Freeman, E. D. & Ipser, A. (2016). Individual differences in multisensory integration and timing. Electronic Imaging, 2016(16), pp. 1-4. doi: 10.2352/ISSN.2470-1173.2016.16HVEI-097

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-097

Individual differences in multisensory integration and timing

Elliot D. Freeman; Alberta Ipser; City University London; London, United Kingdom.

Abstract

The senses have traditionally been studied separately, but it is now recognised that the brain is just as richly multisensory as is our natural environment. This creates fresh challenges for understanding how complex multisensory information is organised and coordinated around the brain. Take timing for example: the sight and sound of a person speaking or a ball bouncing may seem simultaneous, but their neural signals from each modality arrive at different multisensory areas in the brain at different times. How do we nevertheless perceive the synchrony of the original events correctly? It is popularly assumed that this is achieved via some mechanism of multisensory temporal recalibration. But recent work from my lab on normal and pathological individual differences show that sight and sound are nevertheless markedly out of synch by different amounts for each individual and even for different tasks performed by the same individual. Indeed, the more an individual perceive the same multisensory event as having an auditory lead and an auditory lag at the same time. This evidence of apparent temporal disunity sheds new light on the deep problem of understanding how neural timing relates to perceptual timing of multisensory events. It also leads to concrete therapeutic applications: for example, we may now be able to improve an individual's speech comprehension by simply delaying sound or vision to compensate for their individual perceptual asynchrony.

Introduction

The story of this research begins at the very birth of Experimental Psychology in the late 18th century. At that time, ships navigating the British Empire relied on accurate astronomical observations from the Royal Observatory in order to stay on course. Such observations were made by marking time points in the transit of stars and planets relative to the sound of a ticking clock. But there was a problem: different astronomers produced consistently different readings, as if sight and sound were out of synch by different amounts in different observers [1].

These individual differences were considered by the astronomers to be a nuisance variable. They developed clever objective methods to quantify them, only so that they could assign each observer a 'Personal Equation' that could be used to adjust their observations and thus achieve agreement between observers. Modern experimental psychology has inherited both the methodological innovations and this tradition of discounting individual differences, which is usually achieved by averaging. However, revisiting the study of Personal Equations opens up some exciting opportunities, for both theoretical and applied advances that would not be achievable using traditional group-mean based methods.

Sight and sound out of synch: Case PH

The present research began with the chance discovery of a patient for whom sight and sound had suddenly become desynchronised after a minor stroke. Case PH is a retired pilot, aged 68 at time of testing. He presented with the complaint that his vision was delayed relative to his hearing: he could literally hear people speak before seeing their lips move [2]. He reported first noticing the visual delay when watching television, suspecting that the show was poorly dubbed. He asked his daughter to confirm this, but was then surprised to find that she was also poorly dubbed!

In all the standard neuropsychological measures PH was high functioning. He suffers tinnitus, possibly from his piloting days, and myasthenia gravis, though the significance of these are unclear. High resolution MRI identified two lesions in subthalamic nucleus and pontine nucleus. The first, as part of basal ganglia, may be implicated in timing, while the latter may be part of the early auditory system.

To verify PH's remarkable claim, we played videos of a person speaking simple syllables, using a range of different different lip-voice asynchronies. For each video PH was asked to decide whether the voice occurred before or after the lip-movements (i.e. a 'temporal order judgement' task, TOJ; Figure 1a). Concurrently (i.e. in a dual task, based on [3]), he also performed a phoneme-identification task ('did you hear /ba/ or /da/'?). This task was based on the famous McGurk illusion, in which mismatching lip-movements change the phoneme that is heard [4].

PH's performance in the first TOJ task showed that voices had to *lag* lip-movements (by ~200ms) in order to offset his own acquired visual delay, so that they were subjectively synchronous to him. This corroborates PH's subjective report that for naturally simultaneous stimuli, vision lags his audition. However we were surprised to find that in the McGurk task, voices now had to *lead* lips (also by ~200ms) to maximise PH's McGurk illusion (Figure 3a, white circle). Thus PH's lesions caused opposite shifts in perceptual timing for different tasks. Note however that on average across these oppositely-biased measures his timing was still veridical.

