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.

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).

Clifton, 1987; Ganong, 1980; Lucas, 1999; Newman, Sawusch, & Luce, 1997; Pitt & Samuel, 1993; Samuel, 1981, 1991, 1997, 2001; Samuel & Pitt, 2003). For example, if an ambiguous phoneme midway between /g/ and /d/ is presented in association with /?ift/, then most people would report the word "gift" while the opposite would happen with /?art/. This effect can be made stronger when the word is presented in the appropriate sentential context, e.g., "At her birthday, she received a valuable..." (e.g., Connine & Clifton, 1987; Gaskell & Marslen-Wilson, 2001; Samuel, 1981), thereby demonstrating sentential context effects on phoneme identification in words. As such effects demonstrate the influence of "higher," more abstract linguistic processes on decisions that can already be made on the basis of lower, sensory (phonetic) processes, these effects are often called top-down or concept-driven as opposed to bottom-up and data-driven (in the cognitivist terminology, see Engel, Fries, & Singer, 2001).

Although there is no doubt in speech perception research that such top-down effects exist, there has been much debate with regards to how these effects should be process-modeled, in particular whether *feedback projections* from higher levels of speech analysis (lexical, semantic) to lower levels (phonetic) are required to account for them. On the pro side, *interactive models* propose that the activity of word units at the lexical level is fed back to alter the activity at phonemic levels via backward projections. This would imply that the *perceptual encoding* of the phonemes is altered by higher linguistic and cognitive processes. A prominent example of this sort of model is TRACE (McClelland, 1991; McClelland & Elman, 1986).

On the contra side, *autonomous models* propose that speech information can only flow forward (bottom-up) along the various levels of the speech processing hierarchy, as in the McGurk theories described above. Such models can nevertheless account for lexical–semantic effects on phonemic identifications by implementing decision nodes at a (relatively late) processing level whose activity can be influenced not only by phonetic factors but by lexical–semantic levels as well (as in MERGE, see Norris, McQueen, & Cutler, 2000). Likewise, the Fuzzy-Logical Model of Perception (FLMP; Massaro, 1987, 1996, 1998; Massaro & Stork, 1998) is a non-interactive model that formalizes how the information from independent sources is combined and weighted as the system attempts to match current information with stored knowledge. FLMP assumes that this matching process will be held back for as long as possible to assemble all available evidence before a selection is made. Therefore the model is able to account for lexical–semantic influences on phonemic decisions without specifying feedback-connections. Notably, phonemic decision is not an all-or-none classification process in this model, but allows the fuzzy nature of human speech information to be reflected in subjects' evaluations and responses (Massaro & Cohen, 1983a, 1983b, 1991, 1995). However, it should also be noted that FLMP is a statistical model, not a process model in the sense of MERGE (which models *activation* levels, not response probabilities, see Norris et al., 2000, p. 356 f; McClelland, 1991). This model is therefore called "integrative" throughout the remainder of this article.

As evident from this description, a critical aspect in the ongoing disputes between interactive and autonomous process models is whether lexical factors are believed to change phoneme *perception/sensitivity* as opposed to *postperceptual* interpretations and biases (Connine & Clifton, 1987; Massaro, 1996; Norris et al., 2000; Samuel, 1981, 2001; Samuel & Pitt, 2003). If lexical levels can be shown to improve accuracy of phoneme *perception*, not just decision/identification bias, then this would speak against autonomous models, as pointed out by Norris et al. (2000): "To begin to make a case against autonomous models on the issue of sensitivity, one would need to demonstrate that lexical information could actually improve phoneme discriminability," (p. 303). Unfortunately, signal detection approaches and direct phonemic judgments have been considered inadequate for this purpose (Norris et al., 2000), which is why some researchers have turned to indirect measures of phonemic sensitivity in phonemic adaptation and restoration studies (Samuel, 2001; Samuel & Pitt, 2003).

The McGurk illusion may provide another means for addressing these issues. The ambiguous phonemes used in the previous studies to demonstrate lexical–semantic influences on phoneme identification were unimodal stimuli made ambiguous by adding noise, manipulating voice onset time, cross-splicing segments, or in other ways manipulating the acoustic features. In contrast, the McGurk illusion is not ambiguous in this sense. In a McGurk stimulus, both auditory and visual signals can be perfectly clear (in fact they *should* be clear to maximize the illusory effect, see MacDonald, Andersen, & Bachman, 2000). However, the stimulus is nevertheless deceptive as it presents a *conflict* to the observer (with regards to the difference between conflict and ambiguity, see Massaro & Cohen, 1983b). This conflict is then resolved as the two inputs are merged into a novel, unitary representation.

Given these physical differences, is it possible that the unitary perception of the McGurk illusion, generated out of unambiguous auditory and visual speech signals, is more robust to lexical–semantic expectation and cognitive intervention than the inherently ambiguous phonemes examined in the previous studies? The purely feedforward nature of the McGurk theories clearly suggests this assumption. Naturally, these theories could be altered to account for top-down influences by moving phonemic decisions out of the speech processing hierarchy, as in autonomous process models of speech perception (see the discussion of Marslen-Wilson, 2000).

However, the question remains whether this move is really necessary in the case of the McGurk illusion. The subjective effect of lipreading is that it enhances the loudness and clarity of speech perception, apparently by changing activation patterns in the superior temporal sulcus (Calvert, 2001; Calvert et al., 1998; Summerfield, 1992; Wright et al., 2003). Lipreading can therefore be used to *disambiguate* poor auditory speech cues. By integrating lipreading with auditory perception, the McGurk illusion might result in a representation that is more potent, more coherent or more efficiently coded than unimodal ambiguous phoneme perceptions. Its bottom-up component might be so strong that it overrides all top-down influences that would normally interact with speech perception signals. If this were the case, a pure bottom-up description of the McGurk effect would indeed be sufficient; feedback connections or other integrative mechanisms would not be required. This result would confirm previous notions of the autonomy and invariability of the McGurk effect (Calvert et al., 1997; Dekle et al., 1992; Fowler & Dekle, 1991; Green et al., 1991; Langenmayr, 1997; Sams et al., 1998).

On the other hand, higher cognitive processes may have an impact on the McGurk illusion if these processes are strong enough and the dependent variables (measuring the quality and strength of the illusionary experience) are sufficiently sensitive. This assumption seems more plausible from a neuroscience viewpoint: Considering the widespread bilateral connections between early sensory and higher cognitive cortex areas in the auditory domain and elsewhere in the brain, it may seem unlikely that any perceptual phenomenon can be entirely autonomous and cognitively inaccessible (Felleman & Van Essen, 1991; Feng & Ratnam, 2000; Friston, 2002; Pandya, 1995).

To address these issues, the present study presented the McGurk illusion in an unexpected versus expected context during sentence processing and during a nonverbal working memory task, similar to what has been done in the previous studies with ambiguous phonemes. Crucially, these procedures manipulated only subject-specific cognitive variables (Engel et al., 2001), without ever changing the bottom-up input signal (i.e., the auditory or the visual component of the McGurk stimuli, as, e.g., in MacDonald et al., 2000), the immediate sensory context (as in Green & Gerdeman, 1995; Hietanen, Manninen, Sams, & Surakka, 2001; Walker et al., 1995), or the attentional weighting of the auditory vs. visual input signals (as in Massaro & Warner, 1977; Summerfield & McGrath, 1984).

