



A neurocomputational analysis of the sound-induced flash illusion



Cristiano Cuppini^a, Elisa Magosso^a, Nadia Bolognini^{b,c}, Giuseppe Vallar^{b,c}, Mauro Ursino^{a,*}

^a Department of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy

^b Department of Psychology, University of Milano-Bicocca, Milan, Italy

^c IRCCS Istituto Auxologico Italiano, Milan, Italy

ARTICLE INFO

Article history:

Accepted 1 February 2014

Available online 9 February 2014

Keywords:

Neural network modeling
Multisensory integration
Visual–auditory illusions
Sound-induced flash illusion
Spatial processing
Temporal processing

ABSTRACT

Perception of the external world is based on the integration of inputs from different sensory modalities. Recent experimental findings suggest that this phenomenon is present in lower-level cortical areas at early processing stages. The mechanisms underlying these early processes and the organization of the underlying circuitries are still a matter of debate. Here, we investigate audiovisual interactions by means of a simple neural network consisting of two layers of visual and auditory neurons. We suggest that the spatial and temporal aspects of audio-visual illusions can be explained within this simple framework, based on two main assumptions: auditory and visual neurons communicate via excitatory synapses; and spatio-temporal receptive fields are different in the two modalities, auditory processing exhibiting a higher temporal resolution, while visual processing a higher spatial acuity. With these assumptions, the model is able: i) to simulate the sound-induced flash fission illusion; ii) to reproduce psychometric curves assuming a random variability in some parameters; iii) to account for other audio-visual illusions, such as the sound-induced flash fusion and the ventriloquism illusions; and iv) to predict that visual and auditory stimuli are combined optimally in multisensory integration. In sum, the proposed model provides a unifying summary of spatio-temporal audio-visual interactions, being able to both account for a wide set of empirical findings, and be a framework for future experiments. In perspective, it may be used to understand the neural basis of Bayesian audio-visual inference.

© 2014 Elsevier Inc. All rights reserved.

Introduction

The ability of the brain to integrate inputs from different sensory channels is fundamental for the perception of the external world, since our sensory experience is typically characterized by inputs from multiple sensory modalities (Stein and Meredith, 1993). Multisensory integration, under many circumstances, improves human performance, for example by reducing perceptual ambiguity and response latencies (e.g., Corneil and Munoz, 1996; Hughes et al., 1994; Maravita et al., 2008; Schröger and Widmann, 1998), as well as enhancing stimulus detection (Bolognini et al., 2005; Frassinetti et al., 2002; Stein et al., 1996, for reviews see also Alais et al., 2010; Shams and Kim, 2010; Stein, 1998). Crossmodal interactions occur at various cortical levels of the two cerebral hemispheres: they have been ascribed to feedback projections from multimodal brain areas in the parietal or temporal lobes (Buchel et al., 1998; Macaluso et al., 2000; McDonald et al., 2003, 2005), to subcortical nuclei, including the superior colliculus (Meredith, 2002; Meredith and Stein, 1986; Wallace and Stein, 2007),

and to direct lateral connections between early sensory cortices (see Alais et al., 2010; Driver and Noesselt, 2008; Foxe and Schroeder, 2005; Shams and Kim, 2010; Stein and Stanford, 2008, for reviews).

A main area of research on multisensory integration has focused on the interactions between the visual and the auditory modalities. Perceptual illusions have been used in order to elucidate the mechanisms underlying multisensory integration (see for a review: Recanzone, 2009). One such well-known phenomenon is the “ventriloquism” illusion where an auditory stimulus is “captured” by a simultaneous, but spatially disparate, visual signal (Howard and Templeton, 1966). A number of factors affect this illusion, including the timing of the two stimuli, and their spatial disparity (Alais and Burr, 2004; Bertelson and Aschersleben, 1998; Bertelson and Radeau, 1981; Bolognini et al., 2007; Thurlow and Jack, 1973; Vroomen and de Gelder, 2004a; Welch, 1999; Welch and Warren, 1980). According to the “modality appropriateness hypothesis” (Welch and Warren, 1980) when two sensory modalities deliver conflicting information, perception is dominated by the modality with the greater acuity for the task to be performed. Hence, in the case of spatial discrimination, the visual system that specifies location more precisely than audition prevails producing the ventriloquism effect (Welch, 1999; Welch and Warren, 1980). This is also in line with the idea that sensory uncertainty determines the perceptual weight allocated to a given cue during multisensory integration (e.g., Alais and Burr, 2004; Battaglia et al., 2003; Ernst and Banks, 2002; Ernst et al., 2000; Wallace et al., 2004b).

