cortex 49 (2013) 2875–2887

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research report

Sight and sound out of synch: Fragmentation and renormalisation of audiovisual integration and subjective timing☆

Elliot D. Freeman [a,*], *Alberta Ipser* [a], *Austra Palmbaha* [a], *Diana Paunoiu* [a], *Peter Brown* [b], *Christian Lambert* [c], *Alex Leff* [d] *and Jon Driver* [d]

[a] *City University London, United Kingdom*
[b] *John Radcliffe Hospital, Oxford University, United Kingdom*
[c] *St George's, University of London, United Kingdom*
[d] *Institute of Cognitive Neuroscience, University College London, United Kingdom*

ARTICLE INFO

Article history:
Received 19 November 2012
Reviewed 21 December 2012
Revised 10 February 2013
Accepted 19 March 2013
Action editor Robert McIntosh
Published online 1 April 2013

Keywords:
Audiovisual integration
Psychophysics
Individual differences
Illusions
Sensory timing

ABSTRACT

The sight and sound of a person speaking or a ball bouncing may seem simultaneous, but their corresponding neural signals are spread out over time as they arrive at different multisensory brain sites. How subjective timing relates to such neural timing remains a fundamental neuroscientific and philosophical puzzle. A dominant assumption is that temporal coherence is achieved by sensory resynchronisation or recalibration across asynchronous brain events. This assumption is easily confirmed by estimating subjective audiovisual timing for groups of subjects, which is on average similar across different measures and stimuli, and approximately veridical. But few studies have examined normal and pathological individual differences in such measures.

Case PH, with lesions in pons and basal ganglia, hears people speak before seeing their lips move. Temporal order judgements (TOJs) confirmed this: voices had to *lag* lip-movements (by ~200 msec) to seem synchronous to PH. Curiously, voices had to *lead* lips (also by ~200 msec) to maximise the McGurk illusion (a measure of audiovisual speech integration). On average across these measures, PH's timing was therefore still veridical. Age-matched control participants showed similar discrepancies. Indeed, normal individual differences in TOJ and McGurk timing correlated *negatively*: subjects needing an auditory lag for subjective simultaneity needed an auditory lead for maximal McGurk, and vice versa. This generalised to the Stream–Bounce illusion. Such surprising antagonism seems opposed to good sensory resynchronisation, yet average timing across tasks was still near-veridical.

Our findings reveal remarkable disunity of audiovisual timing within and between subjects. To explain this we propose that the timing of audiovisual signals within different brain mechanisms is perceived relative to the average timing across mechanisms. Such renormalisation fully explains the curious antagonistic relationship between disparate timing estimates in PH and healthy participants, and how they can still perceive the timing of external events correctly, on average.

* *Corresponding author*. City University, Northampton Square, London EC1V 0HB, United Kingdom.
E-mail address: elliot.freeman@city.ac.uk (E.D. Freeman).
0010-9452/$ – see front matter © 2014 The Authors. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cortex.2013.03.006

1. Introduction

When a person speaks, we usually expect to hear their voice at the same time as seeing their lips move. Furthermore, if we watch their lips, it often helps us to hear their voice better, via 'speechreading' (Sumby and Pollack, 1954). Two distinct kinds of processes are implied by such observations: synchronisation and integration. Firstly, we are sensitive to when auditory and visual events are occurring at the same time (Alais and Carlile, 2005; King, 2005; Kopinska and Harris, 2004; Sugita and Suzuki, 2003). Secondly, the ability to benefit from the combination of modalities, as in speechreading, requires that auditory and visual information be brought together in the brain and integrated. So automatic and compelling is such integration that artificial mismatches between sound and vision can easily induce illusory changes in the perceived location, timing or actual interpretation of the stimuli (Bertelson and Radeau, 1981; Howard and Templeton, 1966; McGurk and MacDonald, 1976; Witkin et al., 1952).

It is easy to take for granted that audiovisual events are always synchronised and integrated correctly. But here, we present the first ever confirmed case of a patient (PH) who hears peoples' voices *before* he sees their lips move. Testing this individual in comparison with neurologically healthy participants gave us the unique opportunity to address two issues: Firstly, we ask whether PH's auditory leading phenomenon is selective for subjective synchrony or whether his audiovisual integration is also affected. This addresses a current debate over whether optimal integration depends on achieving *subjective* synchrony, or whether integration obeys independent temporal constraints (Arnold et al., 2005; Martin et al., 2013; Munhall et al., 1996; Soto-Faraco and Alsius, 2007, 2009; van Wassenhove et al., 2007). Secondly, PH's pathological desynchronisation might provide insight into the deeper question of how (or indeed whether) sensory synchronisation is normally achieved, which has long perplexed neuroscientists and philosophers (Dennett and Kinsbourne, 1995; Harris et al., 2008; Keetels and Vroomen, 2012; Spence and Squire, 2003; Vroomen and Keetels, 2010; Zeki and Bartels, 1998). We consider this issue first.

