourceforge.net/) with FastICA algorithm. The number of independent components was estimated using minimum-description-length algorithm [20] inbuilt in GIFT.

Before calculating the ISC map, the six realignment parameters and the whole-volume-average signal were fitted to the data of individual gray matter voxels and their effect was removed from the data. Correlation images were calculated, using the first image series, between all 45 subject pairs formed by the 10 subjects: First the Pearson's correlation coefficients were determined voxel-by-voxel and, then, Fisher transformation was applied to convert the correlation coefficients to the normally-distributed variables.

Correlation images for all subject pairs were pooled to search for group-level statistically significant correlations. *T*-statistics exceeding the threshold $p < 0.01$ (false discovery rate corrected) and the extent of 10 voxels (see Figure 1) formed the functional template for IC ordering.

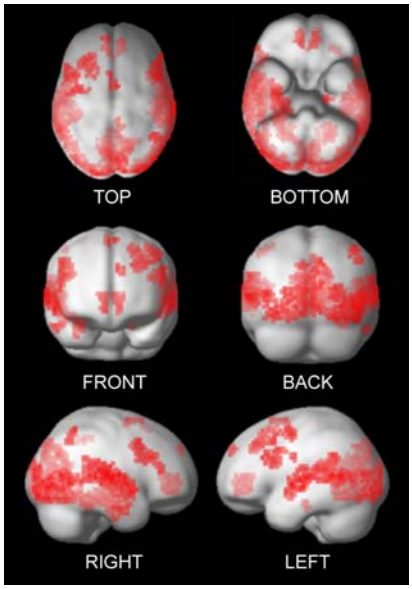

Figure 1. Inter-subject correlation map (ISC) used in the sorting of ICs. Statistically significant correlations are overlaid on a smooth average map (SPM2 template).

For each of the 41 ICs we determined a sorting parameter, defined as the equally weighted sum of the spatial correlation coefficient between the ISC and mean IC ($z > 3.09$) and the number of overlapping voxels relative to the total amount of voxels in the ISC map. The amount of overlapping voxels was determined from binary images: voxel values exceeding a given threshold were set to 1 and others to 0.

2.6 General linear model

An additional analysis, based on the general lineal model (GLM), was later used in SPM2 to compare the observations between data- and model-based approaches. The model *i.e.* design matrix included 8 separate regressors for normal (S) and reversed speech (R) according to their volume; loud (SL/RL), soft (SS/RS), or mute (SM/RM), and for loud and soft tone pips (TL and TS, respectively). Boxcar models were convolved with hemodynamic response function, fMRI time series were high-pass filtered at (1/420 s), and AR(1) model was applied for serial correlations. The mean effect of both stimulus series was statistically tested. Individual contrast images were subjected to random-effects analysis (one sample *t*-test). Activation clusters of 10 voxels or more, each voxel at $p < 0.001$ were considered as statistically significant activation.

3 Results

3.1 Subject reports on stimulus loudness

After the scanning, all subjects were interviewed and asked about how well they heard the voice and/or understood the speech message during loud and soft stimulus presentations. Every subject had heard and understood the loud normal speech with no difficulty. Similarly, all subjects reported that they did hear the soft speech, but 8 out of 10 subjects needed extra effort to understand the message.

3.2 ISC map

The ISC map (Fig. 1) covered widely the occipital visual areas, the superior temporal lobes symmetrically and the lower temporal lobes with right-hemisphere dominance, the precentral areas in a left-hemisphere dominant manner, the inferior frontal gyrus (IFG) in both hemispheres, the medial prefrontal cortex, the right inferior parietal lobe (IPL), and the fusiform area (in the picture projected to the surface of the cerebellum; see the brain viewed from bottom).

3.3 Stimulus-related ICs

Figure 2 shows the 10 most stimulus-related group ICs on three orthogonal slices around their activation maxima; the components are ordered from the most (IC1) to the least (IC10) overlap with the ISC map presented in Fig 1. These 10 ICs with the highest sorting parameters comprised 75% of the sum of all sorting parameters of those ICs that correlated positively with it ISC map (21 out of 41).

IC1 covers the auditory cortex in the superior surface of both temporal lobes, extending along the superior temporal gyrus (STG), and including Heschls's gyrus in both hemispheres.