* Corresponding author at: Department of Electrical, Electronic and Information Engineering (DEI), University of Bologna, Viale Risorgimento 2, I40136 Bologna, Italy. Fax: +39 051 2093073.

E-mail addresses: cristiano.cuppini@unibo.it (C. Cuppini), elisa.magosso@unibo.it (E. Magosso), nadia.bolognini@unimib.it (N. Bolognini), giuseppe.vallar@unimib.it (G. Vallar), mauro.ursino@unibo.it (M. Ursino).

In line with this hypothesis, the prediction can be made that in the temporal domain the auditory system dominates perception, due to shorter neural latencies in the primary auditory cortex (e.g. Recanzone et al., 2000), compared with those in the primary visual cortex (Maunsell and Gibson, 1992): indeed, in the brain auditory stimuli are processed faster than visual stimuli, resulting in a greater temporal acuity of the auditory system, as compared to the visual system.

A number of studies have revealed that the visual modality can be strongly influenced by audition in the temporal domain (Gebhard and Mowbray, 1959; Knox, 1945; Morein-Zamir et al., 2003; Ogilvie, 1956; Sekuler et al., 1997; Stein et al., 1996; Welch et al., 1986). One such example is a cross-modal effect, known as the “temporal ventriloquism”, where a sound, presented in close temporal proximity to a visual stimulus, can alter the perceived timing of the visual onset (Freeman and Driver, 2008; Getzmann, 2007; Parise and Spence, 2008; Vroomen and de Gelder, 2004b). Similarly, the participants' ability to perceive the temporal rate of visual stimuli is affected by auditory stimuli paired to the visual ones (Recanzone, 2003; Shipley, 1964).

Another example of audition dominating vision is an auditory–visual illusion, reported among others by Shams et al. (2000) (but see also Andersen et al., 2004; Innes-Brown and Crewther, 2009; Shams and Kim, 2010), namely, the “sound-induced flash illusion”: when a single flash is presented along with two or more beeps, observers often report seeing two or more flashes, the so-called *fission* illusion. A corresponding *fusion* illusion has also been described, where a single beep causes the fusion of a double flash stimulus (Andersen et al., 2004; Bolognini et al., 2010; Innes-Brown and Crewther, 2009; Shams and Kim, 2010). Therefore, in the sound-induced flash illusions, auditory information determines the phenomenal experience of the number of seen visual stimuli. These functional effects must be implemented by physical connections between different sensory and associative brain regions. Several psychophysical and neuroimaging results show that the auditory-induced alteration of visual perception is associated with a modulation of neural activity in the visual cortex (Shams et al., 2001; Watkins et al., 2006): when a single flash is perceived incorrectly as two flashes, activity in retinotopic V1 is increased; conversely, V1 activity is decreased when a double flash is perceived incorrectly as a single flash (Watkins et al., 2006, 2007). Hence, activity in V1 appears to reflect the participant's subjective perception, rather than the physically presented visual stimulus (Watkins et al., 2006, 2007).

More specifically, visual evoked potentials (VEPs) are modulated by sound in the illusion trials, suggesting that these cross-modal effects occur at the level of the occipital cortex (Shams et al., 2001). Importantly, VEPs associated with illusory flashes are qualitatively very similar to those associated with physical flashes, indicating that similar brain mechanisms underlie the two percepts generated in the visual cortex. In a subsequent fMRI study, capitalizing on the fact that not in all trials the illusory effect occurs, Watkins et al. (2006) found an increased neural activity in the primary visual cortex (V1) when participants experienced the illusion (namely, they reported two flashes), compared to when the illusory perception was absent (namely, they reported one flash), in response to the illusion-inducing audio-visual stimulus (i.e., two beeps, one flash); other areas of increased activity included the posterior segment of the superior temporal sulcus, close to the supramarginal gyrus, and the superior colliculus.