If the McGurk illusion resists these manipulations, this would suggest that the illusion is indeed an essentially autonomous and invariable phenomenon whose bottom-up component is so strong that it overrides lexical–semantic context and biasing influences of cognitive control. If, however, it proves to be susceptible to these manipulations, then this would suggest that the McGurk illusion is treated by the brain no different (at least not fundamentally) than noisy or otherwise acoustically ambiguous phonemes.

In the latter case, it will be crucial to see in what direction the effects of the cognitive manipulation go. No data currently exist on this question since previous attempts to influence the McGurk illusion by cognitive set have been largely unsuccessful, and the theories of the McGurk illusion do not make any specific predictions regarding this question. At the same time, this is exactly the point at which the results may have important implications for the interactive-autonomous debate in speech perception research.

There are two possibilities: The first is that the illusion will tend to be *destroyed* in an expected relative to an unexpected context. When subjects focus on a particular, upcoming stimulus (as opposed to another stimulus), this might help them to *accurately encode* this stimulus perceptually (cf. Samuel, 1981, 1991). As a result, they might be more likely to detect the true nature of this stimulus, including the inherent audiovisual conflict in the case of the McGurk illusion. Consequently, instead of rendering the typical fusion response (/k/), they may be more likely to give the auditory response (/p/), the visual response (/t/), or a combination of the two (/pt/). In other words, subjects might be "disillusioned" more easily under conditions of focused expectations, causing them to give a lexically illegitimate, but phonetically correct response. This would demonstrate effects of (cognitive, semantic) expectation on the *sensitivity* of phonemic perception (as opposed to decision bias). As Norris et al. (2000) and other authors have emphasized (Connine, 1987; Connine & Clifton, 1987; Samuel, 1981, 1997, 2001), such an observation would speak for interactive models of speech processing because it shows that information from semantic levels is propagated backwards via feedback connections to *improve* perceptual acuity at lower auditory levels.

At the same time, this result would be in conflict with the essentially autonomous nature of the current McGurk theories which propose that the audiovisual integration process occurs solely in a feedforward manner. If semantic and other higher cognitive levels can be shown to destroy the audiovisual integration process, thereby enabling subjects to correctly identify the actual sensory input, then audiovisual integration cannot be entirely data-driven.

The second possibility is that the illusion will become *stronger* in expected versus unexpected conditions. In this case, subjects would report the illusory fusion response more often when the illusion confirmed their expectations rather than not. Such a finding would suggest that the processing of the *combined* audiovisual information, and not the distinct sensory components, benefits from enhanced expectation. While this outcome would not be able to decide between autonomous and interactive theories of speech perception, it would nevertheless show that the McGurk illusion is subject to cognitive interpretations, similar to ambiguous phonemes. At which level these interpretations impact on the illusion (perception or decision levels) would then have to be decided on the basis of some other means.

Three experiments were carried out in the present study. In Experiments 1 and 2, subjects' expectations were manipulated by varying the semantic congruency of sentences providing the context for words containing the McGurk illusion. Thus, the McGurk illusion occurred either in a semantically congruent or incongruent context (while the immediate lexical context was the same). In Experiment 3, subjects called on their working memory to actively hold an orthographically presented nonsense syllable in their short term memory for later comparison with a stimulus containing the McGurk illusion. This syllable either matched the illusion or not. In all three experiments, not only was the probability of the McGurk illusion assessed (as reported by the subjects), but also its strength by having subjects rate the "clarity" of their perceptions on a 7-point scale. Presumably, this continuous measure would be more sensitive to any potential effects of expectation than the categorical decisions typically taken in studies on the McGurk effect. In addition, the inclusion of the rating measure allowed us to determine whether the experimental manipulations act on continuous/fuzzy information rather than on stochastic categorical information (Massaro & Cohen, 1983a, 1983b, 1995).

Experiment 1

In this experiment, the McGurk illusion was embedded in two-syllable words so that the typical fusion response according to MacDonald and McGurk (1978) yields the lexically correct word (e.g., auditory /zuper/ and visual /zuter/ should yield the German word /zuker/ = ZUCKER).[2]

In each stimulus, the target consonant (k, g, n, or d) was flanked by two vowels to result in relatively clear McGurk effects. Examples are given in Appendices A and B.

These "McGurk-words" were presented either in the context of a semantically congruent or incongruent sentence. The sentences were quite elaborate and semantically highly constrained as most of them made reference to German sayings and idioms. The probability of the McGurk illusion and its strength (i.e., its "clarity") was assessed. As most McGurk illusions are typically perceived with a probability of about .50–.70 (MacDonald & McGurk, 1978), the responses of the subjects had room to move into both directions; i.e., they could increase or decrease as a function of semantic expectation.

In addition to semantic congruency, the strength of subject's expectations was varied. In one condition, incomplete sentences were used for which the word containing the illusion provided a syntactically and semantically correct ending. For example, the sentence: "I prefer my coffee with milk and..." will make most subjects predict the word 'sugar' (= ZUCKER). This condition was meant to induce highly focused expectations with regards to the sentence-final words. It will be called the "Prediction" condition throughout this article.

In another condition, complete sentences were presented. These sentences could also be either semantically related or unrelated to the McGurk-words. A semantically congruous example of this condition is the sentence: "For making a cake, one needs wheat, butter, milk, and eggs." This sentence primes the word 'sugar' (among other words, e.g., 'salt'), but since the sentence is already complete, the word 'sugar' will not be predicted to the same degree as in the Prediction condition. This condition will be called the "Priming" condition throughout this article. Presumably, these "priming" sentences induce more diffuse semantic network effects relative to the incomplete sentences in the Prediction condition.

Both conditions (Prediction and Priming) were performed with semantically congruent vs. incongruent McGurk-words. Thus the study had a 2×2 factorial design with the two repeated measures Condition (Priming and Prediction) and Congruency (congruent vs. incongruent). Regarding the interaction, the hypothesis was that the effects of semantic congruency on the McGurk illusion would be stronger in the Prediction condition than in the Priming condition because the semantic expectations in the Prediction condition were relatively more focused. Alternatively, if audiovisual integration is essentially autonomous and bottom-up driven, the McGurk illusion should be independent of semantic congruency and predictability in both conditions, Prediction and Priming.

Methods

Subjects

Thirty-three healthy subjects with a mean age of 24.7 years (range 18–46) participated in this study; 23 were female and 10 male. Thirty (90%) of the participants

[2] The McGurk illusion is defined as a change of auditory perception through visual influences. The example most often given is the perception of /d/ in response to combined auditory /g/ and visual /b/, but MacDonald and McGurk (1978) described many others. In this study I used the ten different consonant combinations for which the prominent fusion response was to be expected in at least 50% of the cases according to MacDonald and McGurk (1978). Among them was the example given here (auditory /p/ and visual /t/ yielding /k/), for other examples see Appendices A and B.

were undergraduate students of Psychology who received course credit for participation and the remaining three were University personnel who were not familiar with the aims of the study.

