BRILL

Multisensory Research 31 (2018) 1–6

MULTISENSORY RESEARCH

brill.com/msr

Introduction to the Special Issue: Forty Years of the McGurk Effect

Michael S. Beauchamp *

Baylor College of Medicine, Houston, USA

In this special issue of *Multisensory Research*, we celebrate the fortieth anniversary of the publication in 1976 of the seminal paper by Harry McGurk and John MacDonald entitled "Hearing lips and seeing voices" (McGurk and MacDonald, 1976). The genesis of the special issue was a symposium honoring the anniversary at the fourteenth meeting of the International Multisensory Research Forum, held in Suzhou, China, in June of 2016 (Fig. 1).

In the first article in the special issue, John MacDonald recounts the history of the original publication (MacDonald, 2018). John was a graduate research assistant in Harry McGurk's laboratory at the University of Surrey, which studied the development in infants of the coordination between vision and hearing (what we would now term "multisensory integration"). Videotape technology had been recently developed, making it simpler to prepare and present audiovisual stimuli. To determine how infants perceived the incongruence between auditory and visual speech, Harry instructed John to make recordings of "baba" and "gaga" and arrange for the Audiovisual department to dub together congruent (auditory "baba" + visual "baba", AbaVba) and incongruent (auditory "baba" + visual "gaga", AbaVga) pairings of the syllables. John tells of his panic when on listening to the resulting incongruent recording, he instead heard something entirely different ("dada"). However, other observers experienced the same percept, and the finding was duly written-up and published in *Nature*. John touchingly devotes his article to the memory of his late mentor (who passed away in 1998 at age 62) and graciously reports that he has no hard feelings that the illusion has become known as the McGurk effect.

* E-mail: Michael.Beauchamp@bcm.edu

DOI:10.1163/22134808-00002598

Figure 1. Speakers at the fortieth anniversary symposium held on June 16th, 2016. From left: John MacDonald, Michael Beauchamp, Julia Irwin, Salvador Soto-Faraco. Jean Vroomen was not able to attend but delivered his lecture remotely.

The remainder of the special issue illustrates the broad scope of research that continues to be inspired by the original publication.

One common theme of research on the McGurk Effect examines different stimulus manipulations. Changing playback rate is a common occurrence in online media: a simple click in *YouTube* or other sites can change the playback rate from half-normal to double-normal. In the second article of the special issue, Magnotti, Basu Mallick, and Beauchamp describe the effects of playback rate on McGurk perception (Magnotti *et al*., 2018). Naively, one might expect that slowing playback would increase participants' ability to process visual speech and hence to increase the incidence of the McGurk effect, which depends on visual speech influencing the perception of auditory speech. Surprisingly, the opposite effect was observed: slowing playback rate significantly decreased the frequency of McGurk responses. Unlike the naive prediction, a quantitative Bayesian model of the McGurk effect accurately predicted participants' behavior. Under Bayesian inference, the perceiver weights each modality by its reliability. Slow playback rates increase the reliability of

the visual modality, and subjects therefore weight the modality more strongly and provide more visual responses. The article by Magnotti and colleagues demonstrates the importance of rigorous quantitative models in understanding the illusion and shows that the McGurk Effect still retains the capacity to surprise.

Another common theme of research on the McGurk Effect is the examination of brain processes underlying the illusion. In the third article of the special issue, Irwin, Avery, Brancazio, Turcios, Ryherd and Landi (Irwin *et al*., 2018) describe an EEG examination of audiovisual syllable perception. Irwin and colleagues examine a phenomenon known as phonemic restoration. In phonemic restoration, replacing part of an auditory syllable with noise changes the percept, for instance from a /ba/ to an /a/; pairing the deleted /a/ with a visual /ba/ restores the missing phoneme: subjects perceive "ba" without noticing the missing auditory information. While phonemic restoration, like the McGurk effect, is an example of the powerful influence of visual speech on auditory speech perception, it has the advantage that it does not require the presentation of incongruent auditory-visual information (which can be confusing for clinical populations or children) or the addition of high levels of auditory noise, as in speech-in-noise studies. One of the event-related potentials examined by Irwin and colleagues is the P300, an EEG measure of identification and discrimination that is typically elicited using an oddball paradigm, in which an occasionally-presented stimulus evokes the P300. In their study, the common stimulus was an AbaVba, while the deviant stimulus was AaVba. If the mouth of the visual talker was obscured with pixelation, the AaVba stimulus evoked a larger P300 and subjects detected it more reliably, showing a correspondence between brain and behavioral measures of auditory visual speech perception.

Advanced behavioral methods are also providing new insight into the McGurk effect. In the fourth article of the special issue, Sánchez-García, Kandel, Savariaux and Soto-Faraco combine two advanced techniques (Sánchez-García *et al*., 2018). The first was recording their audiovisual speech syllables with a high-speed visual camera that provided 10-ms resolution. The second was the use of a playback technique known as *gating*, in which a stimulus is presented repeatedly. First, only a short initial segment of the stimulus is presented (the first 10 ms); then more of the stimulus is presented (the first 20 ms) and so on, until the entire stimulus is shown. Participants respond after each presentation, and the stimulus duration on which they converge on their final responses provides a measure of the timing of the information present in the speech stimulus. There was significant variability among phonemes; interestingly, "th" was the only syllable that showed a clear multisensory benefit, with identification possible for shorter audiovisual stimuli than either unisensory auditory or visual syllables.

Another common theme of research on the McGurk effect are cross-cultural studies. This theme is represented in the fifth article of the special issue, in which Burnham and Dodd examine variants of the McGurk effect in native Thai speakers, native English speakers, and native Japanese speakers with increasing levels of English proficiency (Burnham and Dodd, 2018). Cross-cultural studies are valuable because they allow for the dissection of the contribution of different levels of language processing to the illusion. A key distinction is made between phonetic and phonemic representations. Audiovisual speech gestures may be perceptually different but have equivalent meaning in a particular language. For instance, "ch" and "sh" (as in "chop" and "shop") are separate phonemes in English but in Thai they are allophones, meaning they are phonetically distinct but do not impact meaning. Interestingly, Burnham and Dodd find evidence that in the McGurk effect, auditory and visual information is initially integrated at the phonetic level of processing, independent of the constraints of a given language. This suggests that the McGurk effect provides a window onto universal processes in speech perception that are shared regardless of the culture of the perceiver.

The sixth and the final article in the special issue, by Alsius, Pare and Munhall, is a critical review of the literature describing the McGurk effect (Alsius *et al*., 2018). Alsius and colleagues report large variability across the 28 studies of the effect that they examined. One likely contributor to this variability is that the stimuli used in original paper have been lost. Different laboratories have had to construct their own stimuli if they wish to study the effect. These stimuli are likely to vary in the quality of the auditory and visual speech information they contain and thus their ability to evoke the McGurk effect. The precise methods used for collecting and analyzing behavioral responses were not provided in the original paper. Subsequent investigations have thus used a variety of task instructions, task structure, response structure and scoring methods, adding additional sources of variability. This variability across studies of the McGurk effect may have obscured some very interesting observations. While in the original study only a single subject who did not perceive the illusion was reported, later studies have found a much higher proportion of individual who are not susceptible to the effect. Alsius and colleagues rightly point out that a better understanding of why these participants process audiovisual speech information differently than McGurk perceivers could enormously advance our understanding of the mechanisms at work in speech perception. Finally, Alsius and colleagues highlight an important debate in the field. Watching a face that is saying one syllable while hearing the same talker pronounce a different syllable is fundamentally an artificial construct. It remains unclear how the individual differences observed in the McGurk effect translate to individual differences observed in other, more natural audiovisual speech settings, such as speech in noise.

