

Hearing Lips and Seeing Voices: the Origins and Development of the ‘McGurk Effect’ and Reflections on Audio–Visual Speech Perception Over the Last 40 Years

John MacDonald ^{*,**}

Department of Psychology, University of the West of Scotland, Paisley, PA1 2BE, UK

Received 6 November 2016; accepted 20 January 2017

Abstract

In 1976 Harry McGurk and I published a paper in *Nature*, entitled ‘Hearing Lips and Seeing Voices’. The paper described a new audio–visual illusion we had discovered that showed the perception of auditorily presented speech could be influenced by the simultaneous presentation of incongruent visual speech. This hitherto unknown effect has since had a profound impact on audiovisual speech perception research. The phenomenon has come to be known as the ‘McGurk effect’, and the original paper has been cited in excess of 4800 times. In this paper I describe the background to the discovery of the effect, the rationale for the generation of the initial stimuli, the construction of the exemplars used and the serendipitous nature of the finding. The paper will also cover the reaction (and non-reaction) to the *Nature* publication, the growth of research on, and utilizing the ‘McGurk effect’ and end with some reflections on the significance of the finding.

Keywords

McGurk effect, audiovisual speech perception, hearing lips and seeing voices

1. Introduction

This paper is dedicated to the memory of Harry McGurk. Harry was my PhD supervisor and research colleague while I was at the University of Surrey, in Guildford, in the UK, and where the research I shall be describing was carried out. Harry was an extremely dynamic individual and was invigorating to work with. Academically I benefited considerably from the experience of working

* Emeritus Professor of Psychology.

** E-mail: profjohnmac@btinternet.com

with him. He died in 1998 and it is sad that he is not around to share this celebration of the auditory–visual illusion that he and I discovered, subsequently known as the ‘McGurk effect’.

In November 1974, I joined the University of Surrey as a research fellow to work on a project with Harry. The project was entitled ‘Development of coordination between vision and hearing during early infancy.’ It was funded by the then SSRC (Social Science Research Council). The main aim of the project was to assess the capacity of the very young infant to coordinate the perceptual activity involved in looking and listening and to trace the development of this ability over the first year of life. The prevailing view at that time was that the development of the senses entailed moving from a position where each sense was processed independently, and that the task for the baby was to integrate these independent sources to produce a unified perceptual world. We investigated this experimentally. What this entailed was presenting young infants — 3 to 18 months old — with still pictures of objects (people and other objects) and simultaneously playing them sounds (voices or non-speech sounds) and measuring how much visual attention they paid to these combinations. The question was whether the pattern of their visual attention was disrupted when either or both, the picture or sound was changed. Harry’s view was a Piagetian perspective of the development of the senses being one of a process of sensory integration. In 1974 Harry and Michael Lewis (McGurk and Lewis, 1974) had tried and failed to replicate a finding that one-month-old infants were disturbed by the dislocation of their mother’s face and voice (Aronson and Rosenblum, 1971). Even up to seven months the infants were not disturbed by the dislocation. Aronson and Rosenblum (1971) were working within a Gibsonian framework that posited very early unity of the senses and was the main counterpoint to the sensory integrationist position.

The research SSRC project had been running for a year before I arrived. The previous assistant had left in August of that year. As was characteristic of the way Harry worked, the studies had been detailed in the original application and my role was to finish the series of experiments over the remaining two years. The two main paradigms used were habituation and visual selective attention. The visual stimuli were rear projected coloured slides of female faces or coloured abstract patterns and the auditory stimuli were female voices reciting nursery rhymes or continuous musical chimes or tones. This work was subsequently published in a number of papers, for example, McGurk and MacDonald (1978).

2. The Development and Creation of the Stimuli

In the middle of 1975, Harry and I began to think about what the next project could be in order to apply for another research grant. Again this was typical

of Harry thinking ahead. Before applying for a grant he liked to have carried out some pilot work in order to strengthen the case for the proposed research. Although thinking a grant ahead is now the norm for researchers, that was not the case in the mid 1970s.