IC2 coincides with the visual-motion area V5/MT, and IC3 covers the region of the parieto-occipital sulcus (POS). ICs 4 and 5 agree with lateral and central early visual cortices (V1/V2). ICs 2, 4, and 5 reacted to every visual stimulus and their activation was suppressed during tone pips. Instead, IC3 in POS reacted not only to each visual stimulus block, but also at the beginning of the tone-pip series, when the visual stimulus disappeared.

IC6 comprises an extensive cortical network, with activation maxima in the left posterior superior temporal sulcus (STS), left anterior STG, left inferior frontal gyrus (Broca's area), left premotor cortex (BA6), and right IPL (two latter visible in Fig. 3). IC7 comprises posterior cingulate cortex (PCC), medial frontal gyrus (MFG, BA10), right IFG, and right IPL. IC8 shows activation in the right insula and bilaterally in the fusiform gyri and collateral sulci. IC9 covers both insulae and middle cerebellum (see Fig. 3). Finally, IC10 covers the MFG, IFG, IPL and middle temporal gyrus (MTG, not visible in the figure) in the right hemisphere.

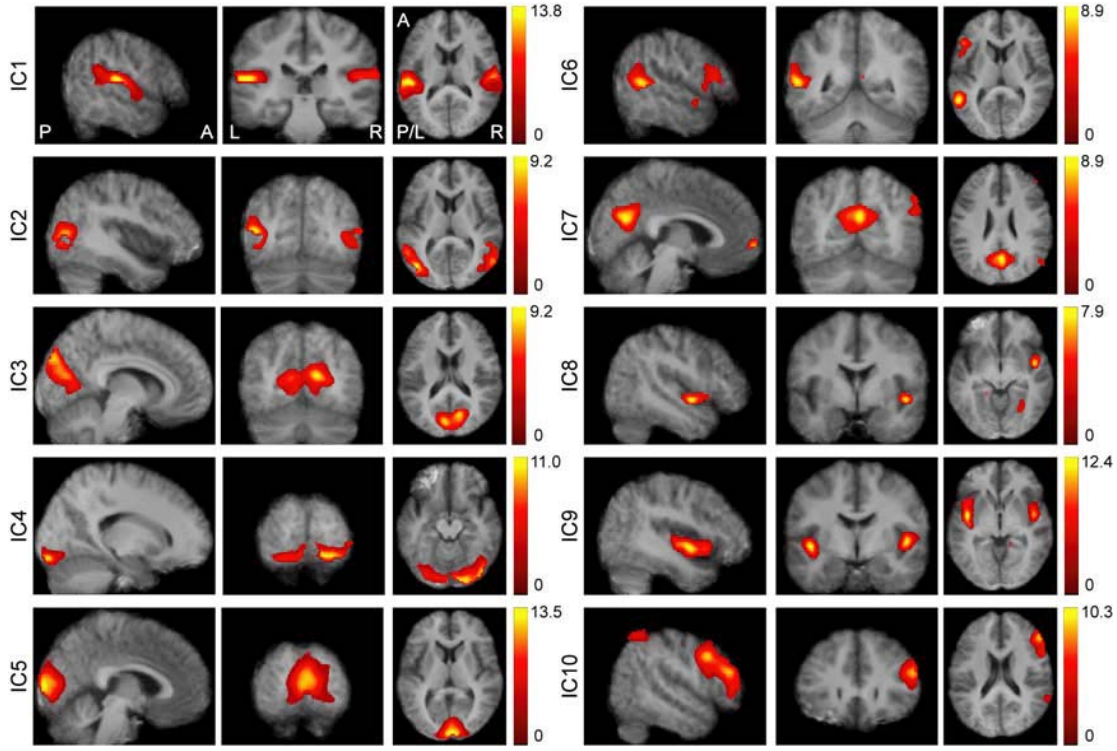

Figure 2. The ICs selected using the ISC map; the components IC1–IC10 are ordered according to their overlap (from largest to smallest) with the map. The color bars indicate *z* scores. P/A, posterior/anterior; L/R, left/right.

Figure 3 shows three ICs (1, 6, and 10; red, green and blue, respectively) in more detail and the corresponding time behaviors. The auditory-cortex IC1 (red) reacts to each sound, and the louder the sound, the stronger the response. Notably, IC1 (and thus the auditory cortex) reacts strongest for the loud reverse speech.

IC6 (green), that is the posterior left STS–anterior STG–IFG–BA6–right IPL network, is sensitive to natural (non-reversed) speech, with no or weak responses for reversed speech and tone pips. It reacts both to well-hearable and soft speech and, occasionally even more strongly for soft than loud and mute sounds.

