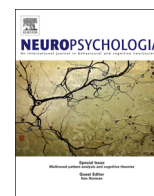




ELSEVIER

Contents lists available at SciVerse ScienceDirect

## Neuropsychologia

journal homepage: [www.elsevier.com/locate/neuropsychologia](http://www.elsevier.com/locate/neuropsychologia)

# The time course of auditory–visual processing of speech and body actions: Evidence for the simultaneous activation of an extended neural network for semantic processing

Georg F Meyer<sup>a,\*</sup>, Neil R Harrison<sup>b</sup>, Sophie M Wuerger<sup>a</sup><sup>a</sup> Department of Psychological Sciences, University of Liverpool, Liverpool L697ZA, UK<sup>b</sup> Department of Psychology, Liverpool Hope University, Hope Park L16 9JD, UK

## ARTICLE INFO

## Article history:

Received 3 December 2012

Received in revised form

16 April 2013

Accepted 20 May 2013

Available online 31 May 2013

## Keywords:

Multisensory

Biological motion

Speech

Semantic processing

EEG

## ABSTRACT

An extensive network of cortical areas is involved in multisensory object and action recognition. This network draws on inferior frontal, posterior temporal, and parietal areas; activity is modulated by familiarity and the semantic congruency of auditory and visual component signals even if semantic incongruences are created by combining visual and auditory signals representing very different signal categories, such as speech and whole body actions. Here we present results from a high-density ERP study designed to examine the time-course and source location of responses to semantically congruent and incongruent audiovisual speech and body actions to explore whether the network involved in action recognition consists of a hierarchy of sequentially activated processing modules or a network of simultaneously active processing sites.

We report two main results:

- 1) There are no significant early differences in the processing of congruent and incongruent audiovisual action sequences. The earliest difference between congruent and incongruent audiovisual stimuli occurs between 240 and 280 ms after stimulus onset in the left temporal region. Between 340 and 420 ms, semantic congruence modulates responses in central and right frontal areas. Late differences (after 460 ms) occur bilaterally in frontal areas.
- 2) Source localisation (dipole modelling and LORETA) reveals that an extended network encompassing inferior frontal, temporal, parasagittal, and superior parietal sites are simultaneously active between 180 and 420 ms to process auditory–visual action sequences. Early activation (before 120 ms) can be explained by activity in mainly sensory cortices.

The simultaneous activation of an extended network between 180 and 420 ms is consistent with models that posit parallel processing of complex action sequences in frontal, temporal and parietal areas rather than models that postulate hierarchical processing in a sequence of brain regions.

© 2013 Elsevier Ltd. All rights reserved.

## 1. Introduction

Information about the environment in which we operate is represented in multiple modalities. The cues that determine how we integrate these sensory representations range from common signal statistics, such as temporal, spatial or structural congruence, to signal semantics. Behavioural data show that temporally, spatially, and semantically congruent information has a facilitatory effect on performance such that congruent bimodal stimuli are

detected and discriminated faster and more accurately than stimuli representing incongruent signals (Meyer, Wuerger, Roehrbein, & Zetzsche, 2005; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Harrison, Wuerger, & Meyer, 2010; Meyer & Wuerger, 2001; Wuerger, Hofbauer, & Meyer, 2003). The facilitatory effect of spatial and temporal congruence can be explained by early neural integration stages that have, for instance, been demonstrated in the superior colliculus of cat (e.g. Meredith & Stein, 1996; Meredith, Nemitz, & Stein, 1987), but a purely signal-statistic-dependent integration of sensory signals, without a representation of stimulus semantics, cannot account for the behavioural consequences of semantic congruency. These semantic effects have been demonstrated behaviourally for biological motion perception (e.g. Brooks,

\* Corresponding author. Tel.: +44 151 794 2579.

E-mail addresses: [georg@liverpool.ac.uk](mailto:georg@liverpool.ac.uk), [georg@liv.ac.uk](mailto:georg@liv.ac.uk) (G. Meyer).

van der Zwan, Billard, Petreska, Clarke & Blanke, 2007; Wuerger, Crocker-Buque, & Meyer, 2012a) as well as in speech perception (Soto-Faraco, Navarra, & Alsius, 2004) in EEG (Teder-Sälejärvi, Di Russo, McDonald, & Hillyard, 2005) and fMRI studies (Werner & Noppeney, 2010; Noppeney, 2009; van Atteveldt, Formisano, Goebel, & Blomert, 2004; Meyer, Greenlee, & Wuerger, 2011; Wuerger et al., 2012a,b).

fMRI studies have consistently identified an extended network of brain areas that contribute to semantic processing (Doehrmann & Naumer, 2008; Meyer et al., 2011; Noppeney, 2009; Noppeney, Ostwald, & Werner, 2010; Phillips, Humphreys, Noppeney, & Price, 2002; Skipper, van Wassenhove, Nusbaum, & Small, 2007; van Wassenhove, Grant, & Poeppel, 2005). Meyer et al. (2011) showed semantically incongruent stimuli, consisting of two meaningful speech and body action signals that were presented simultaneously in different modalities, causes increased activation in a subset of brain areas, including posterior temporal (pSTS), inferior frontal (left IFG) and parasagittal areas (SMA). These incongruity effects were not seen when a meaningful signal was paired with a meaningless, but statistically matched, signal. This suggests that overlapping networks are used to process meaningful speech and body-action signals.

