Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 50 (2004) 212–230
Journal of Memory and Language
www.elsevier.com/locate/jml

Effects of sentence context and expectation on the McGurk illusion

Sabine Windmann*

Institute of Cognitive Neuroscience, Ruhr-University Bochum, Biopsychology, GAFO 05, Bochum D-44780, Germany

Received 7 July 2003; revision received 30 September 2003

Abstract

Visual speech cues presented in synchrony with discrepant auditory speech cues are usually combined to a surprisingly clear unitary percept that corresponds with neither of the two sensory inputs (the McGurk illusion). This audiovisual integration process is commonly believed to be highly autonomous and robust to cognitive intervention, unlike the processing of ambiguous phonemes which has been shown to be dependent on lexical–semantic context and other higher cognitive variables. To investigate this issue, three experiments were carried out in which subjects' expectations were varied as they were presented stimuli containing the McGurk effect. In Experiments 1 and 2, the illusion was embedded in real words that were presented in semantically congruent vs. incongruent sentential contexts. In Experiment 3, nonlexical stimuli containing the McGurk illusion either matched or did not match subjects' prior expectations. Results show that the clarity of the illusion, and to some extent the probability of the illusion, was significantly influenced by subjects' expectations. Thus perceptions that are based on audiovisually integrated speech cues are not immune to cognitive influences; rather, they seem to be subject to the same functions and variations as ambiguous phonemes.
© 2003 Elsevier Inc. All rights reserved.

Keywords: Speech perception; Semantic context; Sensory integration; Audiovisual; Illusion; Top-down

Perception of heard speech can be facilitated or altered by visual observation of the speaker's lip movements. A striking demonstration of this fact is the McGurk illusion (McGurk & MacDonald, 1976). This refers to the following phenomenon: When the auditory syllable /ba/ is presented in synchrony with a speaker mouthing /ga/, subjects typically report understanding /da/. Thus, discrepant auditory and visual speech cues are integrated into a unified percept that corresponds with neither the auditory nor the visual stimulus originally presented. Such perceptual fusion occurs most frequently when labial auditory consonants are paired with nonlabial visual consonants (MacDonald & McGurk, 1978). By contrast, combinatorial responses (such as "bga") are usually induced by the inversed kind of pairing (auditory /ga/ and visual /ba/).

Several theories have been proposed to account for the McGurk illusion. Notably, all these theories are feedforward models with strong emphasis on the bottom-up flow of information, prior to lexical–semantic analysis. In essence, these theories describe how the visual and the auditory input signals are propagated forward to a common representational space whose format (or code) allows for their integration (Schwartz, Robert-Ribes, & Excudier, 1998). However, the theories differ with regard to the type of information this space conveys. In the "fuzzy-logical model of perception" (FLMP; Massaro, 1987, 1998), this space represents linguistic knowledge. The model states that visual and auditory speech cues are first analyzed separately and continuously up to the level of phoneme analysis. The results of these two independent processes are then

* Fax: +49-234-321-4377.
E-mail address: sabine.windmann@ruhr-uni-bochum.de.

0749-596X/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2003.10.001

combined and matched with stored phoneme prototypes based on a relative goodness rule (Massaro, 1987, 1996, 1998; Massaro & Stork, 1998).

The second type of theory assumes that the information integrated in the McGurk effect includes knowledge about speech gestures, vocal tract configurations and speech production programs. Prominent examples of these types of models are the "motor theory" (Liberman & Mattingly, 1985) and the "direct realist theory" of speech perception (Fowler, 1986; Rosenblum, 1989). These theories assume that listeners decode the gestures responsible for producing a given speech signal as they try to identify a spoken phoneme that is both heard and seen. Empirical studies with infants and adults clearly suggest that such sensory-to-motor mapping plays an important role in the acquisition and fine-tuning of speech *production* skills (Houde & Jordan, 1998; Kuhl & Meltzoff, 1988, 1996), but its specific role in speech perception is less clear.