A number of factors coalesced at this time. The first was that we both agreed that it would make more sense if we could use dynamic visual stimuli, principally to increase the realism of the stimuli and provide a more ecologically valid perceptual situation. In addition, video technology had progressed such that it was now more affordable, smaller and within the reach of non-specialists in terms of use. We decided therefore to concentrate on social stimuli (people) and manipulate faces and voices in a controlled manner and observe infants' reactions to these stimuli. The basic idea being to have 'congruent' stimuli, where the face and voice would present the same speech token and comparing these with 'incongruent' stimuli where each modality would present different speech tokens. Would the infants react differently to the 'incongruent' stimuli from the 'congruent' stimuli? Our reasoning was that if the infants had a coordinated audio-visual space then the 'incongruent' stimuli would disrupt their visual attention (in the same way that adults experience a 'discomfort' when viewing badly dubbed or out of synchrony images on film or television). Harry's original idea was to use similar materials to those in the SSRC study, i.e., nursery rhymes as the auditory speech. However, I had been independently looking at the infant speech perception literature and argued that we should try and control the relationship between the face and voice in a more systematic manner. I argued that if we used the nursery rhymes, although we could start the audio and visual at the same time, when the audio nursery rhyme was different to the visual nursery rhyme, they may go in and out of synchrony in unpredictable ways, unless we could very carefully control what was being said in each modality. My argument was that we should use simpler speech sounds, i.e., consonant-vowel (CV) combinations, which would allow us to both control onset and duration in each modality. We could easily adjust the length of the stimulus presentation by having repetitions of each CV combination. We anticipated from the previous studies that to show differential behaviour in the infants we would need to be able to extend the basic trials to around 15–20 s. As there had been recent research published on infant speech perception (Eimas, 1974), and a wealth of research from the Haskins and other laboratories, we initially limited ourselves to the stop consonants and nasals. Also it is not uncommon for adults to speak these simple sounds to babies and the early sounds that babies produce are simple consonant-vowel combinations. Hence we arrived at the basic stimulus that we used in our eventual experiments — e.g., /baba/ /baba/ /baba/.

Although we were able to present videotape materials, we did not have the facilities to record and edit high quality materials. Hence we enlisted the help

of the University Audio–Visual Aids (AVA) department to record master versions of the stimuli. We also recruited Ms Susan Ballantyne as our model. We chose Susan as she had some amateur dramatic experience and did not feel awkward about repeating these simple phrases to camera, she had a southeast of England accent (which our anticipated baby participants would be familiar with), and critically Harry knew her and her husband socially. The initial stimuli (the consonants with /a/ as the following vowel) were duly recorded in the AVA studio, one evening. I then provided the AVA technician with the stimulus combinations we wished. These included the incongruent versions (e.g., audio /baba/ dubbed onto a visual /gaga/) and congruent versions (audio /baba/ dubbed onto visual /baba/). Hence the congruent stimuli would be second-generation, as would the incongruent stimuli.

The AVA unit was an independent department of the University; therefore we had to wait until they could fit our request into their other commitments. In addition, Harry and the Head of the AVA unit had some disagreement in the past so that it was left to myself and the AVA technician to liaise. This then was the background to why we did what we did, fully anticipating running some pilot studies with infants along the lines of those we had conducted in the SSRC project.

When eventually the tapes were ready I went to view them in the AVA unit studio. My immediate reaction was that some were fine, but the incongruent ones did not sound right. I wasn't hearing what I was expecting. I anticipated that the 'incongruent' stimuli would seem discordant in some way, but that was not what I experienced. I heard sounds but they didn't sound quite right, they sounded like different speech sounds.

My immediate reaction was panic. Had something gone wrong in the recording or the dubbing process? Would we have to rerecord everything? Telling Harry this news did not bear thinking about. I went back to the office and confronted him with the news that the tapes were ready, but that maybe he should come and view them as there may be a problem with the recordings. His initial reaction was exasperation as this was another example of the AVA unit not working to the standards he expected. We went and viewed the materials and to his credit, he did not explode when he too thought that the AVA technicians had 'messed up'. It was he who took the crucial step of listening to the dubbed tapes with his eyes closed and then open. Sure enough there was nothing wrong with the soundtrack, it was simply that we heard a different sound when watching the incongruent stimuli than when we only listened. In the audio–visual conditions we were experiencing an auditory illusion. A number of versions of our discovery have grown up over the years with some exaggeration having crept in (some of it from Harry himself). His response was measured and calm. No one was threatened with the sack as is sometimes reported.