IC10 (blue), a right-lateralized network comprising right MFG, IFG, IPL and MTG, reacts more to the reversed speech than does the left-hemisphere IC6 network. Moreover, IC10 shows a notable 0.06 Hz periodicity, clearly related to block periodicity. Similarly, the IC3 (POS) and IC8 (right insula, fusiform gyri, and collateral sulci) had most of their signal power at 0.06 Hz.

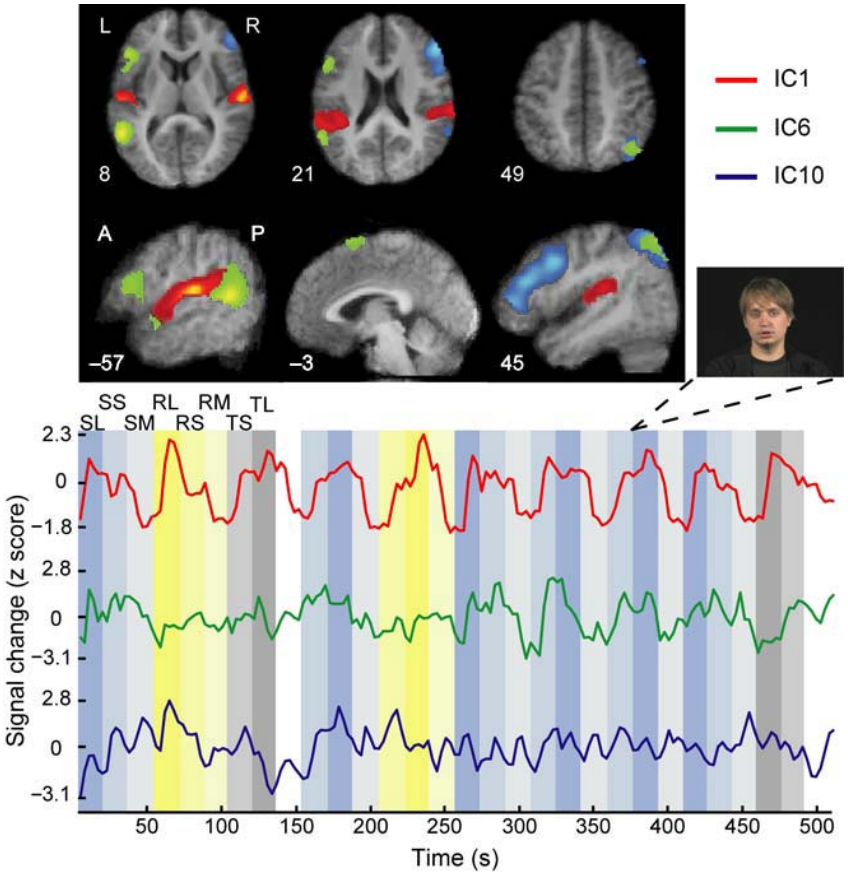

Figure 3. ICs related to auditory or audiovisual speech processing. The color of the time trace corresponds to the color of the spatial IC. Behind the traces, the color bars indicate the stimulus order: SL = speech loud, SS = speech soft, SM = speech mute, RL = reverse speech loud, RS = reverse speech soft, RM = mute reverse speech, TL = tones loud, TS = tones soft. Numbers next to axial and sagittal slices are MNI *z*- and *x*-coordinates, respectively.

3.4 GLM analysis

We tested certain selected hypotheses, arising from the ICA results, also with a separate model-based analysis. ICA suggested that: *(i)* the auditory cortex (in IC1) reacts stronger to loud stimuli than to other stimulus levels. This observation was tested with three separate contrasts, one for each stimulus type: SL > SS + SM, RL > RS + RM and TL > TS, *(ii)* auditory-cortex activation is stronger for loud reversed speech than loud normal speech (contrast RL > SL), *(iii)* IC6 indicated sensitivity to normal speech (contrasts S > R and S > T), and *(iv)* IC6 indicated also stronger responses for soft than loud or mute speech (SS > SL + SM). In addition to these observations, we searched for areas that were more active during silent visual speech than during heard speech (SM > SL + SS).