1.1. *Multisensory synchronisation*

The problem of synchronisation is exemplified by the maxim known as Segal's law: '*With one clock you always know the time; with two you are never sure*'. Does the brain also have multiple clocks, and if so, does this create a similar uncertainty? There are many multimodal convergence zones in the brain (Bushara et al., 2001; Cappe et al., 2009; Driver and Noesselt, 2008; Ghazanfar and Schroeder, 2006; Macaluso and Driver, 2005; Meredith et al., 1987; Stevenson et al., 2010), and to get there, auditory and visual signals must traverse different routes and distances, thus most likely arriving at different times (Arnold et al., 2001; Aschersleben and Prinz, 1995; Halliday and Mingay, 1964; Moutoussis and Zeki, 1997; Stone et al., 2001). Consequently each area will have different information about when visual and auditory events occurred (Scharnowski et al., 2013). This entails a 'multiple-clocks' uncertainty for knowing the absolute and relative timing of external events.

Despite such systemic and intrinsic asynchrony, subjects still often recognise when auditory and visual sources are approximately synchronous (Harris et al., 2008), at least for proximal if not always for distal stimuli (Alais and Carlile, 2005; Arnold et al., 2005; Heron et al., 2007; King, 2005; Kopinska and Harris, 2004; Stone et al., 2001; Sugita and Suzuki, 2003; Vroomen and Keetels, 2010). Shifts in subjective simultaneity following adaptation to asynchrony are consistent with the existence of mechanisms functioning at least locally to resynchronise temporal discrepancies between modalities (Fujisaki et al., 2004; Hanson et al., 2008; Miyazaki et al., 2006; Yamamoto et al., 2012). However, individuals differ widely with respect to the objective audiovisual asynchrony which they perceive as subjectively synchronous (the Point of Subjective Simultaneity – PSS; Stone et al., 2001). This may depend intrinsically on the time for neural conduction and processing of signals, which may differ between stimuli and individuals (Arnold et al., 2001; Aschersleben and Prinz, 1995; Halliday and Mingay, 1964; Moutoussis and Zeki, 1997; Stone et al., 2001), though attentional biases may also account for some apparent individual differences in multisensory timing (Spence and Parise, 2010; Spence et al., 2001). Furthermore, even within the same subjects given the same stimuli, different tasks produce uncorrelated estimates of PSS (van Eijk et al., 2008) though such variations may depend on strategic variables (García-Pérez and Alcalá-Quintana, 2012; Schneider and Bavelier, 2003; van Eijk et al., 2008). Thus synchronising mechanisms, if they exist (Zeki and Bartels, 1998), may not function perfectly.

If there were a single specialised mechanism for multisensory synchronisation, one might expect to find individuals for whom different modalities have been chronically desynchronised following a brain trauma. Loss of acuity for temporal order has been observed following temporal lobectomy (Sherwin and Efron, 1980), but the lack of selective impairments in temporal processing is inconsistent with the notion of a unitary specialised mechanism underlying timing perception (Wiener et al., 2011). Indeed, there is only one previously reported case of apparently acquired sensory desynchronisation (Hamilton et al., 2006). Hamilton et al. (2006) described patient AWF who claimed to experience '*a perceived temporal mismatch*' (Abstract). However they did not specify whether vision actually preceded or lagged audition, and did not formally quantify the temporal mismatch using objective measures, for example by measuring performance across a range of audiovisual asynchronies. Thus to date, evidence that sensory synchronisation can be pathologically impaired rests largely on AWF's subjective report, which is not very specific.

1.2. *Dependence of integration on synchronisation*

While investigations of synchronisation have typically focused on temporal relationships between modalities (e.g., Harris et al., 2008), the multiple-clocks problem also logically applies more generally between different processes. Here we consider two such notional processes, supporting subjective temporal judgements versus those that serve to integrate inputs from different modalities. We ask whether sound and vision are optimally integrated when they are subjectively