Materials

Twenty bisyllabic words were chosen containing two vowels connected by a consonant that served as the target for the McGurk manipulations. These "McGurk-words" are shown in Appendix A. They provided the final words to 20 highly constrained incomplete sentences used in the Prediction condition, and were semantically related to 20 complete sentences used in the Priming condition. In addition, 10 highly constrained incomplete and 10 semantically coherent complete sentences were constructed that semantically centered around different words (i.e., words not contained in the stimulus set). These were presented in the semantically incongruent condition.

The same 10 consonant pairs were used to generate the McGurk words in the congruent vs. the incongruent conditions (e.g., GLOCKE vs. ZUCKER; LÖHNE vs. SAHNE). This was done to balance out any potential differences between the various McGurk illusions across the experimental conditions. These 20 McGurk words were repeated once in the Prediction and Priming conditions, but these repetitions were cross-balanced across conditions and subjects. That is, for example, ZUCKER was presented in a congruent context in the Priming condition and was repeated in the incongruent context in the Prediction condition for half of the subjects ("Form A" in Appendices A and B), but the reverse was true for the other half of the subjects ("Form B"). Note that this procedure worked against the hypothesis of significant differences: If subjects simply repeated their responses each time the McGurk stimuli was presented, then this would decrease the differences between the experimental conditions. This would pose a problem for the interpretation of null results, but not for the interpretation of significant differences. Note in particular that within the Prediction condition, each McGurk word occurred *only once*.

Thus, each subject was presented with 10 sentences plus the McGurk-words in both tasks and conditions, resulting in a total of 40 trials. Furthermore, four control trials were added, two of which contained "inverse" McGurk stimuli (in which the auditory and the visual consonants of the McGurk illusion were interchanged so that fusion is unlikely) while the other two were undubbed videos. This was done solely to control for subjects' maintaining attention and will not be considered further.

To create the McGurk stimuli, a female speaker (S.W.) producing both the auditory and the visual stimuli in front of a plain white background was filmed with a high resolution digital video camera. This film was later cut into segments of 3 s duration. The audio tracks of these segments (recorded with a sampling rate of 44.1 kHz) were then dubbed onto the video tracks using the software Adobe Premiere. In this procedure, the original speech waveform served as a visual aid to ensure proper synchronization. The new segments were cut once more to align the edges properly. The resulting clips of approximately 2.8 s duration were saved and exported into Motion Pictures Expert Group (MPEG) format with a size of 352 × 288 pixels, a bit rate of 1,100,000 per second, and a frame rate of 25 frames per second. For presentation, the videos were enlarged to fit the entire 14" TFT display so that the mouth had a horizontal extension of about 3–4 cm. Only the lower part of the face was visible because a black mask was used to cover the upper half of the screen and the edges, leaving a window of approximately 26 × 9 cm. This was done to prevent subjects from looking at the speaker's eyes or elsewhere other than the lips (cf. Summerfield, 1979). The written instructions, the sentences, and the typed-in responses were also presented in this window. The auditory stimuli were played with a loudness of ca. 64 dB via two loudspeakers placed at a distance of approximately 70 cm from the subject.

Procedures

Subjects were tested individually in a light- and sound-attenuated chamber. They were seated in front of a laptop computer at a comfortable distance of approximately 50 cm from the screen. They were then told the cover story: They were asked to imagine that they did an internship in a film studio, where their task was to catalogue a number of videos showing a female person speaking two-syllable words. They were told that these videos had originally presented meaningful words, but that many of them were damaged or not improperly synchronized so that their quality was often poor. Specifically, they were told that the spoken words might be phonetically unclear, syntactically incorrect, or pronounced badly; sound and picture might be poorly synchronized, the words might not match the context in which they appeared, and any mixture of all these defects might occur. Their task was to rate the quality and the contents of the stimuli.

Due to the probabilistic nature of the McGurk-effect, evaluating the quality of the videos seemed natural to the subjects. Even reporting the words exactly the way they had understood them (including their flaws) presented no problem. In fact, it turned out that only about 50% of the words were evaluated as being correctly pronounced.

On each trial, subjects first read aloud the context sentence presented to them on the computer screen and pressed the space bar when ready. They were then presented the video with the McGurk-word. They were asked to exactly repeat what they had understood and to type this in, even if it was a nonword. If they were unsure, they were encouraged to type in whatever came next to what they had understood. Thereafter, they were asked to rate the “clarity” of the spoken word on a 6-point rating scale; i.e., they indicated how close the spoken word was to its correct pronunciation (the target word was written on the screen in case it had not been identified; e.g., ZUCKER).

In these instructions, no specific reference was made to either the auditory or the visual modality. Subjects were simply asked to indicate what they had understood, but they were warned explicitly that nonlexical items would appear on the majority of trials. Finally, subjects were asked to indicate on a 6-point rating scale (from 1 to 6) how well the spoken word matched the sentence context in which it had been presented. This latter rating served to assess the effectiveness of the semantic congruency manipulation. It proved highly successful as the statistical comparison of the ratings in the congruent vs. incongruent condition showed ($F(1,32) = 2069.24$, $p < .001$, $\eta^2 = .985$). This was more true for the prediction condition (5.95 vs. 1.44) than for the priming condition (5.573 vs. 1.71; interaction $F(1,32) = 33.04$, $p < .001$, $\eta^2 = .51$). Cloze probabilities or other measures of sentence constraint were not obtained (but see Experiment 2).

All responses were given via the computer keyboard. Subjects received five practice trials prior to the actual experimental session.

Data analysis

The probability of the McGurk illusion was determined by counting the relative number of times that subjects gave the typical fusion response for the critical McGurk-phoneme. Deviations from the lexically correct illusory response which occurred in about 10% of all cases (for example, “HAGE” instead of “LAGE”) were ignored as only the critical McGurk-phonemes were relevant (i.e., whether visual /b/ and auditory /k/ were combined into /g/). Excluding these nonlexical responses from the analysis did not change the results pattern. However, it should be noted that subjects gave nonword responses in more than 50% of all trials (677 of 1320), indicating that they had complied with the instructions to precisely report what they had understood, even if it was a nonword.

A 2 × 2 analysis of variance with repeated measures was performed for both measures, response probability and clarity rating. *Partial* η^2 is reported as a measure of effect size, reflecting the proportion of the variability in the dependent variable that is explained by the independent variable (when the variability of all other factors are partialled out). For moderate effect sizes, η^2 is approximately .10 according to Cohen (1988).

Results

Fig. 1A shows the results in terms of the probability of the typical illusory fusion response. Visual and auditory responses are not depicted. These responses were rendered in less than 8.5% in all conditions, and did not vary significantly as a function of semantic congruency or prediction/priming (all $p > .20$). Combinatorial responses were not observed.

On average, the McGurk illusion occurred in about 60% of all trials as in the original studies (MacDonald & McGurk, 1978, left quadrant of Table 1). No main effects of condition (Prediction vs. Priming) were observed. However, there was a strong main effect of semantic congruency; $F(1,32) = 10.43$, $p < .001$, $\eta^2 = .25$; indicating that the illusion occurred more often when the words were congruent as opposed to incongruent. Fig. 1A suggests that this difference is somewhat more pronounced in the Prediction condition (where it was in fact significant in a paired t-test; $t(32) = 2.50$, $p < .005$) than in the Priming condition (where it was not significant; $t(32) = 0.758$). Nonetheless, this pattern did not result in a statistically significant interaction of Congruency × Condition.