Figure 4a (focusing on right-hemisphere activation) shows that loud sounds activated the auditory cortex bilaterally more than did soft sound and silence; the same was true for reversed speech and tone pips. Figure 4b shows that auditory cortex was activated more for loud reverse than loud normal speech. Beyond the bilaterally activated auditory cortex, other activations were highly right-lateralized. Additional clusters, seen in areas covered also by IC10, were in the right MTG, right IFG, and right IPL.

Figure 4c shows that the left posterior STS was activated more for normal natural than reversed speech (S > R), and more fore speech than tone pips (S > T). The same brain region, together with the right IPL, was also activated more for soft speech than for loud or silent speech. No statistically significant differences were observed between soft *vs.* other levels of reversed speech or tone pips.

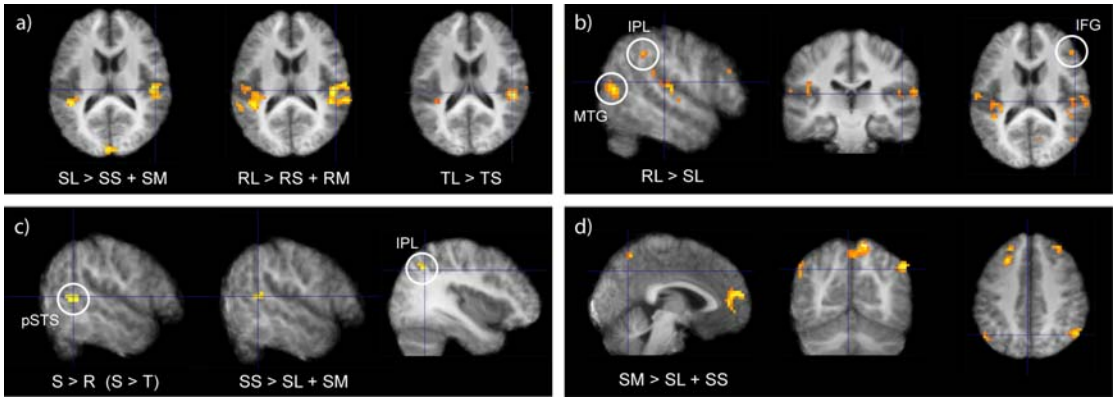

Figure 4. Results from GLM analysis. a) Auditory cortex activations for comparison of loud stimulus with other stimulus levels, b) activations elicited by reversed speech compared with natural speech, c) activations related to natural speech processing and decreased intelligibility, and d) silent speech compared with loud and soft speech. Abbreviations: S = natural speech, R = reversed speech, T = tone pips, MTG = medial temporal gyrus, IPL = inferior parietal lobule, IFG = inferior frontal gyrus (in the right hemisphere), pSTS = posterior superior temporal sulcus (in the left hemisphere), for others see Fig. 3.

Figure 4d shows that silent *vs.* loud and intermediate speech activated IPL and medial prefrontal cortex bilaterally, ACC (BA10), and the precuneus adjacent to PCC. The IPL, MFG and BA10 activations were also seen in the IC7.

4 Discussion

Modification of the intelligibility of audiovisual speech unraveled *(i)* a left-hemisphere-lateralized sensorimotor network for natural speech comprehension, with equal or occasionally even enhanced activation during decreased speech intelligibility, and *(ii)* a right-lateralized network, which was more sensitive for time-reversed than normal speech. These results relied on ICs that overlapped with the stimulation-related ISC map. Certain hypotheses were also tested with GLM-based analysis, which replicated the findings on the right-hemisphere network, but showed only parts of the left-lateralized sensorimotor speech comprehension network: namely the temporal and parietal areas, but no premotor activation.

4.1 IC sorting with ISC map

In our new IC-sorting procedure, the ISC map served as a stimulation-specific functional template and it helped to reveal brain areas that reacted similarly across subjects to the external stimuli. Thereby ICA offered more detailed information of the spatially independent sub-networks within this extensive ISC map. The ISC map is a conservative approximation of the extent of the extrinsic processing circuitry, and it includes neuronal networks that are modulated similarly across subjects, either directly by the stimulus, or indirectly via stimulus-related task or stimulus-induced cognitive or motor processing.

The strength of the combined ISC–ICA approach is that it does not require assumptions about the stimulation features nor predetermined spatial locations (regions of interest) of the activations. Instead, the underlying assumptions are related to the methods themselves: similar temporal modulation of responses in corresponding brain areas across subjects in ISC, and the requirement of spatial independence and linearity in ICA. Consequently, sorting the ICs with the help of the ISC map provided a relatively simple and straightforward way to facilitate the selection of stimulus-related ICs from the whole set of ICs, although the stimulation was complex and continuously changing.