We subsequently showed the tapes to a number of colleagues and students in the Psychology Department who confirmed our observations. What we found was that the overwhelming majority of people reported something other than the ‘sound’ presented and their responses were largely consistent. In one example, when the face is mouthing ‘gaga’ and the voice is saying ‘baba’, the report is predominantly ‘dada’. You ‘hear’ something that isn’t there.

3. What Was This Phenomenon?

Neither Harry or I expected this, nor did we know whether this was already a well-known phenomenon in the speech literature. Harry reasoned that if this was a novel finding then we should be the first to report it. I searched the science literature, which at that time entailed manually going through paper-based citations and citation indices. My search involved the ventriloquism literature, lip-reading and journal abstracts, including the *Journal of the Acoustical Society of America*, and found nothing like it reported elsewhere. What we did not do was contact other speech perception researchers — that came later. During this period we continued to run some further experiments with the initial stimuli and further combinations of sounds.

4. Publication

Having run the experiments and confirmed that the ‘illusory’ response was robust we discussed writing a paper and submitting for publication. My inclination was to send to a mainstream perception journal, such as *Perception and Psychophysics*. Harry wanted to send to *Nature*, the most prestigious scientific journal. I was sceptical but agreed on the basis that *Nature* had very fast decision processes and that if it was rejected we could quickly turn it round and send somewhere else. It would also mean that we would get feedback from referees who would confirm or otherwise whether this finding was already known in the speech community or a genuinely novel finding, or even worse that there was some fundamental flaw in what we had done and been unaware of.

We sent the paper to *Nature* on the 14th of July 1976. We received a positive response and after some minor changes it was accepted in November and published in the last week of December 1976. I often wondered whether the editors thought that a quirky paper like this was suitable for the Christmas edition (McGurk and MacDonald, 1976). The paper states that “Appropriate analyses confirm that the various effects reported for the auditory–visual condition are statistically significant”. I have been asked on a number of occasions what these analyses were. My recall is that we tried a number of procedures centred round chi-square, comparing the observed errors in the auditory–visual

conditions with expected or estimated errors taken from the auditory only conditions. The argument was that if there were no influence of the visual stimuli then the errors should be comparable under the two conditions. However, after 40 years and with none of my original notes available, I cannot be sure. The original draft submitted to *Nature* is no more informative.

5. The Impact

Once we knew that the paper had been accepted, we then contacted UK and US speech researchers to alert them to what we had found and the paper's publication. At that time there were two main centres in the UK — Adrian Fourcin at University College, London, and Mark Haggard and Quentin Summerfield at the Institute of Hearing Sciences in Nottingham University. The reaction we received was a mixture of surprise — it was a novel finding, and I think a slight feeling of resentment that this had been discovered by two psychologists with little background in speech research. It was at this time that we were alerted to the work of Barbara Dodd (Dodd, 1977) and her research, which also reported a significant influence of visual speech in her dubbed condition where vision and hearing were in competition. Her technique was to present the stimuli 'live' to her participants, hence she never experienced the conflicting stimuli. Harry and I often wondered what would have happened had her stimulus conditions been pre-recorded. Would she have been the first to 'discover' the 'illusion'?

From a scientific perspective the rest as they say is history, although in the early years there were few citations and follow up of our initial findings, apart from ourselves (MacDonald and McGurk, 1978; MacDonald *et al.*, 1978).

However, there was interest from a variety of unexpected directions. Some popular science journalists picked up on the *Nature* paper. One was Bryan Silcock who wrote for the *Sunday Times* and included a piece in the Magazine section on 1 May 1977. Although this was in a prominent publication it went largely unnoticed as the main feature was an account of the David Frost–Richard Nixon interviews which were eagerly anticipated assuming Nixon would come clean regarding his activities. We found out later that the *Sunday Times* received a number of letters in response to our piece. None were complimentary and tended to fall into one of three categories: (i) this finding is simply not true, we hear perfectly well without vision; (ii) this research is showing something we all know, is obvious and therefore redundant (one correspondent even reporting that she put on her spectacles while answering the telephone), and (iii) this is nonsense research and why are Universities wasting taxpayers' money funding it. The University also received some letters in a similar vein, one of which I think called for our dismissal.