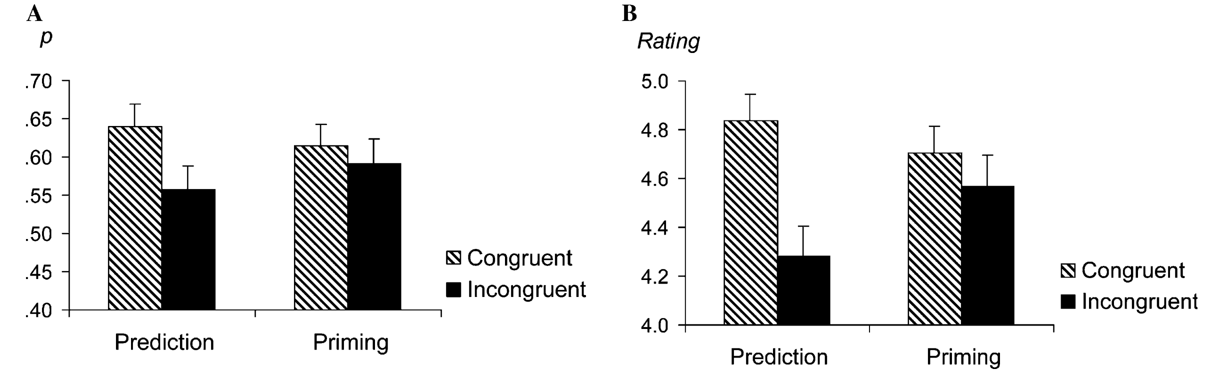

Fig. 1. (A) Probability and (B) rated clarity of the McGurk illusion presented in semantically congruent vs. incongruent lexical–semantic contexts in the predicion and priming conditions (Experiment 1).

Subjects gave "atypical" fusion responses (that is, responses that reflected neither the auditory/visual signal nor the illusion) in approximately 20-30% of the trials. These percentages varied as a function of semantic congruency and condition; they were 20.9% vs. 31.2% in the congruent vs. incongruent Prediction condition, and 26.4% vs. 27.4% in the congruent vs. incongruent Priming condition, respectively. These proportions showed a significant effect of congruency ($F(1,32) = 15.89$, $p < .001$, $\eta^2 = .33$); an effect that was greater in the Prediction condition than in the Priming condition; $F(1,32) = 4.75$, $p < .05$, $\eta^2 = .13$. Post hoc tests revealed that the congruency effect was significant in the Prediction condition ($t(32) = 4.15$, $p < .001$) but not in the Priming condition ($t(32) = 0.34$).

Taken together, these results suggest that in the Prediction condition, semantically incongruent as compared to congruent contexts made subjects' reports shift from the typical illusory fusion response to "atypical" fusion responses. This means that the illusory experience was lost on a significant proportion of trials when the context was incongruent. By contrast, *accurate* responses reflecting the true nature of the visual and auditory signals were largely unaffected by semantic expectations.

Fig. 1B shows the results for the rating variable. Pronunciation of the McGurk words was rated as clearer for a semantically congruent as compared to incongruent sentence context; $F(1,32) = 70.33$, $p < .001$, $\eta^2 = .69$. This difference was larger in the Prediction condition than in the Priming condition as revealed by a significant interaction of Congruency × Condition; $F(1,32) = 4.48$, $p < .05$, $\eta^2 = .123$. When only the Prediction condition was considered, the effect of semantic congruency was significant; $t(32) = 5.34$, $p < .001$. This was not true for the priming condition ($t = 1.24$).

Discussion

In summary, semantic congruency had a significant impact on the McGurk illusion. Across both conditions (Prediction and Priming), the McGurk illusion was experienced more often and was rated as clearer in the semantically congruent condition relative to the incongruent condition. This tendency seemed somewhat stronger for the Prediction condition than for the Priming condition.

Taken together, these results suggest that the McGurk illusion is not entirely autonomous. Rather, it can be significantly influenced by higher cognitive variables such as semantic expectation. Apparently, subjects were more able to fuse and further process the auditory and visual signals into a unitary percept when it corresponded with their expectations than when it was incongruent with their expectations. By contrast, when the McGurk-words came by surprise, the illusion was destroyed on a significant proportion of trials.

This effect seemed slightly larger in the Prediction condition than in the Priming condition, although not significantly. It is therefore somewhat unclear whether these conditions represent different parameter values of the same dimension (expectedness) or qualitatively different processes (a priori prediction vs. a posteriori integration). Experiment 2 was carried out to clarify this question.

Experiment 2

This experiment was carried out to replicate and extend the results of Experiment 1. I wanted to find out what had caused the effects in the Prediction condition to be so strong. On the one hand, they could be due to highly focused semantic expectations invoked by the incomplete sentences as had been hypothesized before. On the other hand, *syntactic* violations may have contributed to the size of the effects because the sentence-final words in the incongruent condition were selected simply by chance out of the pool of McGurk-words. This resulted in incongruent sentences such as: "On an orbit in space you find Mars, the Earth and every other TEETH." There were multiple violations of German grammar in these sentences, especially with regards to gender. This may have caused subjects to be highly surprised, beyond any semantic expectancy violations.

Therefore, I tried to replicate the effects found for the Prediction condition in Experiment 1 with new incomplete sentences to which the McGurk-words provided syntactically correct endings in both the congruent and the incongruent conditions. The McGurk words were now possible endings in both conditions, but they were far less plausible in the incongruent than in the congruent condition (see Appendix B). To quantify this difference, a measure of semantic constrainedness was obtained: Subjects were asked to complete the sentences presented to them by naming the sentence final word they found most appropriate (before they were presented with the McGurk-words). The probability of the most frequently given word (= cloze probability, Taylor, 1953) was then computed for each sentence in the congruent vs. incongruent conditions.

Methods

Subjects

Twenty-five healthy subjects participated in this study (mean age 27.0 years, range 20–47). Eighteeen were undergraduates of Psychology, seven were graduate students or University staff members who were blind to the goals of the study. Seven participants were male, 18 were female. All student participants received course credits for participation. None of the subjects had participated in Experiment 1.

Materials

The same McGurk-words were used as in Experiment 1 and the same incomplete sentences in the congruent condition. New sentences were formulated for the incongruent condition. These matched the McGurk-words syntactically and to some extent also semantically, although the McGurk words were still very unusual endings to the sentences.

Cloze probability of the most frequently chosen sentence-final word was .90 ($SD = .15$) in the high-constraint-condition. This shows that the sentences in this condition were in fact highly constrained. In the low-constraint condition, average cloze probability of the most frequently named final word was only .45 ($SD = .29$). This difference was statistically significant ($t(38) = 6.01$, $p < .001$). Note, however, that the predictability of the McGurk words actually presented differed even more because the most frequently named word was always presented in the congruent condition, but never in the incongruent condition (i.e., cloze probability of the McGurk words was 0 in the incongruent condition).