An MEG study investigating the time course of the cross-modal interactions involved in the sound-induced flash illusion (illusion trials vs. visual-alone and auditory-alone trials) showed an early cross-modal effect (35–65 ms from the onset of the visual stimulus) in occipital and parietal scalp locations, as well as at later time intervals (≈ 150 ms post-stimulus), in the occipital, parietal and anterior regions; the early occipital cross-modal effect suggests that the sound-induced flash illusion is mainly based on a direct projection from the primary auditory to the primary visual cortex, rather than on feed-back projections from higher-order associative areas, which may, nevertheless, exert some modulatory role (Shams et al., 2005a).

The conclusion that multisensory interactions occur early in time post-stimulus presentation is in line with experiments using different cross-modal paradigms. For example, Giard and Peronnet (1999) found a modulation of activity in the primary visual cortical areas by sound during a pattern recognition task, with an onset of 40–50 ms from the stimulus presentation. Molholm et al. (2002), in a reaction-time task to audio-visual stimuli, compared to unimodal stimuli, recorded high-density event-related potentials (ERPs) indicating cross-modal interactions in parieto-occipital regions, with an onset as early as 46 ms, essentially simultaneous with the onset of visual cortical processing (for a review, see Driver and Noesselt, 2008). As response latencies of higher-level association cortices are expected to be slower, the observed early responses are unlikely to be based on feedback loops involving non-primary sensory regions (e.g., Schroeder and Foxe, 2005).

Taken together, these sources of evidence support models of brain organization in which multisensory convergence may occur during early perceptual processing, rather than being based on parallel and independent unisensory streams, converging only later in higher-level associative cortical areas (Foxe et al., 2000; for a review, see Driver and Noesselt, 2008; Foxe and Schroeder, 2005; Ghazanfar and Schroeder, 2006; Schroeder and Foxe, 2005). Anatomical studies in nonhuman primates support this view: direct projections have been identified from the primary and association auditory cortices to sub-regions of V1 and V2, where the periphery of the visual field is represented (Falchier et al., 2002; Rockland and Ojima, 2003). Such connections may directly link sensory-specific cortices without involvement of intervening multisensory regions (see also Cappe and Barone, 2005). Moreover, recent findings in humans based on diffusion tensor imaging suggest that fiber tracts from Heschl's gyrus terminate in the occipital pole where the (para)foveal visual field would be represented (Beer et al., 2011), while fMRI evidence has revealed that the resting state activity in the visual cortex of humans correlates with the activity in auditory cortex, but not with activity in parietal or other temporal brain areas (Eckert et al., 2008). On the other hand, data from the macaque suggest fewer direct connections between sensory-specific cortices than feedback connections from conventional multisensory areas, such as the Superior Temporal Sulcus (STS, see Falchier et al., 2002). Therefore, the feedback influences on sensory-specific cortex from multisensory convergence zones may play a major role in modulating crossmodal interactions in primary sensory areas. To exemplify, a fMRI study in humans shows that audio-visual correspondence in temporal pattern may induce feedback influences from multisensory STS upon primary visual and auditory areas (Noesselt et al., 2007).

Our understanding of the mechanisms subtending visual–auditory illusions can greatly benefit from the use of neural network models. Such models allow the mass of behavioral and neural data accumulated on visual–auditory interactions to be summarized into a coherent theoretical structure and the underlying neural circuits to be analyzed in detail (Craik, 1945). Under this rationale, we have recently realized an original model (Magosso et al., 2012) devoted to the simulation and exploration of the ventriloquism effect in the spatial domain. The model is able to emulate a number of physiological and psychophysical data on visual–auditory spatial interactions. Importantly, the model is focused merely on the spatial properties of the visual and auditory stimuli, without considering their temporal characteristics, being therefore not suitable to emulating the temporal aspects of visual–auditory interactions. The present work aims at extending the model of Magosso et al. (2012), by including the temporal patterns of the visual and auditory inputs, in order to overcome its previous limitations: the extended model should be able to simulate not only the spatial ventriloquism illusion, but also other illusions in the temporal domain, such as the sound-induced flash illusion (Shams et al., 2000).

Here we aim at demonstrating that both spatial and temporal aspects of different audio-visual illusions, namely the sound-induced *fission* and *fusion* illusions, as well as the *ventriloquism* illusion, can be explained within a single theoretical framework, based on two main

Download English Version:

<https://daneshyari.com/en/article/6027604>

Download Persian Version:

<https://daneshyari.com/article/6027604>

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