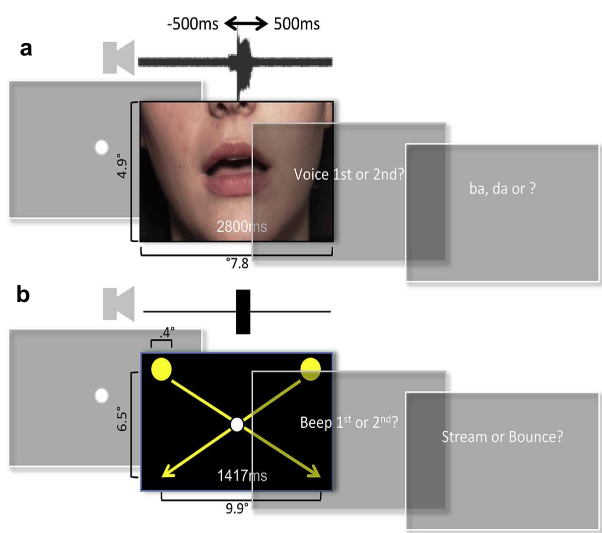

Figure 1. Trial sequence and stimuli for (a) McGurk and (b) Stream-Bounce illusions.

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-097

Audiovisual asynchrony in healthy individuals

Comparison of PH with healthy control participants yielded further surprises. Individual differences were very large, and some individuals performed almost as deviantly as PH. Also like PH, some people needed an auditory *lag* for subjective simultaneity but an auditory *lead* for maximal McGurk; other people showed the opposite pattern (Figure 3a). Consequently, individual differences in TOJ and McGurk tasks correlated with each other *negatively*. And also like PH, average perceptual timing across tasks and participants was close to veridical. A similar significant negative correlation was observed between two analogous measures of the Stream-Bounce illusion in which the apparent trajectory of two colliding balls is altered in the context of an appropriately-timed sound (Figure 1b and Figure 3b). Thus this phenomenon is therefore general to speech and non-speech, and for both directions of influence between modalities.

To confirm that this negative correlation is not a response bias artefact resulting from the use of a dual-task, we have since also successfully replicated it using measures obtained on separate sessions rather than concurrently in a dual task (Freeman, Ipser et al, in prep).

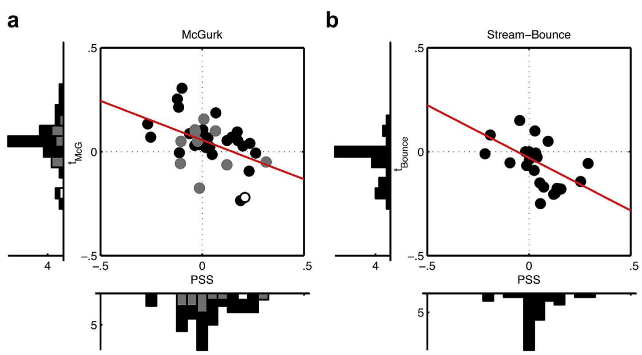

Figure 3. Point of Subjective Simultaneity for the TOJ task (x-axis, sec) plotted against asynchrony for maximum a) McGurk (open circle: PH; grey: older; black: young) and b) bounce illusions.

Temporal Renormalisation Theory

This apparently antagonistic relationship between different measures of perceptual timing seems opposed to good sensory resynchronisation, however note that timing across tasks was still near-veridical when averaged across tasks. It also runs against dominant theories which assume that temporal discrepancies between different brain networks can be minimised or discounted, thus preserving unity of perceptual timing across multisensory events [5]–[9]. Such theories would predict that measures of perceptual timing should correlate positively across different tasks, if at all, but never negatively.