We also received invitations to speak about the illusion, at some surprising (to us) conferences and seminars, including a seminar on 'Letterforms as Articulation Diagrams' at the School of Oriental and African Studies in the University of London, November 1977, and the Cybernetics Society in June 1985.

I have not yet tracked down when the term 'McGurk effect' first came into being and am sometimes asked if I feel resentful that my name rarely appears when it is mentioned. Generally I don't. Where I do get a little annoyed is when descriptions of the effect, either fail to mention me at all, or misname me. For example, in one early account, in *New Scientist* (March, 1977), although I am John MacDonald at the beginning of the article by the end of it I have morphed into someone called Johnson. Neither Harry nor I ever worked with anyone of that name. I have now got used to the 'a' in 'Mac' getting routinely excised, and the 'D' being decapitalised to 'd', even in references where one would imagine the editorial processes would pick up such errors.

Since 1976, there have been 4800+ citations of the paper published in *Nature*. Generally the subsequent research has dealt with three aspects: when does the illusion occur (the evidence), where does it occur (the neurophysiology) and why does it occur (the theory)? For a succinct summary of some issues raised by the McGurk effect, see Bernstein *et al.* (2002). One aspect that is clear in retrospect is how the illusion moved the focus of visible speech research beyond being the issue of compensatory information in the deaf and hard of hearing to the realisation that speech is a multimodal phenomenon and later as a key demonstration of multisensory processing and integration.

One aspect that puzzled us at the time was the prevalence of the illusion across participants. In the first study published in *Nature*, illusory responses varied between 98% in adults to ~50% in children, and varied with the auditory–visual combinations used. This variation was even more pronounced when we expanded the range of consonant combinations used and reported in the second study, published in *Perception and Psychophysics* (MacDonald and McGurk, 1978). At the time, and lacking any other explanation, we attributed this to being due to differences in the stimuli used, variation in the presentation conditions, or response biases and/or variation in attention on the part of the participants. We did not try to uncover the source of this variability. However, we did note that people who experienced the illusion continued to experience it even when it was pointed out to them. (See Nath and Beauchamp, 2012, for more recent work on individual susceptibility to the illusion.) Unfortunately, the original stimuli appear to have been lost and it is not possible therefore to rerun the first experiments with the original stimuli. (Both Harry and I moved institutions, video formats changed and I suspect the tapes were simply lost or ditched.)

Although we both kept some interest in the effect over subsequent years, both Harry and I moved on in terms of careers, research interests and geography. Harry's academic interests had always been underpinned by a strong sense of socially relevant child and family research and he pursued this in his move, first to the Thomas Coram Research Unit in the University of London, and subsequently at the Australian Institute of Family Studies at the University of Melbourne.

I moved from a research position at Surrey to a full-time lecture position at Portsmouth University, which initially left little time for research and my interests moved to collaboration with new colleagues. I did occasionally go back to the 'McGurk effect', particularly when students needed a relatively ready-made project to conduct.

One question that had intrigued us from the start was what type and level of detail of facial information was required for the effect to occur? The traditional assumption always seemed to be that phonetic or phonological information was being extracted from the visual information and combined with that from the auditory component to form the resultant percept. However, lip reading research showed that normally hearing people in general were relatively poor at identifying speech information from lip movements (especially where context did not allow disambiguation of the alternatives), and the hard of hearing were only marginally better. It seemed to us then that whatever information was being extracted was at a low level, i.e., not detailed.

In 1998 a new colleague, Dr Tallis Bachmann joined the Department in Portsmouth. His previous research had used spatial quantisation techniques to investigate perceptual processes. With him and an undergraduate student (Søren Andersen) we carried out a set of studies where we systematically varied the level of detail in the visual image. We found that although the prevalence of the illusion reduced at the higher levels of quantization it did not disappear until the visual stimulus was no longer recognized as a face. Hence, we concluded that the information being extracted from the face was not fine level detail about the face and lip movements, but relied only on them picking up relatively gross features of movement. (MacDonald *et al.*, 1999, 2000, 2001).

6. How and why Does the Illusion Occur?

This is still an open question. In 1976 the dominant approaches in speech perception were to try and explain our perceptual experience by relating it to the physics of the speech sound waves — an auditory perspective — the psychoacoustic theories (Blumstein, 1986; Diehl and Kluender, 1989).