These data support the view that semantic processing of speech (Skipper et al., 2007) and actions (Noppeney, 2009) is based on large-scale dynamic interactions between temporal cortical regions, involved in pattern matching, multisensory integration and higher order semantics, the fronto-parietal action system, and executive regions in prefrontal and inferior frontal regions (Meyer et al., 2011; Noppeney, 2009). Doehrmann and Naumer (2008) argue for a functional differentiation of temporal and frontal cortical regions for semantic processing because temporal areas tend to be more responsive to semantically congruent signals while frontal areas respond preferentially to semantically incongruent audiovisual (AV) stimulation.

The aim of this paper is to explore the temporal dynamics of semantic processing using EEG to identify the timing and broad location of semantic interactions and to differentiate between hierarchical and interactive processing architectures. If processing of the audiovisual signals depends on a hierarchy of anatomically separated processing centres, we expect sequenced activations of different areas, starting out in the sensory cortices. Alternatively, an active perception model, suggested by Skipper et al. (2007) for audiovisual speech perception, predicts simultaneous activity in a network of areas including inferior frontal, superior temporal and parietal areas. ERPs, by virtue of their excellent temporal resolution, enable us to test these competing hypotheses empirically. Furthermore, if a functionally differentiated network of inferior frontal and posterior temporal brain areas is involved in the crossmodal integration of speech and body signals (Doehrmann & Naumer, 2008), we would predict differential ERP responses, with increases magnitude ERPs to congruent stimuli in temporal areas and more sustained activity in (inferior) frontal areas for incongruent than for congruent signals.

Electrophysiological correlates of audiovisual signal presentation that may be linked to semantic processing have been demonstrated previously. The most widely studied electrophysiological signature of semantic processing is the N400 component, which is seen when incongruent semantic signals are compared with congruent stimuli. The N400 component is a monophasic negativity between 200 and 600 ms that is largest over centroparietal sites. In the language domain a lateralisation to the right hemisphere is commonly observed (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980). While the N400 potential was first demonstrated in language tasks, equivalent responses for a wide range of semantic congruency effects have been shown for non-linguistic stimuli, including face processing (Bentin & Deouell,

2000), music perception (Steinbeis & Koelsch, 2008), gesture (Kelly, Kravitz, & Hopkins, 2004), and action perception (Bach, Gunter, Knoblich, Prinz, & Friederici, 2009), which suggests that N400-like components are linked to a much more general contextual semantic integration than purely linguistic processing (Kutas & Federmeier, 2011).

While the N400 potential is well established as a correlate of semantic processing, is it still an open question to what extent *earlier* audiovisual interactions reflect semantic processing or sensory facilitation effects. The earliest audiovisual interactions that cannot be explained by anticipatory slow potentials (Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002) have been linked to the N100 component of auditory evoked potentials (around 100 ms after signal onset). The component is suppressed and speeded up when the auditory signal is accompanied by congruent lip movements (Besle, Fort, Delpuech, & Giard, 2004; Klucharev, Möttönen, & Sams, 2003; van Wassenhove et al., 2005) and the observed cortical deactivation to bimodal speech is associated with behavioural facilitation, that is, faster identification of bimodal syllables than auditory-alone syllables (Besle, Fort, & Giard, 2004; Klucharev et al., 2003).

A second commonly reported time window for audio-visual interaction that may represent semantic integration has been reported around 170–200 ms (the visual P2, or auditory P200 component; Teder-Sälejärvi et al., 2002) and is likely to emanate from cortical areas in the occipito-temporal stream. Giard and Peronnet (1999) explained this effect as an auditory modulatory influence on predominantly visual processing.

Teder-Sälejärvi et al. (2002), consistent with Stekelenburg and Vroomen (2007), report a major audiovisual interaction around 250 ms after signal onset that was accounted for by a dipole pair in anterior temporal peri-sylvian cortex. It is unclear whether the facilitation in these early analysis time windows is due to integration of semantic, speech-specific, information or due to common signal statistics, independent if the stimulus semantics. An example for this sensory integration without recourse to semantic representations is the facilitatory effect of spatial and temporal congruence in very early neural integration stages (Meredith & Stein, 1996; Meredith et al., 1987). Stekelenburg and Vroomen (2007) argue that, for speech signals, the N1 component modulation by congruent visual signals is not contingent on the meaning of the speech signal and demonstrate similar effects for non-speech visual signals, provided that there was visual anticipatory motion. They argue that there is no difference in early AV integration effects between speech and non-speech provided the temporal dynamics of the audiovisual stimuli are matched.

A common method to assess multisensory interactions using event-related potentials is to subtract the sum of the responses to two unimodal signals from the response to a bimodal cue, i.e., AV-(A+V), the so-called 'additive model' (Besle et al., 2004; Besle, Bertrand, & Giard, 2009; Giard & Besle, 2010). This approach closely matches neurophysiological methods used to identify sensory signal interactions (Meredith et al., 1987) but harbours two potential problems for ERP analysis:

(1) It is likely that the (two) unimodal trials on one side of the equation contain 'common' activity (activity related to attention, response selection, manual response etc.), while this is only present in the 'bimodal, AV' side of the AV-(A+V) equation once (Besle et al., 2009). For this reason it is common to restrict the interaction analysis to putatively sensory components prior to approximately 200 ms after stimulus onset (Hillyard, Vogel, & Luck, 1998) although this may not rule out spurious interactions (Teder-Sälejärvi et al., 2002).

(2) anticipatory potentials, also are present twice in the unimodal conditions, so that very early differences between (A+V)

Download English Version:

<https://daneshyari.com/en/article/10464761>

Download Persian Version:

<https://daneshyari.com/article/10464761>

[Daneshyari.com](https://daneshyari.com)