We have proposed an alternative theory relating neural to perceptual timing. We assume that discrepancies of timing within different neural sub-networks are not minimised, but *normalized* relative to the average timing across the whole network [2]. For example, if vision is particularly delayed in one sub-network, then the more vision will seem to lead audition in others, because it is perceived *relative* to the biased average of audiovisual asynchronies across the whole active neural network. Such *Temporal Renormalization* fully explains the curious antagonistic relationship between disparate timing estimates from different tasks, and how despite this apparent disunity of timing, individuals can still perceive the timing of external events correctly, at least on average.

Asynchrony and Dyslexia

What significance do individual differences in perceptual timing have for higher cognitive abilities? Our ongoing research has been examining how this may impact on reading and also speech comprehension (next section). We hypothesise that poor audiovisual synchronization during development might impair accuracy of audiovisual speech interpretation [10], and thus delay the process of learning to disambiguate unfamiliar phonemes by integrating visual lip-movement cues. This might subsequently impact on learning to relate phonemes to written graphemes. To test this, we compared normal adult readers with adults who were independently diagnosed with dyslexia. In support of our hypothesis we found that individual asynchronies for integrating audiovisual speech correlated significantly with measures of reading ability (Figure 2). In the dyslexics only, better reading ability was predicted by a shift in the optimal asynchrony for experiencing the McGurk effect towards greater auditory lags (Ipser & Freeman, in prep). One possible explanation is that some dyslexics have learned to pay more attention to visual lip-movement cues in order to disambiguate auditory phonemes, but these cues take longer to process. If so, then training with delayed video might help to reinforce the audiovisual association between voice and lips, thus improving phonemic awareness.

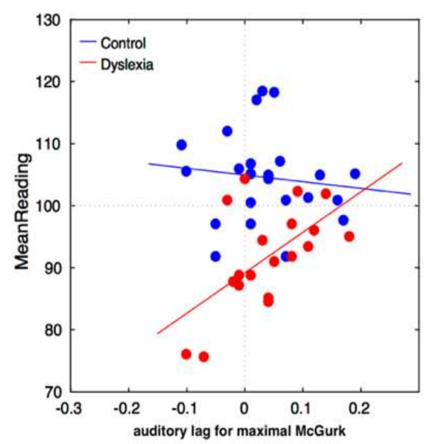

Figure 2. Reading scores (Woodcock battery) against Auditory lag for maximum McGurk (sec), for dyslexic and control participants.

Asynchrony and Speech comprehension

While the McGurk illusion is a strong measure of audiovisual integration it suffers from poor ecological validity, because it involves repeated presentations of a restricted set of single syllables, in the context of incongruous lip-movements. It is therefore arguable that the large individual differences in timing observed in the above McGurk-based tests could reflect participants' unfamiliarity with this context. For greater ecological validity, we have therefore modified the paradigm to assess comprehension of degraded natural speech in the context of asynchronous but congruent lip movements (Ipser et al, in prep). For each subject, we found the critical asynchrony that produced maximum accuracy in identifying unique noise-vocoded word stimuli. This study also aimed to assess whether individual differences in our measures had the trait-like retest-reliability characteristic of 'Personal Equations' from the classical literature.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-097

In support of such reliability, we found significant positive correlations between measures of optimal asynchrony obtained from two repetitions of the McGurk task, and a similar correlation for the Degraded Speech task (Figure 5). The stability of our measurements is encouraging from the applied therapeutic perspective, because we may be able to improve speech comprehension by finding an individual's Personal Equation and artificially delaying sound or vision to compensate for it (see next section).

Surprisingly, we found no correlation between optimal asynchronies measured for McGurk and Degraded Speech tasks. This lends further support to the argument that each of us has different Personal Equations associated with different tasks and stimuli. This result argues against theories of unified timing mechanisms across the brain [5,6,8,9], and supports instead an explanation based on temporally independent neural networks supporting these two superficially similar speech-integration tasks.

The null correlation between tasks here contrasts with the negative correlation observed in the previous study with McGurk and TOJ tasks, which were performed on exactly the same stimuli. This suggests that neural asynchronies associated with different tasks may only be renormalized relative to each other, when they are evoked by the same stimuli. This provides further constraints for conceptualising and modelling the renormalization process.

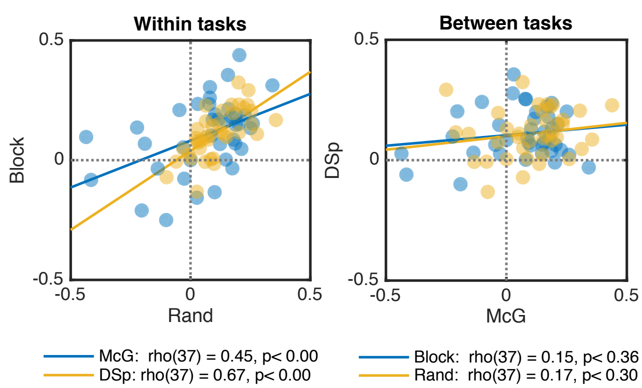

Figure 5. Scatterplots of individual measures of the asynchrony for maximal integration (tMax) for McGurk and Degraded Speech tasks. a) Positive correlations across repetition of the same task with randomized versus blocked order of asynchronies; b) Null correlations between tasks.

Practical applications

One concrete therapeutic implication from the present data is that speech comprehension may actually be enhanced for some individuals with naturally asynchronous perception. Our results from the Degraded Speech task show that on average an auditory delay of 107ms (SD 8.8ms) results in correct identification of 16.4 more words in every 100 (SD 10). The highest 50% of our sample showed improvements of 21 percentage points (Figure 4).

To obtain such benefits outside the lab, one would first need to measure each individual's personal asynchrony for speech comprehension, and then apply a compensatory auditory delay. This could be implemented via a programmable delayed-sound hearing-aid or even via headphones driven by a smartphone. A similar approach could be used improve speech comprehension and audiovisual integration in telecommunication, media streaming, telepresence, television, and gaming (patent pending [11]). Our findings with dyslexia additionally points to a possible educational application for delayed-sound therapy.

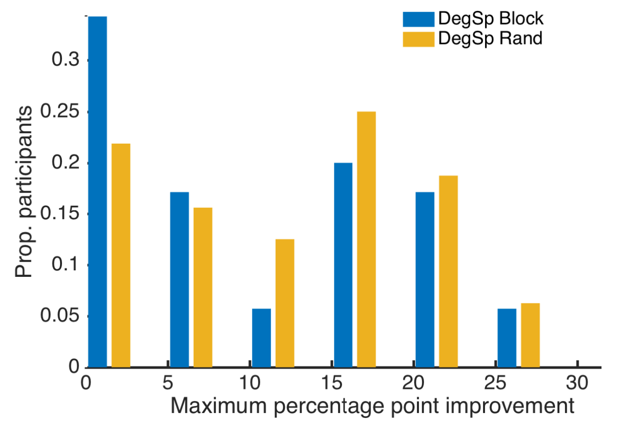

Figure 4. Histogram showing the distribution of percentage point increases of McGurk effect (blue, open) and comprehension accuracy (yellow) when stimuli are desynchronized to match each individual's optimal asynchrony.

Discussion

Though it has been known for centuries that people can err in their judgment of the timing of light relative to sound, the causes of this have been debated for just as long. The dominant view ascribes these errors to response biases or attentional 'prior entry' biases which might selectively speed vision relative to audition or vice versa [12]. But most previous studies have relied on explicit subjective measures of timing which are notoriously prone to such biases. As our new paradigm is objective and implicit (measuring asynchrony indirectly via performance on phoneme or word identification tasks), it escapes from this response bias problem to some extent. Our observations using this paradigm, that perceptual asynchronies measured for different tasks are uncorrelated or even negatively correlated, undermine the prior entry explanation which would assume that all tasks are subject to the same basic bias towards vision versus audition. That assumption would falsely predict positive correlations across tasks.

The alternative that needs now to be taken seriously is that these persistent personal asynchronies are related to delays associated with the propagation and processing of information within different neural networks subserving different perceptual functions. These delays have a demonstrable impact on higher cognitive abilities such as reading and speech comprehension. But the good news is that these delays are relatively stable and can be quickly and easily measured using a simple automated procedure. The resulting Personal Equations can then be used to artificially resynchronise the senses and consequently enhance speech comprehension and presumably many other perceptual and cognitive abilities.

Conclusions

In summary, our new observations allow us to begin to (1) fractionate 'multisensory perception' functionally and structurally into distinct mechanisms, (2) to relate complex abilities such as reading and speech-reading to a profile of easily-measured perceptual variables, leading to potential remedial applications, and (3) to understand how, though sight and sound cannot be perfectly synchronised, we can nevertheless still perceive the timing of events in our multisensory world accurately, on average.

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-097

References

[1] J. Mollon and A. Perkins, "Errors of judgement at Greenwich in 1796.," *Nature*, vol. 380, no. Mar 14, pp. 101–102, 1996.

[2] E. D. Freeman, A. Ipser, A. Palmbaha, D. Paunoiu, P. Brown, C. Lambert, A. Leff, and J. Driver, "Sight and sound out of synch: Fragmentation and renormalisation of audiovisual integration and subjective timing," *Cortex*, vol. 49, no. 10, pp. 2875–2887, Nov. 2013.

[3] S. Soto-Faraco and A. Alsius, "Conscious access to the unisensory components of a cross-modal illusion.," *Neuroreport*, vol. 18, no. 4, pp. 347–350, Mar. 2007.

[4] H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, vol. 264, no. 5588, pp. 746–748, 1976.

[5] R. B. Ivry and R. M. C. Spencer, "The neural representation of time.," *Curr. Opin. Neurobiol.*, vol. 14, no. 2, pp. 225–232, Apr. 2004.

[6] D. Bueti, "The Sensory Representation of Time," *Front. Integr. Neurosci.*, vol. 5, no. August, pp. 1–3, 2011.

[7] W. Fujisaki, S. Shimojo, M. Kashino, and S. Nishida, "Recalibration of audiovisual simultaneity.," *Nat. Neurosci.*, vol. 7, no. 7, pp. 773–8, Jul. 2004.

[8] J. V. M. Hanson, J. Heron, and D. Whitaker, "Recalibration of perceived time across sensory modalities.," *Exp. Brain Res.*, vol. 185, no. 2, pp. 347–352, Feb. 2008.

[9] L. R. Harris, V. Harrar, P. Jaekl, and A. Kopinska, "Mechanisms of simultaneity constancy," in *Issues of Space and Time in Perception and Action*, R. Nijhawan, Ed. Cambridge University Press, 2008, pp. 232–253.

[10] J. C. Ziegler, C. Pech-Georgel, F. George, and C. Lorenzi, "Speech-perception-in-noise deficits in dyslexia," *Dev. Sci.*, vol. 12, no. 5, pp. 732–745, 2009.

[11] E. D. Freeman, "GB2511909 - Method & apparatus for calibrating and correcting for neural asynchronies," 2014. https://www.ipo.gov.uk/p-ipsum/Case/ApplicationNumber/GB1400335.4.

[12] C. Spence, D. I. Shore, and R. M. Klein, "Multisensory prior entry," *J. Exp. Psychol. Gen.*, vol. 130, no. 4, pp. 799–832, 2001.

Author Biography

Elliot Freeman received his BSc in Psychology at Cardiff University (1992) and PhD in Psychology at University of Bristol (1998) with Professor Tom Troscianko. He worked as a post-doctoral fellow at the Institute of Cognitive Neuroscience, University College London with Professor Jon Driver until 2007. During that period he held visiting postdoctoral fellowships at Weizmann Institute, Salk Institute and Smith-Kettlewell Eye Research Institute. In 2007 he was appointed Lecturer at Brunel University, and in 20009 moved to City University London where he is a Senior Lecturer (Associate Professor) in Psychology, and member of the Cognitive Neuroscience Research Unit. Research interests include audiovisual integration, synaesthesia, individual differences, perceptual multistability, visual attention and perceptual grouping; methods include psychophysics, EEG, fMRI, and transcranial electrical stimulation. His home page can be found at http://www.staff.city.ac.uk/~sbbf269