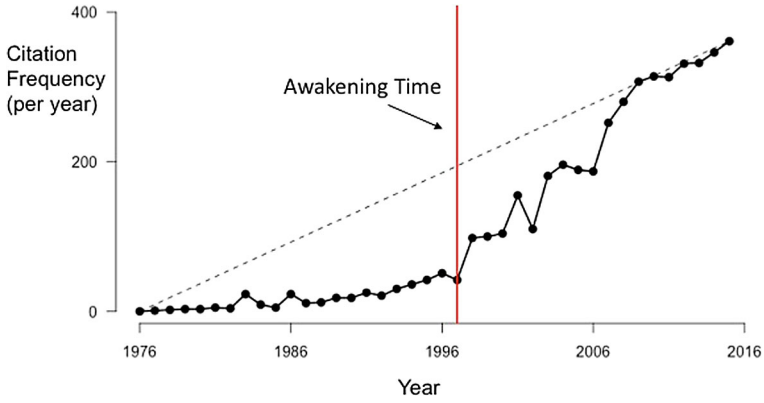

Figure 2. For each year since 1976, the number of papers published in that year citing the original study, according to *Google Scholar*. Awakening time estimated as in (Ke *et al*., 2015).

In summary, the special issue offers an excellent overview of the thriving community of researchers interested in uncovering the mysteries of the McGurk effect, more than 40 years after it was discovered. An interesting phenomenon in the sociology of science is known as the "Sleeping Beauty," in which the importance of a paper is not recognized for several years after publication (Ke *et al*., 2015). Initially, the number of citations to these papers is very low, but then grows rapidly after an "awakening time". This stands in distinction to the citation pattern for most papers, which are cited most often in the first few years after they are published, and then less so over time, presumably because their findings have been assimilated into the knowledge base or superseded by new discoveries. As shown in Fig. 2, the McGurk and MacDonald *Nature* paper displays the citation characteristics of a Sleeping Beauty. There were relatively few citations for many years after it was first published, followed by a rapid increase to the current rate of several hundred a year; the awakening occurred around 1997.

The history of the McGurk effect provides a few teachable points. The first is the power of serendipity. While McGurk and MacDonald purposefully set out to create incongruent AbaVga stimuli, they had no inkling that the resulting stimuli would produce a percept different than either the auditory or visual components of the stimulus. The second is the importance of keeping an open mind. If McGurk and MacDonald had discarded the unexpected result as useless or confusing, they would have missed their chance for a page in the history books. A third is the importance of staying the course and continuing all of the way to the finish line to produce a published paper. At the same time as McGurk and MacDonald, scientists in the USA were studying similar phenomena (which they termed the "fox-box illusion") but it was only ever published as a conference abstract and was essentially forgotten (Yonovitz *et*

al., 1977). Finally, the time series of the citations to the original study, like that of other scientific Sleeping Beauties, emphasizes the necessity of taking a long view of impact and significance. It may take decades for the importance of a discovery to be fully realized.

References

Alsius, A., Pare, M. and Munhall, K. G. (2018). Forty years after hearing lips and seeing voices: the McGurk effect revisited, *Multisens. Res.* **31**, 111–144.

Burnham, D. and Dodd, B. (2018). Language-general auditory-visual speech perception: Thai–English and Japanese–English McGurk effects, *Multisens. Res.* **31**, 79–110.

Irwin, J., Avery, T., Brancazio, L., Turcios, J., Ryherd, K. and Landi, N. (2018). Electrophysiological indices of audiovisual speech perception: beyond the McGurk effect and speech in noise, *Multisens. Res.* **31**, 39–56.

Ke, Q., Ferrara, E., Radicchi, F. and Flammini, A. (2015). Defining and identifying sleeping beauties in science, *Proc. Natl Acad. Sci. USA* **112**, 7426–7431.

MacDonald, J. (2018). Hearing lips and seeing voices: the origin and development of the 'McGurk effect' and reflections on audio-visual speech perception over the last 40 years, *Multisens. Res.* **31**, 7–18.

Magnotti, J. F., Basu Mallick, D. and Beauchamp, M. S. (2018). Reducing playback rate of audiovisual speech leads to a surprising decrease in the McGurk effect, *Multisens. Res.* **31**, 19–38.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**, 746–748.

Sánchez-García, C., Kandel, S., Savariaux, C. and Soto-Faraco, S. (2018). The time course of audio-visual phoneme identification: a high temporal resolution study, *Multisens. Res.* **31**, 57–78.

Yonovitz, A., Lozar, J. T., Thompson, C., Ferrell, D. R. and Ross, M. (1977). "Fox-box illusion": simultaneous presentation of conflicting auditory and visual CV's, *J. Acoust. Soc. Am.* **62**, S3. DOI:10.1121/1.2016172.