Procedures and data analyses

Procedures were the same as in the case of the incomplete sentences in Experiment 1 with one exception: Before subjects were shown the McGurk videos, they were asked to name the word that they thought would complete the sentence best. Furthermore, each McGurk word was shown only once, either in the congruent or in the incongruent condition, counterbalanced across subjects.

As in Experiment 1, word responses were given in less than 50% of the trials (238 of 500) indicating that subjects understood the task and were willing to yield perceptually accurate responses, even if these were lexically illegitimate. For the following analyses, like in Experiment 1, only the critical phonemes are considered, whether they occurred in a word or in a nonword response. Results were not affected by this procedure.

Again I analyzed (1) the probability of experiencing the illusion and (2) the "clarity" ratings. Auditory and visual responses occurred in less than 11 percent; auditory responses were given significantly more often ($F(1,24) = 4.90$, $p < .05$) than visual responses, but there were no significant differences between the conditions (all $p > .20$).

Results

The typical fusion response was again given in approximately 60% of all trials. It occurred slightly more often when the sentence-final words were expected as opposed to unexpected (see Fig. 2A), but this difference did not reach statistical significance ($F(1,24) = 2.02$). Likewise, the "atypical" response category showed no significant difference between the two conditions (in contrast to Experiment 1).

However, the rating measure did indicate a significant difference between the two experimental conditions: Pronunciation of the McGurk-words was rated as clearer (i.e., closer to correct) in the congruent context as compared to the incongruent context; $F(1,24) = 7.01$, $p < .05$, $\eta^2 = .226$ (see Fig. 2B).

Additional analysis

When data from Experiment 1 (Prediction condition) and Experiment 2 were combined, the effect of semantic congruency on the probability of the McGurk illusion was clearly significant; $F(1,56) = 6.53$, $p < .01$, $\eta^2 = .105$, with virtually no differences between the two experiments (congruency × experiment interaction: $F(1,56) = 0.77$). This suggests that the differential results pattern found in the separate analyses is of quantitative rather than of qualitative nature.

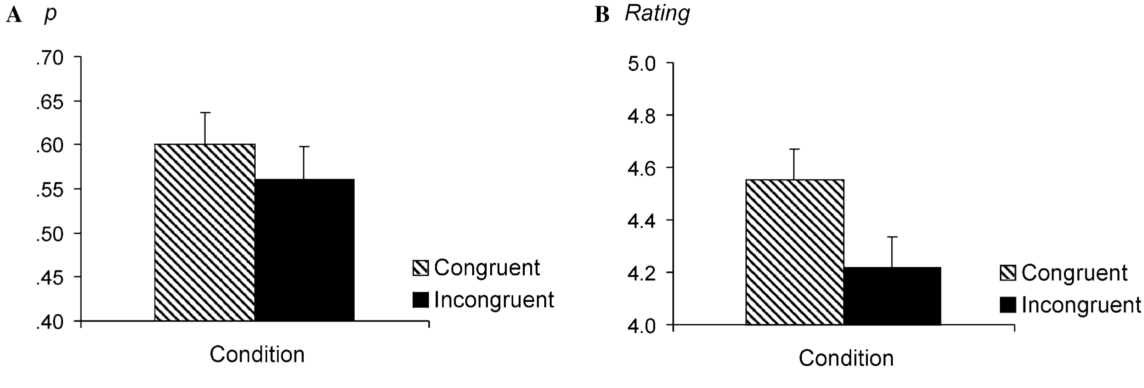

Fig. 2. (A) Probability and (B) rated clarity of the McGurk illusion presented in semantically congruent vs. incongruent lexical–semantic contexts in Experiment 2.

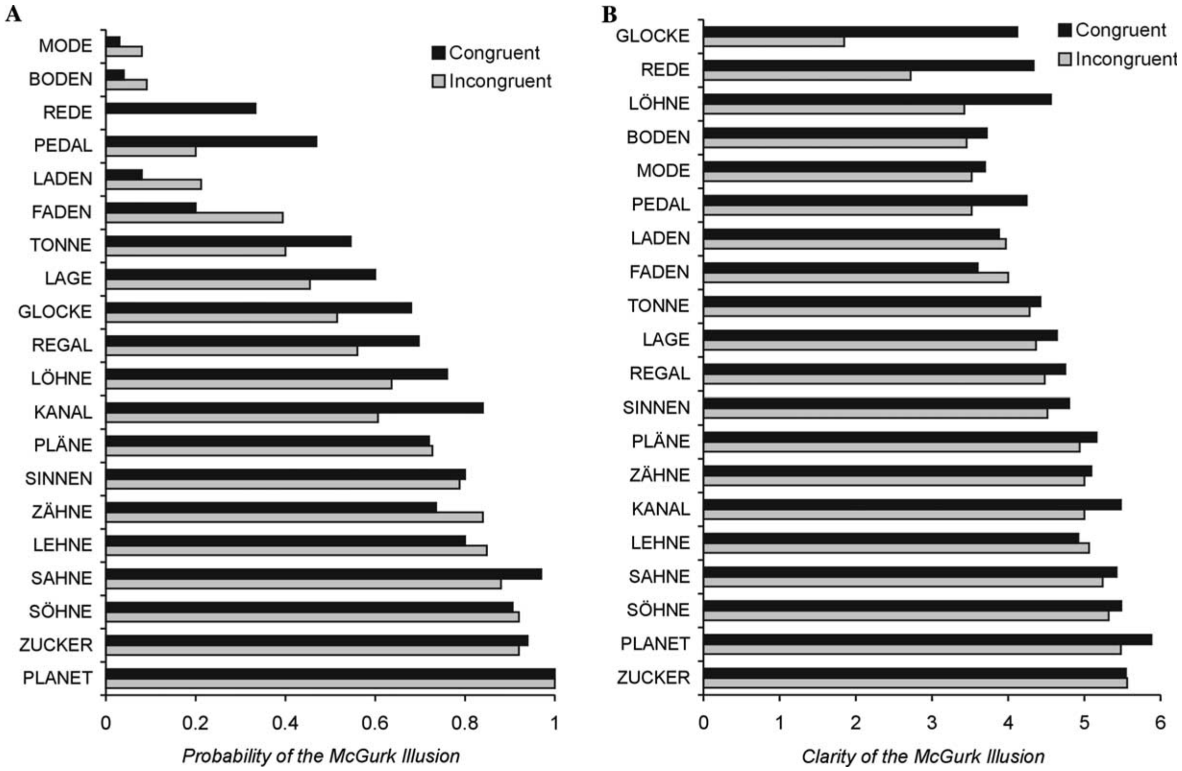

Fig. 3. (A) Probability and (B) clarity rating for the 20 different “McGurk-words” presented as semantically congruent vs. incongruent sentence-final words; data from Experiments 1 and 2 combined.

Fig. 3A shows for the combined data set how the probability of the illusion varied across the 20 McGurk words as a function of semantic congruency. Although there was considerable variability in the sensitivity of the items to the congruency manipulation (yielding a non-significant $F(1,19) = 1.57$, $p = .22$, for the $N = 20$ items available), the effect does not seem to be restricted to only a few items.