The apparent difficulty of the present approach is, however, that user still needs to decide how many of the ordered components are selected for further consideration. Here, our threshold was set such that selected ICs needed to explain 75% of the total sum of all sorting parameters. One possibility to overcome this difficulty would have been to constrain the ICA

to brain areas covered by the ISC map. Then, all ICs would have been known to express activity common to the whole subject group and thus most likely related to processing of the external stimulation. We, however, decided to apply ICA totally independently of the ISC analysis. ICA, which requires statistical independence utilizing higher-order statistics, can theoretically detect nodes of the functional networks even outside the ISC map. To capture whole functional networks, ISC map should facilitate rather than constrain the IC selection.

ICs related to blood-oxygen-level-dependent (BOLD) variations in the fMRI signals can also be identified by classifying the components according to their spatiotemporal properties [21]. Successful classification can separate several types of artifacts from task-related and default-mode [22] ICs. In future fMRI studies, this type of segregation of BOLD-related ICs, followed by the ISC-map-based IC sorting, might improve the detection of task- or stimulus-related components and allow more accurate differentiation between components related to extrinsic and intrinsic processing. ISC as such is a rather conservative estimate of the extrinsic processing circuitry and thus the ISC-based sorting of ICs pinpoints the intrinsic or default-mode components only if they are modulated by the given stimulus.

Although we were able to differentiate functional networks better with ISC–ICA than GLM approach, still one shortcoming of the ISC–ICA method is that the differences seen in the IC time courses do not reflect statistical significance as such. Albeit the strength of the totally data-driven approach to provide information beyond predetermined models, currently it can only complement, but not totally substitute, the traditional model-based fMRI analysis. Below the implications provided by the ISC-ICA about the involvement of the functional networks in the processing of the audiovisual stimuli are discussed in more detail.

4.2 Left-lateralized speech-comprehension network

The left-hemisphere-dominant network comprised nodes at the left posterior STS, left anterior STG, left IFG, right IPL, and left medial BA6. The frontal and posterotemporal nodes of the speech-comprehension network agree with areas that are anatomically connected via the arcuate fasciculus and other dorsal and ventral white matter tracts [23]. An additional pathway running lateral to the classical arcuate fasciculus into IPL has been demonstrated in humans by diffusion tensor imaging [24], and IPL connections to ventral premotor cortex have been traced by autoradiography methods in monkeys [25].

Superior temporal, inferior frontal and premotor areas are known to be activated more during perception of audiovisual speech than of either auditory or visual speech alone [3]. Our results suggest this network, beyond being activated during natural well-hearable audiovisual speech, shows enhanced activation when speech understanding requires additional effort due to decreased voice loudness.

Both data- and model-based analysis methods revealed the crucial involvement of superior temporal plane and IPL [26, 27] in normal speech processing. The involvement of the posterior STS in perception of audiovisual speech is also well known, with increased activation at low signal-to-noise ratio [28–31]. Accordingly, we found, with both ISC–ICA and GLM-based analyses, that the decreased sound level of the audiovisual speech was associated with enhanced activation of the left posterior STS, and that this area overlapped with activation related to observation of normal speech. The manipulation of the sound level of reversed speech or tone pips did not result in any statistically significant differences in brain activations.

Contrary to visual cortices, left posterior STS and right IPL—areas related to audiovisual speech processing and identified with all three analysis methods—the frontal nodes of the network were seen only with ISC and ICA. In an earlier fMRI study by Wilson et al. [17], ISC analysis revealed extensive IFG activations during comprehension of narrative speech, although the cluster sizes did not reach statistical significance in standard model-based analysis. According to the authors, IFG failed to show consistent signal changes during narrative comprehension, whereas some stimulus features resulted in systematic fluctuations. Our observations corroborate these findings: IFG was well visible in both the ISC map and in ICA, but only with very liberal statistical threshold (uncorrected $p = 0.02$) in the GLM

analysis. Altogether, these findings support the involvement of IFG in the processing of natural and continuous speech.

Although IFG is not a direct sensory projection area, its clear appearance (both on left and right) in the ISC map and ICA network indicates systematic, across-subjects dependence on the external audiovisual speech. ICA results further confirmed the involvement of IFG in comprehension of audiovisual speech also in challenging hearing conditions. This conclusion is in line with previous observations of IFG function: IFG is activated, together with superior and middle temporal areas, during perception of distorted speech [2] and during lip reading [32, 33]. Along similar lines, stronger activation of Broca's region within the IFG occurs *e.g.* during perception of incongruent than congruent audiovisual speech [29, 34], in dyslexic subjects *vs.* normal-reading subjects who passively view words [35], and in patients with a cochlear prosthesis listening to their native language relative to normal-hearing subjects [36]. In all these conditions, the involvement of Broca's region increased when the task required increased effort for understanding the sensory input (for a review, see Nishitani et al. 2005).

4.3 Activation for time-reversed speech and the right MTG–IFG–IPL network

The strongest activation in the auditory cortex was seen for loud reversed speech. Attention is known to increase activity of the auditory cortex [37–40], and thus a likely reason was the enhanced attention that the subjects had to pay to these perplexing audiovisual stimuli.

The strong left-hemisphere lateralization of normal speech compared with reversed speech agrees with recent findings of a left-hemisphere network activated by intelligible *vs.* time-reversed speech (auditory word pairs without visual input) [41]. Reversed speech activated more than the normal speech the right-hemisphere MTG–IFG–IPL circuitry.

The right IPL, a known multimodal association area, was a part of both the left-lateralized sensorimotor and of the right-lateralized MTG–IFG–IPL network. Its activation was increased during soft normal speech as well as during reversed speech, *i.e.* during audiovisual stimuli of low or absent intelligibility. This effect could be related to increased effort in word retrieval [42], increased attentional demands [40], or recruitment of working memory (WM) processes.

The synchronous activation of MTG, IFG, and IPL, right-hemisphere lateralized in our study, resembles the auditory WM network, which occupies the middle temporal cortices, superior, middle and inferior frontal gyri, as well as the posterior parietal area [43]. This network showed a peculiar temporal behavior, with transient increases of activity during every block, as though the state of this fronto-temporo-parietal network would have changed each time when the WM was updated (cf. [44]). Similar reactivity to block transitions occurred also in the POS region. Interestingly, POS is activated after eye blinks [45, 46] and saccades [47], two phenomena during which the visual input is briefly interrupted, and thus visual continuity has to be supported internally.

4.4 Default-mode network and silent speech

The posterior cingulate cortex (PCC), medial prefrontal cortex, and lateral parietal cortices are commonly described as a default-mode (DM) network that reflects brain connectivity in the absence of sensory stimulation or tasks [22]. The activity of the DM network decreases during mental tasks [48] and can be modulated by external stimulation [49–51]. Our ISC–ICA analysis picked up a DM network among the ten most stimulus-related ICs, meaning that the audiovisual stimulation modified its function, which is well understandable because the DM is known to be less strongly activated during tasks and various stimuli. In contrast, the GLM analysis failed to reveal any PCC activation in contrasts between silent normal speech and conditions, in which voice was included; instead, activations were observed in other areas of the DM network, *i.e.* the inferior parietal and medial prefrontal cortices, precuneus adjacent to PCC, and ACC. It is possible that the visual stimulation, present in both silent-speech and heard-speech conditions, affected similarly the PCC activation and, thus no difference was seen in subtraction analysis (silent *vs*. heard speech).

5 Conclusions

The combined ISC–ICA analysis allowed us to identify a sensorimotor left-hemisphere network involved in the comprehension of continuous audiovisual speech; the network reacted to both natural and well-hearable speech, and occasionally even stronger for speech of diminished intelligibility. Corresponding GLM analysis revealed only parts of the network and no involvement of motor areas. The ISC–ICA approach thus seems suitable for fMRI experiments involving naturalistic and continuous stimuli.

6 Acknowledgements

This work was financially supported by the Academy of Finland (National Centers of Excellence Programme 2006–2011, and NEURO research grant number 111721), the Ministry of Education Finland via the Finnish Graduate School of Neuroscience, the Sigrid Jusélius Foundation, Jenny and Antti Wihuri Foundation, and ERC Advanced Grant (grant number 232946). We thank Marita Kattelus and Nuutti Vartiainen for expert help in fMRI experiments.

References

[1] Price CJ (2000) The anatomy of language: contributions from functional neuroimaging. *J Anat* 197 Pt 3:335–59.

[2] Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. *J Neurosci* 23:3423–31.