The major alternative theoretical approach was the 'motor' theory (Lieberman and Mattingly, 1985; Mattingly and Studdert-Kennedy, 1991). In this

account it is proposed that we perceive speech by detecting, from the auditory input, what articulatory gestures produced the stimulus being presented. How that was achieved was rather underspecified in the early version of this theory, which was proposed in the 1960s and predates our 1976 study by a decade or so. However, the *Nature* paper, which showed an influence of visible speech and facial movements, was eagerly taken up by the ‘motor’ theorists as strong (unequivocal support) for their position.

A third theoretical position that has a controversial history in the psychology of perception (Gibson, 1979), but has only relatively recently been applied to audiovisual speech perception is that of ‘direct perception’ (Fowler, 1996). Here it is proposed that the function of our sensory systems is to perceive the causes of the sensory input we receive. For example, in the case of speech the cause of the sensory stimulus is the vocal tract activity of the speaker, i.e., what did you do to make that sound. The major challenge for this type of theory was to explain how listeners did this from solely auditory information. However, for our purposes this theory has no problem with audiovisual speech perception. Visible speech information simply provides another (potential) source of information about the speaker’s vocal tract activity.

Clearly there is much further work to be done to distinguish between these and other theoretical positions that have been developed more recently [e.g., FLMP (Massaro, 1998)]. See, Van Wassenhove (2013) for a fuller discussion of these issues.

7. What Does It All Mean?

However, one thing that puzzled us from the outset was why normally hearing individuals are influenced by the facial information, when the visual stimulus is not necessary to perceive the auditory speech. We formulated this as ‘Why does the brain do this when it generally doesn’t need to?’. It is only under very restricted circumstances that the auditory information is so poor that visible speech information would be useful, and generally in such circumstances contextual information would help resolve any uncertainty about what was being said. For a number of reasons I think this question is misconceived, and it should be framed as ‘Why would the brain not use such available information when it is built so to do?’. The behavioural evidence from the illusion experiments and the neurophysiological studies both show that more than the auditory pathways are involved in speech processing and perception. An interesting perspective on this is to view speech from an evolutionary viewpoint.

8. The Evolution of Speech and Language

The idea that speech perception should be thought about more as vocal tract activity, or vocal gesture, rather than a purely acoustic phenomenon is gaining much ground. Recently Professor Michael Corballis, of the University of Auckland, has written about the evolution of speech and language (Corballis, 2002). If one views speech as an auditory problem then one might propose that human speech has evolved from animal cries. However, in primates there is a fairly limited repertoire of cries and the vocal apparatus of most primates is limited in range and complexity. In contrast, Corballis has advanced the alternative view that speech and language has evolved not from animal cries and sounds but from gesture, firstly, through the use of manual gesture and then over the course of evolution through facial expression and facial gesture. Early hominids and primates have good manipulative abilities. There would be an evolutionary advantage to moving gesture to another area (i.e., the face). It would release the hands for other tasks. With associated changes in vocal tract structure and complexity, these visual gestures would be augmented by sounds that would allow a more complex set of speech tokens to be used. Sound of course has distinct advantages over visual information — you can use sound at a distance and in the dark, and the listener does not have to be watching you. In this scenario, our current use of and sensitivity to visible speech information is a residue of our evolutionary history. The fact that we often still use facial and manual gesture in our speech lends support to this view.

The ‘McGurk illusion’ shows that facial speech information is not simply an adjunct of auditory speech but is an intrinsic component of normal speech perception. This also fits very well with the kind of account that Corballis is putting forward regarding the evolution of speech and this theory provides an explanation of why we process and are affected by visible speech information and why we experience this illusion.

9. Final Reflections

At various times, both I (more so) and Harry (less so) had anxieties about what we had found. Did it mean or show anything significant about human speech perception? We waxed and waned between feelings of it being a significant finding to that it was perhaps only a ‘quirk’ of the experimental circumstances we had created. However, subsequent replications and citations by esteemed colleagues and laboratories across the world have to a large extent allayed these doubts. The fact that the illusion is robust to the range of methodologies that have been used — natural and artificial voices and faces, stimulus tokens (CV, VCV etc.), language groups, non-tonal and tonal language speakers — I think attests to the importance of the finding. It has turned out to be a much more important study than we realized at the time.