Fig. 3B shows the same item analysis for the rating measure. The congruency effects in this measure seemed relatively small, but were very consistent across the 20 items and therefore highly significant; $F(1,19) = 15.162$, $p < .001$, $\eta^2 = .45$. They were somewhat larger for items that received ratings below 3 in the unexpected condition, maybe due to a possible ceiling effect in the other cases in which relatively high ratings were given in both conditions.

Discussion

This experiment showed that semantic factors alone are sufficient to influence the McGurk illusion, and that any additional syntactic or other more fundamental violations of sentence discourse are not necessary. The sentence final words in this experiment were both syntactically correct and semantically meaningful, though much more predictable in the expected condition than in the unexpected condition. This difference alone sufficed to significantly enhance the clarity ratings for the McGurk illusion, and to increase the probability of the illusion by almost 10% (although this increase did not reach significance).

This pattern of findings is novel as the McGurk effect has never before been shown to be influenced by semantic levels of analysis. Although several other studies have successfully presented the McGurk illusion in lexical contexts (Dekle et al., 1992; Sams et al., 1998; but see Easton & Basala, 1982; and the discussion in Massaro, 1987, p. 49 f.), there was only one attempt (to my knowledge) to subsequently present these words in the context of sentences to prompt semantic processing. Sams et al. (1998) presented words and nonwords containing the McGurk illusion either as the first or the last word of incomplete sentences in a study with a rather complex design. The authors reported that the McGurk effect was so strong that it often destroyed the lexical–semantic analysis of the word or the sentence in which it appeared. Hence, they concluded that the illusion was independent of lexical–semantic context. It should be noted, however, that the experimental manipulations in this study may not have been strong enough to produce the desired effects. For example, the authors used somewhat ambiguous very short sentences which may have resulted in much less focused predictions than in the present study (unfortunately, the authors give only the following example in the text: "In tennis, we need a ..."). Second, the combination of auditory /pa/ and visual /ka/ was used on all trials of the Sams et al. study, and was expected to result in the illusion of /ta/ or /ka/. This is somewhat surprising because MacDonald and McGurk (1978) reported /pa/ to be the most likely response to this combination (70%), followed by /ta/ (10%), /ka/ (10%) and tha (10%). Finally, the authors took only categorical decisions where rating measures might have been more sensitive as suggested by the present results. Hence the discrepancies between the present results and the ones reported by Sams et al. (1998) are probably due to methodological differences.

The results of the present experiments resemble more the many previous observations made with intrinsically ambiguous phonemes (Connine, Blasko, & Wang, 1994; Elman & McClelland, 1988; Ganong, 1980; Gaskell & Marslen-Wilson, 2001; Newman et al., 1997; Samuel, 1981, 1997, 2001; Samuel & Pitt, 2003). These studies have shown that the identification of ambiguous speech segments is influenced by phonemic, lexical, and semantic context, as shown in the present study for the McGurk illusion. Whether these variables exert their influence on perception, interpretation, or decision processes is still a matter of debate (Norris et al., 2000), although recent reports suggest that perceptual effects can be involved (Samuel, 2001; Samuel & Pitt, 2003).

In his earlier investigations on phonemic restoration, Samuel (1981) used the signal detection approach to show lexical effects on phonetic sensitivity. Insofar as McGurk phonemes are processed like the inherently ambiguous phonemes used in these studies, his reports may aid in interpreting the present results pattern. Phonemic restoration refers to the fact that subjects tend to "fill in" missing or noisy phonemes in spoken words, such that they compensate for the defective phoneme in favor of an intact word perception (Warren, 1970; Warren & Sherman, 1974). By manipulating the context of these processes, Samuel (1981) found an interesting dissociation between measures of bias and perceptual accuracy: While sentence context increased the *bias* of his subjects to restore the deficient word stimuli (Experiment 3 in Samuel, 1981), lexical–syntactic factors reduced the perceptual *accuracy* with which subjects processed these words (they had to differentiate between two variants in which noise either *overlaid* a critical phoneme or *replaced* it). Similar effects of semantic meaning on lexical decision bias, but not on lexical discrimination performance, have been reported for affective stimuli presented visually (Windmann, Daum, & Güntürkün, 2002a; Windmann & Krüger, 1998; Windmann & Kutas, 2001).[3]

In these experiments, a similar disconnect was observed between the rating measure on the one hand and the probability measure on the other. In Experiment 2 as well as in the Priming condition of Experiment 1 (where semantic congruency was varied outside of syntactical factors), subjects primarily changed their *ratings* in accordance with their semantic expectations. In light of Samuel's findings, this might indicate that subjects had *biased* their phonemic interpretations depending on the semantic context in these conditions. By contrast, subjects showed altered category decisions – in addition to the ratings – in the Prediction condition of Experiment 1 (where sentence syntax was violated in addition to semantic congruency). Perhaps this condition had changed phonemic *perceptions* in addition to interpretations/decision biases. Although some of the differences between Experiments 1 and 2 seem to be of quantitative rather than of qualitative nature, this interpretation is consistent with some authors' views (e.g., McQueen, 1991) that semantic processes act at a postperceptual guessing stage (by biasing phonemic decisions) whereas syntactic factors might influence earlier, perceptual processes, perhaps because they are more specific and more focused and therefore stronger.

To test this possibility, a third experiment was devised with nonlexical stimuli. This experiment tested whether non-semantic violations of predictions that specifically focus on phonetic features as opposed to lexical–semantic meaning can significantly and qualitatively influence the classifications of the illusory McGurk phonemes. The hypothesis was that this manipulation might bring out clearer changes in the probability measure than the semantic manipulations in Experiment 2.

Experiment 3

Three-letter nonsense bisyllable tokens (vowel-consonant-vowel) were used as McGurk-stimuli in this experiment. Subjects' phoneme expectations were manipulated by informing them that they would be presented with [1] the stimulus they were expected to hear if the McGurk illusion was perceived (this will be referred to as the "illusion expected" condition), or [2] the auditory signal ("auditory expected"), or [3] the syllable mouthed by the speaker ("visual expected").

[3] In fact this latter study showed that for items with an emotional connotation, semantic information affected both, lexical decision accuracy and bias in the right hemisphere, while only the bias was affected in the left hemisphere.

In all these conditions (except in [4], a control condition, see below), subjects were in fact presented the McGurk stimulus.

This procedure is similar to the "nonword priming" experiment by Samuel (1981, Experiment 2) and follows a similar rationale. By manipulating subjects' expectations, the different modalities (auditory, visual, integrative) stimulated by the McGurk illusion were "primed." If this priming facilitated *accurate* perceptual encoding (due to feedback connections), then subjects should give more auditory and visual responses in conditions [2] and [3], respectively, relative to condition [1] in which the illusion was primed. If, however, the "priming" established a nonspecific *bias* towards the expected cue, then subjects' responses should be affected equally in all three conditions.