[3] Skipper JI, Nusbaum HC, Small SL (2005) Listening to talking faces: motor cortical activation during speech perception. *NeuroImage* 25:76–89.

[4] Calhoun VD, Adali T, Pearlson GD, Pekar JJ (2001) A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp* 14:140–51.

[5] Malinen S, Hlushchuk Y, Hari R (2007) Towards natural stimulation in fMRI—issues of data analysis. *NeuroImage* 35:131–9.

[6] Calhoun VD, Pekar JJ, McGinty VB, Adali T, Watson TD, Pearlson GD (2002) Different activation dynamics in multiple neural systems during simulated driving. *Hum Brain Mapp* 16:158–67.

[7] Kansaku K, Muraki S, Umeyama S, Nishimori Y, Kochiyama T, Yamane S, Kitazawa S (2005) Cortical activity in multiple motor areas during sequential finger movements: an application of independent component analysis. *NeuroImage* 28:669–81.

[8] McKeown MJ (2000) Detection of consistently task-related activations in fMRI data with hybrid independent component analysis. *NeuroImage* 11:24–35.

[9] McKeown MJ, Makeig S, Brown GG, Jung TP, Kindermann SS, Bell AJ, Sejnowski TJ (1998) Analysis of fMRI data by blind separation into independent spatial components. *Hum Brain Mapp* 6:160–88.

[10] Moritz CH, Carew JD, McMillan AB, Meyerand ME (2005) Independent component analysis applied to self-paced functional MR imaging paradigms. *NeuroImage* 25:181–92.

[11] Ylipaavalniemi J, Savia E, Malinen S, Hari R, Vigario R, Kaski S (2009) Dependencies between stimuli and spatially independent fMRI sources: towards brain correlates of natural stimuli. *NeuroImage* 48:176–85.

[12] Calhoun VD, Maciejewski PK, Pearlson GD, Kiehl KA (2008) Temporal lobe and "default" hemodynamic brain modes discriminate between schizophrenia and bipolar disorder. *Hum Brain Mapp* 29:1265–75.

[13] van de Ven VG, Formisano E, Prvulovic D, Roeder CH, Linden DE (2004) Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. *Hum Brain Mapp* 22:165–78.

[14] Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. *Science* 303:1634–40.

[15] Golland Y, Bentin S, Gelbard H, Benjamini Y, Heller R, Nir Y, Hasson U, Malach R (2007) Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation. *Cereb Cortex* 17:766–77.

[16] Golland Y, Golland P, Bentin S, Malach R (2008) Data-driven clustering reveals a fundamental subdivision of the human cortex into two global systems. *Neuropsychologia* 46:540–53.

[17] Wilson SM, Molnar-Szakacs I, Iacoboni M (2007) Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cereb Cortex*.

[18] Bartels A, Zeki S (2004) The chronoarchitecture of the human brain—natural viewing conditions reveal a time-based anatomy of the brain. *NeuroImage* 22:419–33.

[19] Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113.