If Harry had lived he would now be 80 years old and the ‘McGurk illusion’ would have been around for half of his life. He was a cultured man who enjoyed poetry and I think he would have been pleased to have the following as a fitting epitaph to this part of his academic career.

Hauf his soul a Scot maun use
Indulgin’ in illusion
And hauf in getting rid o them
And comin’ to conclusions.
(Hugh McDiarmid, 1928)

Acknowledgements

Thanks are due to a number of people who have been critical to this research, namely, the technicians (at Surrey, Keith and Kevin of the AVA, Dominic and Jenny of the Psychology Department; at Portsmouth, Dave of the Psychology Department); the models (Susan Ballantyne and Rachel Seymour); academic colleagues (Talis Bachmann and Soren Anderson), and most importantly, Harry.

References

- Aronson, E. and Rosenblum, S. (1971). Space perception in early infancy: perception within a common auditory–visual space, *Science* **172**, 1161–1163.
- Bernstein, L. E., Burnham, D. and Schwartz, J.-L. (2002). Special session: Issues in audio-visual spoken language processing (When, Where, and How?), in: *Proceedings of the 7th International Conference on Spoken Language*, Denver, CO, USA, pp. 1445–1448.
- Blumstein, S. E. (1986). On acoustic invariance in speech, in: *Invariance and Variability in Speech Processes*, J. S. Perkell and D. H. Klatt (Eds), pp. 178–197. Erlbaum, Hillsdale, NJ, USA.
- Corballis, M. C. (2002). *From Hand to Mouth: the Origins of Language*. Princeton University Press, Princeton, MA, USA.
- Diehl, R. L. and Kluender, K. R. (1989). On the objects of speech perception, *Ecol. Psychol.* **1**, 121–144.
- Dodd, B. (1977). The role of vision in the perception of speech, *Perception* **6**, 31–40.
- Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants, *Percept. Psychophys.* **16**, 513–521.
- Fowler, C. A. (1996). Listeners do hear sounds, not tongues, *J. Acoust. Soc. Am.* **99**, 1730–1741.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA, USA.
- Lieberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception, *Cognition* **21**, 1–33.
- MacDonald, J. and McGurk, H. (1978). Visual influences on speech perception processes, *Percept. Psychophys.* **24**, 253–257.

- MacDonald, J., Dwyer, D., Ferris, J. and McGurk, H. (1978). A simple procedure for accurately manipulating face-voice synchrony when dubbing speech onto videotape, *Behav. Res. Meth. Instrum.* **10**, 845–847.
- MacDonald, J., Andersen, S. and Bachmann, T. (1999). Hearing by eye: Visual spatial degradation and the McGurk effect, in: *Proceedings of Eurospeech '99*, G. Olaszy G. Nemeth and K. Erdohegyi (Eds), Vol. 3, pp. 1283–1286. European Speech Communication Association, Bonn, Germany.
- MacDonald, J., Andersen, S. and Bachmann, T. (2000). Hearing by eye: how much spatial degradation can be tolerated? *Perception* **29**, 1155–1168.
- MacDonald, J., Andersen, S. and Bachmann, T. (2001). Read my lips, but not too closely: What face information is used in the perception of speech? in: *XII ESCOP and XVIII BPS Cognitive Section Conference*, Edinburgh, UK.
- Massaro, D. W. (1998). *Perceiving Talking Faces: from Speech Perception to a Behavioral Principle*. MIT Press, Cambridge, MA, USA.
- Mattingly, I. G. and Studdert-Kennedy, M. (Eds) (1991). *Modularity and the Motor Theory of Speech Perception*. Erlbaum, Hillsdale, NJ, USA.
- McGurk, H. and Lewis, M. (1974). Space perception in early infancy: perception within a common auditory–visual space? *Science* **186**, 649–650.
- McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**, 746–748.
- McGurk, H. and MacDonald, J. (1978). Auditory–visual co-ordination in the first year of life, *Int. J. Behav. Dev.* **1**, 229–239.
- Nath, A. R. and Beauchamp, M. (2012). A neural basis for interindividual differences in the McGurk effect: a multisensory speech illusion, *NeuroImage* **59**, 781–787.
- Van Wassenhove, V. (2013). Speech through ears and eyes: interfacing the senses with the supramodal brain, *Front. Psychol.* **4**, 388. DOI:10.3389/fpsyg.2013.00388.