As in the other two experiments, subjects were told a cover story beforehand. They were asked to imagine that they did an internship in a film study requiring them to evaluate short film clips relative to a reference cue which would be indicated to them prior to each trial. They were presented this "priming cue" orthographically on the computer screen. It informed them about what the speaker in the to-be-evaluated film was supposed to say. This cue was varied on a trial-to-trial basis according to the experimental conditions [1] through [3]. Subjects were then presented the McGurk-illusion. They were asked to compare this video with the orthographic cue they had been presented before. They were asked to give a categorical evaluation ("What did you understand? Select from the following four choices"), and to rate how clearly the syllable was pronounced on the video using a scale from 0 to 6 (e.g., "How close was this to the syllable you had expected (ADA)?").

Thus, subjects had to keep a cue in mind over a short delay and then compare a target stimulus with this cue, as in a standard working memory task. Numerous animal and human studies have provided insight into the neural mechanisms underlying this cognitive function (Elliott, Dolan, & Frith, 2000; Fuster, 2000, 2001; Goldman-Rakic, 1995; Miller & Cohen, 2001; Rainer, Rao, & Miller, 1999). This research has shown that such a task involves wide-range interactions between higher attentional control centers (prefrontal cortex) and lower sensory areas, where neural activity is altered according to the subjects' expectations and attentional settings in a stimulus-specific manner (Corbetta & Shulman, 2002; Desimone, 1998; Desimone & Duncan, 1995; Frith & Dolan, 1997; Hopfinger, Buonocore, & Mangun, 2000; Miller & Cohen, 2001; Mottaghy, Gangitano, Krause, & Pascual-Leone, 2003; Mull & Seyal, 2001; Rees, Frackowiak, & Frith, 1997; Von Stein, Chiang, & König, 2000; for the auditory domain see, e.g., Feng & Ratnam, 2000; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001).

Methods

Subjects

Twenty-three subjects participated in this study (17 female); the majority of them were students who received course credit for participation; the others (6) were doctoral students and one technical assistant. Mean age was 26.1 years (range 18–47).

Materials

Using the techniques described in Experiment 1, eight vowel-consonant-vowel-type of McGurk-stimuli were created (IDI, ADA, AKA, INI, ANA, AGA, ODO, UNU). In addition, eight "inverse McGurk stimuli" were generated (stimuli in which the auditory and the visual consonant of the McGurk illusion are interchanged such that the auditory stimulus is presented visually and vice versa) for control purposes.

Procedures

Subjects were told the cover story as before. On each trial they were then asked to read out aloud the orthographic cue on the screen (e.g., ADA) and to keep it in mind for later comparison with what the speaker said on the video. After pressing the space bar the McGurk-stimulus was shown. Subsequently, subjects were asked (by a question written on the screen) what they had understood. They were presented four response alternatives lined up on the bottom of the screen from which they had to choose one (by clicking on it with the mouse cursor). These response alternatives represented the auditory syllable, the visual syllable, the McGurk stimulus, and a fourth stimulus that contained the same vowels but whose consonant was randomly chosen from all possible auditory, visual, and McGurk-stimuli. This "atypical" choice alternative was not expected to be chosen frequently, it was meant only to accentuate the differences between the four response alternatives to maintain (and control) subjects' continuous attention. The position of the four response alternatives on the screen varied randomly from trial to trial.

Finally, subjects were asked to rate on a 7-point scale how close this syllable was to the one they had expected (the cue). This expected syllable was written on the screen (in case it had been forgotten). Subjects gave all responses via the keyboard.

Data analysis

As in Experiments 1 and 2, the probability of the illusion and the "clarity rating" were analyzed for significant differences between the three experimental conditions (visual, auditory, and illusion) using MANOVA. Control trials (inverse McGurk effects) and the control responses were not included in the analyses. The control responses were given with a proportion of 7.1, 3.8, 3.3, and 7.1% in the "auditory expected," "visual expected," "illusion expected," and "control" conditions, respectively. These differences are not significant.

Results

Fig. 4A shows the probability of the McGurk illusion when subjects expected the illusory stimulus to occur as compared to the auditory or visual stimulus. This probability was about .50 with relatively little variation across the three conditions ($F(1, 22) = .352$, n.s.). However, the rating variable did yield a significant effect of condition; $F(1, 22) = 7.74$, $p < .05$, $\eta^2 = .42$. This effect resulted from the fact that the illusion was rated as strongest when it was expected (see Fig. 4B); it was significantly lower in the "auditory expected" condition ($F(1, 22) = 4.621, p < .05$) while there was no significant difference between the "auditory expected" and the "visual expected" conditions ($F(1, 22) = .653$, n.s.).

Discussion

On the one hand, the results of this study failed to show any significant influence of expectation on the probability of the illusion, consistent with previous studies (e.g., Summerfield & McGrath, 1984). Hence the hypothesis that violations of highly specific phonetic expectations would influence perceptual encoding of the McGurk stimuli more than lexical–semantic violations was not confirmed. In fact, the effects in the categorical measure were even *smaller* than in the two previous experiments. Although the experiments are not directly comparable, these results might indicate that lexical–semantic effects on the McGurk illusion are stronger than more general effects of cognitive set involving working memory. As audiovisual integration of speech cues is typically performed in the context of sentence discourse processing in natural situations, lexical–semantic levels of processing may have established special access to the neural mechanism underlying this phenomenon.

On the other hand, this experiment demonstrated once more that the McGurk illusion is not entirely robust against cognitive interventions because the rating measure did covary significantly with the experimental manipulations. When subjects' expectations matched the illusion (e.g., /aga/), they rated the syllable as significantly closer to their expectation than when they expected either the auditory (/aba/) or the visual component (/aka/) of the McGurk stimulus. This is remarkable because in the latter case, the expectation was *objectively* closer to what had actually been presented, but subjects nevertheless evaluated the illusion as closer.

General discussion

The three experiments reported in this article demonstrate effects of expectation and prediction on the McGurk illusion. The first two experiments showed effects of lexical–semantic expectation on the McGurk effect embedded in real words, and the third one showed effects of specific phonemic expectations on the illusion, outside of lexical–semantic factors. The latter effects were reflected only in the continuous rating measure evaluating the strength or intensity of the illusion, whereas the first experiment (and to some degree also the second) yielded additional evidence for categorical shifts in the identification of the critical consonants.

Overall, this evidence suggests that the McGurk illusion is not more autonomous or robust to higher cognitive interpretations than the various inherently ambiguous speech stimuli that have been used to investigate lexical effects on phonemic identification and restoration (Connine, 1987; Connine & Clifton, 1987; Lucas, 1999; McClelland & Elman, 1986; Newman et al., 1997; Norris et al., 2000; Pitt & Samuel, 1993; Samuel & Pitt, 2003; Samuel, 1981, 2001). This conclusion is inconsistent with previous descriptions emphasizing the autonomy and cognitive inaccessibility of the