[20] Li YO, Adali T, Calhoun VD (2007) Estimating the number of independent components for functional magnetic resonance imaging data. *Hum Brain Mapp* 28:1251–66.
[21] De Martino F, Gentile F, Esposito F, Balsi M, Di Salle F, Goebel R, Formisano E (2007) Classification of fMRI independent components using IC-fingerprints and support vector machine classifiers. *NeuroImage* 34:177–94.
[22] Fox MD, Raichle ME (2007) Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci* 8:700–11.
[23] Friederici AD (2009) Pathways to language: fiber tracts in the human brain. *Trends Cogn Sci* 13:175–81.
[24] Catani M, Jones DK, ffytche DH (2005) Perisylvian language networks of the human brain. *Ann Neurol* 57:8–16.
[25] Petrides M, Pandya DN (2009) Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol* 7:e1000170.
[26] Hickok G, Okada K, Serences JT (2009) Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J Neurophysiol* 101:2725–32.
[27] Iacoboni M, Wilson SM (2006) Beyond a single area: motor control and language within a neural architecture encompassing Broca's area. *Cortex* 42:503–6.
[28] Hein G, Knight RT (2008) Superior temporal sulcus–It's my area: or is it? *J Cogn Neurosci* 20:2125–36.
[29] Ojanen V, Möttönen R, Pekkola J, Jääskeläinen IP, Joensuu R, Autti T, Sams M (2005) Processing of audiovisual speech in Broca's area. *NeuroImage* 25:333–8.
[30] Sekiyama K, Kanno I, Miura S, Sugita Y (2003) Auditory-visual speech perception examined by fMRI and PET. *Neurosci Res* 47:277–87.
[31] Szycik GR, Tausche P, Münte TF (2008) A novel approach to study audiovisual integration in speech perception: localizer fMRI and sparse sampling. *Brain Res* 1220:142–9.
[32] Okada K, Hickok G (2009) Two cortical mechanisms support the integration of visual and auditory speech: a hypothesis and preliminary data. *Neurosci Lett* 452:219–23.
[33] Paulesu E, Perani D, Blasi V, Silani G, Borghese NA, De Giovanni U, Sensolo S, Fazio F (2003) A functional-anatomical model for lipreading. *J Neurophysiol* 90:2005–13.
[34] Szycik GR, Jansma H, Münte TF (2009) Audiovisual integration during speech comprehension: an fMRI study comparing ROI-based and whole brain analyses. *Hum Brain Mapp* 30:1990–9.
[35] Salmelin R, Service E, Kiesilä P, Uutela K, Salonen O (1996) Impaired visual word processing in dyslexia revealed with magnetoencephalography. *Ann Neurol* 40:157–62.
[36] Naito Y, Okazawa H, Honjo I, Takahashi H, Kawano M, Ishizu K, Yonekura Y (1995) Cortical activation during sound stimulation in cochlear implant users demonstrated by positron emission tomography. *Ann Otol Rhinol Laryngol Suppl* 166:60–4.
[37] Grady CL, Van Meter JW, Maisog JM, Pietrini P, Krasuski J, Rauschecker JP (1997) Attention-related modulation of activity in primary and secondary auditory cortex. *Neuroreport* 8:2511–6.
[38] Hari R, Hämaläinen M, Kaukoranta E, Mäkelä J, Joutsiniemi SL, Tiihonen J (1989) Selective listening modifies activity of the human auditory cortex. *Exp Brain Res* 74:463–70.
[39] Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004) Attentional modulation of human auditory cortex. *Nat Neurosci* 7:658–63.
[40] Pugh KR, offywitz BA, Shaywitz SE, Fulbright RK, Byrd D, Skudlarski P, Shankweiler DP, Katz L, Constable RT, Fletcher J, Lacadie C, Marchione K, Gore JC (1996) Auditory selective attention: an fMRI investigation. *NeuroImage* 4:159–73.

[41] Leff AP, Schofield TM, Stephan KE, Crinion JT, Friston KJ, Price CJ (2008) The cortical dynamics of intelligible speech. *J Neurosci* 28:13209–15.
[42] Drager B, Jansen A, Bruchmann S, Forster AF, Pleger B, Zwitserlood P, Knecht S (2004) How does the brain accommodate to increased task difficulty in word finding? A functional MRI study. *NeuroImage* 23:1152–60.
[43] Martinkauppi S, Rämä P, Aronen HJ, Korvenoja A, Carlson S (2000) Working memory of auditory localization. *Cereb Cortex* 10:889–98.
[44] Roth JK, Serences JT, Courtney SM (2006) Neural system for controlling the contents of object working memory in humans. *Cereb Cortex* 16:1595–603.
[45] Bristow D, Frith C, Rees G (2005) Two distinct neural effects of blinking on human visual processing. *NeuroImage* 27:136–45.
[46] Hari R, Salmelin R, Tissari SO, Kajola M, Virsu V (1994) Visual stability during eyeblinks. *Nature* 367:121–2.
[47] Jousmäki V, Hämäläinen M, Hari R (1996) Magnetic source imaging during a visually guided task. *Neuroreport* 7:2961–4.
[48] Esposito F, Bertolino A, Scarabino T, Latorre V, Blasi G, Popolizio T, Tedeschi G, Cirillo S, Goebel R, Di Salle F (2006) Independent component model of the default-mode brain function: Assessing the impact of active thinking. *Brain Res Bull* 70:263–9.
[49] Fox MD, Snyder AZ, Zacks JM, Raichle ME (2006) Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. *Nat Neurosci* 9:23–5.
[50] Greicius MD, Menon V (2004) Default-mode activity during a passive sensory task: uncoupled from deactivation but impacting activation. *J Cogn Neurosci* 16:1484–92.
[51] McKiernan KA, Kaufman JN, Kucera-Thompson J, Binder JR (2003) A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *J Cogn Neurosci* 15:394–408.