In the third type of model, the "dominant recording model" according to Schwartz et al. (1998), auditory representations are dominant, but can be influenced by visual information (Calvert, Brammer, & Iversen, 1998; Diehl & Kluender, 1989), perhaps via the direct connections that exist between A1 and V1 (Bavelier & Neville, 2002). From a neuropsychological perspective, this is the most plausible account because lipreading and McGurk-like effects have been shown to modify activity in primary auditory cortex (Calvert, 2001; Calvert et al., 1998; Mottonen, Krause, Tiippana, & Sams, 2002; Sams et al., 1991; Wright, Pelphrey, Allison, McKeown, & McCarthy, 2003; but see Olson, Gatenby, & Gore, 2002). These data seem inconsistent with models proposing that audiovisual interaction occurs at higher linguistic levels or amodal levels of representation at which no direct crosstalk between primary auditory and visual signals occur (cf. Schwartz et al., 1998).

Nonetheless, in addition to the question of where in the brain audiovisual integration takes place, the underlying mechanism remains to be specified, both with respect to its temporal dynamics and in relation to other cognitive processes. In line with the essentially feedforward character of all theoretical accounts, most authors believe that the integration occurs automatically at very early speech processing levels, prior to phoneme identification (Calvert et al., 1997; Dekle, Fowler, & Funnell, 1992; Green, 1998; Green, Kuhl, Meltzoff, & Stevens, 1991; Langenmayr, 1997; Sams, Manninen, Surakka, Helin, & Kättö, 1998). This notion conforms with experimental reports that have described the McGurk illusion as extraordinarily robust and resistant against cognitive interventions. For example, informing subjects about the audiovisual discrepancy does not eliminate the illusion (Massaro, 1987), neither does practice (Summerfield & McGrath, 1984), or controlled attempts to report only one of the two modalities (Massaro & Cohen, 1983b; Massaro, 1998, pp. 244–250). Likewise, the illusion remains stable when the auditory and the visual cues are separated temporally (up to 180 ms) or spatially (Jones & Munhall, 1997; Massaro & Cohen, 1993; McGrath & Summerfield, 1985; Munhall, Gribble, Sacco, & Ward, 1996), and when subjects do not fixate the speaker's lips (Pare, Richler, ten Hove, & Munhall, 2003). Most strikingly, the illusion remains largely unchanged when the gender of speaker and voice are different, that is, when the voice of a female speaker is dubbed onto the face of a male speaker or vice versa (Green et al., 1991).[1]

Hence, even under conditions in which subjects realize that the information they receive cannot stem from a single source, they still fuse the visual and auditory information, suggesting that higher cognitive functions have little access to the integration process (Langenmayr, 1997). Finally, McGurk-like phenomena have been successfully demonstrated in preverbal infants (Kuhl & Meltzoff, 1982; Rosenblum, Schmuckler, & Johnson, 1997), and even in monkeys (Ghazanfar & Logothetis, 2003), suggesting that the effect must at least to some degree be independent of lexical–semantic capabilities. Although some theoretical formulations nonetheless consider such influences possible (Massaro, 1987, p. 49 ff.; Massaro, 1998), the empirical attempt of Sams et al. (1998) to demonstrate lexical–semantic influences on the illusion failed. As Dekle et al. (1992, p. 361) conclude, "the conditions under which the McGurk effect occurs or fails to occur must be described phonetically, not lexically or semantically" (or cognitively, one might add).

In summary, the McGurk illusion is widely considered to a highly autonomous phenomenon (Calvert et al., 1997; Dekle et al., 1992; Fowler & Dekle, 1991; Green et al., 1991; Langenmayr, 1997; Sams et al., 1998). Attentional control and higher cognitive context seem to have little, if any, modulatory impact on the frequency and strength on the illusion (Dekle et al., 1992; Langenmayr, 1997; Sams et al., 1998). This impression is likewise given by many visual illusions (Eagleman, 2001): No matter how hard one tries, the illusory effect cannot be prevented.

If this notion about the stability and the robustness of the McGurk illusion is true, one might suggest that audiovisually integrated speech cues result in perceptual representations that are stronger, more coherent and perhaps more bottom-up driven than perceptions of noisy or otherwise ambiguous phonemes. Unlike the McGurk illusion, such ambiguous phonemes have indeed been shown to be influenced by lexical–semantic context in numerous studies (Connine, 1987; Connine &

[1] Note, however, an interaction of this finding with face familiarity found by Walker, Bruce, and O'Malley (1995).

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات