

# TIMING OF AUDIOVISUAL INPUTS TO THE PREFRONTAL CORTEX AND MULTISENSORY INTEGRATION

L. M. ROMANSKI<sup>a,\*</sup> AND J. HWANG<sup>a,b</sup>

<sup>a</sup>Department of Neurobiology & Anatomy, University of Rochester School of Medicine, Rochester, NY, USA

<sup>b</sup>Brain & Cognitive Sciences, University of Rochester, Rochester, NY, USA

**Abstract**—A number of studies have demonstrated that the relative timing of audiovisual stimuli is especially important for multisensory integration of speech signals although the neuronal mechanisms underlying this complex behavior are unknown. Temporal coincidence and congruency are thought to underlie the successful merging of two inter-modal stimuli into a coherent perceptual representation. It has been previously shown that single neurons in the non-human primate prefrontal cortex integrate face and vocalization information. However, these multisensory responses and the degree to which they depend on temporal coincidence have yet to be determined. In this study we analyzed the response latency of ventrolateral prefrontal (VLPFC) neurons to face, vocalization and combined face–vocalization stimuli and an offset (asynchronous) version of the face–vocalization stimulus. Our results indicate that for most prefrontal multisensory neurons, the response latency for the vocalization was the shortest, followed by the combined face–vocalization stimuli. The face stimulus had the longest onset response latency. When tested with a dynamic face–vocalization stimulus that had been temporally offset (asynchronous) one-third of multisensory cells in VLPFC demonstrated a change in response compared to the response to the natural, synchronous face–vocalization movie. Our results indicate that prefrontal neurons are sensitive to the temporal properties of audiovisual stimuli. A disruption in the temporal synchrony of an audiovisual signal which results in a change in the firing of communication related prefrontal neurons could underlie the loss in intelligibility which occurs with asynchronous speech stimuli. © 2012 IBRO. Published by Elsevier Ltd. All rights reserved.

**Key words:** frontal lobe, multisensory integration, vocalizations, faces, temporal coincidence, neurophysiology.

\*Corresponding author. Address: Department of Neurobiology & Anatomy, University of Rochester, Rochester, NY 14642, USA. Tel: +1-585-273-1469; fax: +1-585-756-5334.

E-mail address: [Liz\\_romanski@urmc.rochester.edu](mailto:Liz_romanski@urmc.rochester.edu) (L. M. Romanski).  
*Abbreviations:* A, auditory; AVs, auditory plus static visual image; AVm, audio-visual dynamic movie; Async, asynchronous audio-visual movie; ISI, inter-spike interval; SDF, spike density function; SD, standard deviation; SOA, stimulus onset asynchrony; SI, surprise index; VLPFC, ventrolateral prefrontal cortex; Vm, dynamic visual movie stimulus; Vs, visual–static face.

## INTRODUCTION

Timing is everything – not only in our interactions with the world around us but also during neuronal interactions within the brain. The timing of auditory and visual events is especially important during multisensory integration in which multiple sources of information may converge in the brain (Meredith et al., 1987; Schroeder et al., 2008; Lakatos et al., 2009). External factors, such as the timing in the presentation of an auditory and a visual stimulus, as well as internal factors, such as the synaptic transmission of information by the auditory and visual systems, can both affect the integration of audiovisual information in the cortex. It has been shown that external sensory stimuli with temporal and spatial coincidence (i.e. signals in close temporal and spatial proximity) are more readily fused into a single perceptual unit (Munhall et al., 1996; Bertelson and Radeau, 1981). Temporal coincidence or synchrony, is especially important for the integration of audiovisual speech signals (Dixon and Spitz, 1980; Munhall et al., 1996; de Gelder and Bertelson, 2003; Macaluso et al., 2004). In fact the temporal coincidence of conflicting crossmodal information contributes to the McGurk and ventriloquist effects (McGurk and MacDonald, 1976; Driver, 1996; Bertelson and Aschersleben, 2003) which demonstrate some important principles of audiovisual integration.

The neuronal mechanisms which underlie the integration of audiovisual information are less clear. Physiological recordings in a non-speech model system – the superior colliculus of the cat (Stein and Meredith, 1993), have delineated important features which may underlie sensory integration in a general sense. Stein and Meredith have determined that both stimulus congruence and temporal coincidence of audiovisual stimuli evoke characteristic neuronal changes which are correlated with optimum perception (Stein and Meredith, 1993). Recordings in the superior colliculus (Meredith et al., 1987; Rowland et al., 2007) which examined response latencies to simple auditory and visual stimuli, determined that in most cases, simultaneous presentation of multiple sensory stimuli results in optimal integration of these stimuli. In contrast, auditory and visual events that occur farther away in time (temporally disparate) evoke less optimal responses. Furthermore, it has been suggested that multisensory integration may result in shorter response latencies relative to the unimodal condition (Rowland et al., 2007). Understanding the relative neuronal timing of complex stimuli including faces and vocalizations as they converge on specific brain regions could aid

in understanding the neuronal mechanisms which underlie complex behaviors including communication and recognition which involve integration.

It has been previously shown that information regarding vocalizations and facial gestures combines to affect neuronal processing in single neurons of the primate auditory cortex, superior temporal sulcus and ventrolateral prefrontal cortex (Barraclough et al., 2005; Ghazanfar et al., 2005; Sugihara et al., 2006; Kayser et al., 2010). In the ventrolateral prefrontal cortex (VLPFC) studies have shown that neurons respond to species-specific vocalizations (Romanski et al., 2005) and to faces (O'Scalaidhe et al., 1999). Moreover, it has been shown that single neurons in the macaque VLPFC integrate species-specific faces with their corresponding vocalizations (Sugihara et al., 2006). Non-linear multisensory neurons exhibit enhancement or suppression to combined presentations of faces and corresponding vocalizations compared with their response to unimodal presentations (Sugihara et al., 2006).

Prefrontal neurons which integrate complex, socially relevant events such as faces and their corresponding vocalizations may be sensitive to large changes in the temporal synchrony, or simultaneous presentation, of a facial gesture and the corresponding vocalization. Stimuli which are not perceived as coincident or close in time may be perceived as being asynchronous and may not be integrated. For example it has been shown that increasing the naturally occurring stimulus onset asynchrony (SOA) between a speech sound and the accompanying visual mouth movement, can result in a loss of intelligibility of speech or a loss in the integration of sounds and visual stimuli (Massaro et al., 1996; van Atteveldt et al., 2007). It is possible that this lack of intelligibility and integration is due to a decrease in neuronal activity in areas which are essential in communication including VLPFC. For this reason we examined the response latency of prefrontal neurons to unimodal and multisensory stimuli and also asked whether prefrontal neurons were sensitive to changes in the synchronous onset of dynamic vocalization stimuli.

While in our previous study we have documented multisensory responsive neurons and the types of multisensory interactions which occur in ventral prefrontal cortex, in the current study, we have analyzed and documented the response timing of these neurons. Furthermore, we hypothesized that brain regions such as VLPFC which are involved in the perception and integration of audiovisual communication would have neurons which are sensitive to the synchrony of audiovisual stimuli. Our results show that in a proportion of VLPFC neurons, asynchronous audiovisual face–vocalization stimuli suppress multisensory responses.

## EXPERIMENTAL PROCEDURES

### Subjects and surgical methods

In the following study we have re-analyzed 487 cells from Sugihara et al. (2006) and have added 282 new cells ( $n = 769$  total cells), from the same three adult rhesus monkeys (*Macaca mulatta*), 1

female (6.7 kg) and 2 males (8.0 and 12.0 kg) previously recorded. All methods were in accordance with National Institutes of Health standards and all experimental protocols were approved by the University of Rochester Care and Use of Animals in Research committee. Recording cylinders were placed over VLPFC to maximize recordings in areas 12/47 and 45 (Preuss and Goldman-Rakic, 1991; Petrides and Pandya, 2002). Recordings were made in the left hemisphere of 2 animals and the right hemisphere of one of the animals.

### Apparatus and general recording procedure

All training and recording was performed in a sound-attenuated room lined with Sonex (Acoustical Solutions). Auditory stimuli (65–75 dB SPL measured at the level of the subject's ear) were presented by either a pair of Audix PH5-vs speakers (frequency response  $\pm 3$  dB, 75–20,000 Hz) located on either side of a center computer monitor, or a centrally located Yamaha MSP5 monitor speaker (50–40,000 Hz), located 76 cm from the monkey's head and placed just below the computer monitor.

Each day the animal subjects were brought to the laboratory where they performed perceptual and memory tasks for juice reward. In the present study animals performed a fixation task where auditory, visual or audiovisual stimuli were presented while animals maintained fixation. The trial began with a 500 ms pretrial fixation period where only the fixation point was visible on the screen followed by the stimulus presentation (which lasted 500–1100 ms), and then a 500 ms post-stimulus fixation period. A juice reward was delivered at the termination of the post-stimulus period, and the fixation requirement was released concurrently with the juice reward. A 2–3 s inter-trial interval preceded the start of the next trial which began by presenting the fixation point. When the subject voluntarily fixated the central point the trial would commence. Since we used fixation as a measure of attention, breaking fixation during the fixation or stimulus periods resulted in an aborted trial. The fixation window was 2–4 degrees during the pre-stimulus fixation period and enlarged during the stimulus period to the same size as the visual stimuli which subtended 7–10 degrees. Eye position was continuously monitored using an ISCAN infrared pupil monitoring system (ISCAN, Inc., Woburn, MA).

During recordings a parylene-coated Tungsten microelectrode (0.8–2.0 M $\Omega$  at 1 kHz) was lowered into the target region by a hydraulic microdrive (Narishige MO-95C), which fit over the recording cylinder. The neuronal activity was amplified (BAK MD-4 amplifier), filtered (Krohn-Hite, 3500, Avon, MA), discriminated (BAK DIS-I Window Discriminator) and displayed on an oscilloscope. Discriminated spikes were digitized and saved online. Simultaneous isolation of two units was possible with dual window discriminators. The timing of the behavioral contingencies, acquisition and storage of TTL spike data, presentation of all stimuli, delivery of reward, and monitoring of eye position were controlled by a computer running CORTEX (NIH derived software, dual-computer mode).

### Stimuli and testing

As described previously (Sugihara et al., 2006) the library of audiovisual stimuli consisted of the audio and video portions of short vocalization movies which were made from recordings of familiar conspecifics in our home colony or from familiar and unfamiliar humans. All stimuli were generated from digitally recorded movies and were processed using Adobe Premier (Adobe Systems, Inc.) and Jasc Animation studio (Jasc Software, Inc.), as well as several custom and freeware programs. Audio and visual components of recorded movies were divided into .wav and .mpeg streams for filtering. The auditory stimuli were filtered and edited using MATLAB (Mathworks, Inc.) and SIGNAL (Engineering Design, Cambridge, MA). Static visual images of monkey faces were created from single frames of

Download English Version:

<https://daneshyari.com/en/article/6275706>

Download Persian Version:

<https://daneshyari.com/article/6275706>

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