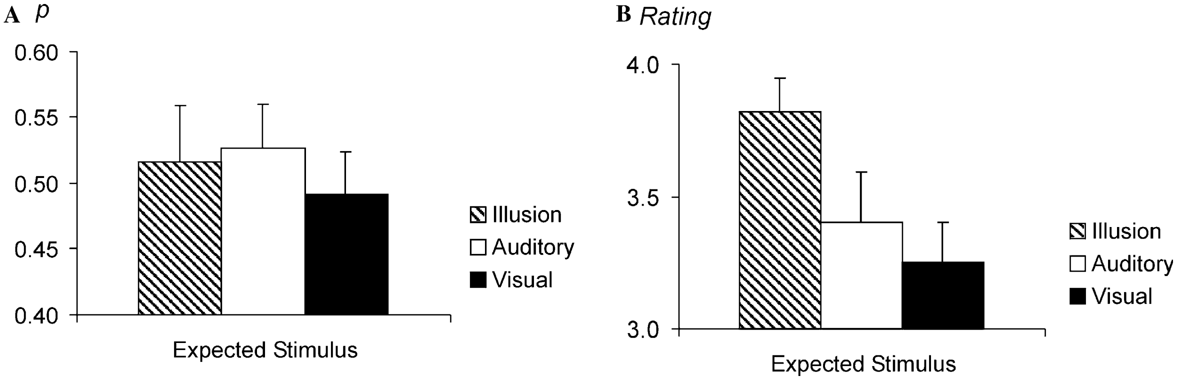

Fig. 4. (A) Probability and (B) rated clarity of the illusory McGurk-stimuli as a function of expectation (Experiment 3).

phenomenon (Dekle et al., 1992; Green et al., 1991; Langenmayr, 1997; Sams et al., 1998). Rather, it suggests that theories on the McGurk illusion must be integrative, as is true for speech perception models in general. That is, they must give higher cognitive processes access to representations of audiovisually integrated speech cues. Whereas the Fuzzy logical model of perception (FLMP) explicitly considers these interactions (e.g., Massaro, 1996; Massaro & Stork, 1998), these interactions as not well specified in amodal theories which attribute the McGurk effect to nonlinguistic modules in the brain, located, e.g., the insula (Bushara et al., 2003) or the claustrum (Olson et al., 2002). Moreover, the FLMP is also consistent with the fact that the effects were highly reliable in the rating measure reflecting a continuous measure of the strength of the illusion (Massaro & Cohen, 1983a, 1983b, 1995). This suggests that the processes affected by the expectancy manipulations encoded fuzzy information, not just distinct phoneme classes. From this it can be inferred that the McGurk illusion is not an all-or-none phenomenon with a probabilistic distribution but is subject to continuous variation, unlike many visual illusions whose impact cannot be countered by cognitive control (cf. Churchland & Churchland, 2002; Eagleman, 2001; Kolb & Braun, 1999). Thus, the McGurk illusion seems to be based on probabilistic, experience-dependent processes (Green, 1998), rather than being a universal, genetically determined function that automatically emerges from the hard-wiring of the system and therefore resists cognitive intervention.

What seems less clear from the present pattern of results is whether these top-down influences changed *perceptual processes*, as has been shown for ambiguous phonemes (Samuel, 2001; Samuel & Pitt, 2003) not just post-perceptual interpretations and decision biases. Such evidence would be of great theoretical importance because it would speak against autonomous process-models of speech perception (Norris, 1994; Norris et al., 2000) in favor of interactive or reentrant models (Elman, 1990; Gaskell & Marslen-Wilson, 1997; Grossberg & Stone, 1986; McClelland & Elman, 1986; Mercado III, Myers, & Gluck, 2002). On the one hand, some of the effects were not only gradual (as would be expected from a bias measure) but seem to have involved categorical changes in phoneme identifications as well. Subjects who reported the illusory fusion when it occurred in a semantically congruent context did not report it when it came by surprise on a significant number of trials in Experiment 1 (and to some degree also in Experiment 2). Such categorical effects are often attributed to perceptual factors (see, in particular, Connine & Clifton, 1987; Ganong, 1980).

On the other hand, these effects were relatively weak compared to the effects in the rating measure. The probability measure did not show any meaningful covariation with the expectancy manipulation in Experiment 3, was not significant in Experiment 2 alone, and did not prove to be very consistent in the item analysis of the data from Experiments 1 and 2 combined. Thus, quantitatively, the effects do not seem overly reliable, despite the fact that the expectation manipulation was relatively strong (as can be inferred from the differences in cloze probability in Experiment 2).

In addition, even if the effects in the probability measure had been stronger, it would still be unclear whether they are related to perceptual processes (as opposed to postperceptual/ postlexical processes) because of the direction in which these effects went. The congruent expectancy condition seems to have *decreased* rather than increased accurate phoneme identification (as increased accuracy would have *destroyed* the illusion rather than strengthened it, thereby prompting more auditory and visual responses, not more "atypical" fusion responses as observed in Experiment 1). This means that subjects' responses shifted *away* from perceptually accurate phonetic identifications in the congruent compared to the incongruent condition, towards a context-appropriate response. This finding is more consistent with a post-perceptual interpretation bias account than the inverse pattern would have been: Subjects may have reported the illusion in the expected context more often not because they have *perceived* it more often, but because it made more sense (semantically). Autonomous process models such as MERGE (Norris et al., 2000) would account for this behavior simply by inserting phonetic decision nodes after lexical–semantic processing levels.

To definitely answer the question of whether perceptual changes were affected by the current expectancy manipulations, it would be essential to repeat these experiments with event-related potentials (ERP) or magnetencephalography (MEG) to examine whether the relevant effects are early and sensory in nature or late and decision-related (cf. Mottonen et al., 2002; Sams et al., 1991; Windmann, Urbach, & Kutas, 2002b). In the former case, early components of the waveform should be influenced by the expectedness of the McGurk-words; in the latter case, later components reflecting higher cognitive analysis and semantic interpretation should be affected. Electrophysiological recordings in vivo are also feasible to directly observe the neuronal activity involved in these processes (Ghazanfar, personal communication, see Ghazanfar & Logothetis, 2003; Leopold & Logothetis, 1999).

If the results of such studies supported the view that category shifts in the evaluation of the McGurk consonants are based on perceptual, prelexical processes, perhaps at the level of A1, then this would mean that the audiovisual fusion process itself, not just its interpretation, can be facilitated or impeded by top-down processes. Accordingly, feedback projections carrying this influence would have to be added to process models of audiovisual integration. In addition, interactive models of speech perception would have to be modified so as to include mechanisms whereby phonetic *acuity/sensitivity* is increased, not just *biased* towards existing knowledge. This would discount the major criticism targeted at McClelland and Elman's (1986) interactive TRACE model (see Norris et al., 2000), and would complement recent studies pointing into this direction (Feng & Ratnam, 2000; Friston, 2002; Noesselt, Shah, & Jäncke, 2003; Pitt and Samuel, 2003; Samuel, 2001). On the basis of the present data, however, the question of whether or not feedback projections are involved in the cognitive modification of the McGurk illusion cannot be resolved.

In conclusion, the present study showed that the McGurk illusion is subject to the same cognitive modifications as heard speech cues. Apart from FLMP, theories of the McGurk illusion have developed somewhat independently from general models of higher language comprehension which may be a conceptual weakness according to the present results. In particular, amodal/motor accounts of the illusion seem to be inconsistent with the observation that the reports of the McGurk illusion depended on sentence context in a situation where phonetic, phonemic, and even syntactic factors at the lexical level (determining the proximate sensorimotor context of the illusion) were held constant (the same McGurk words were shown in the congruent and the incongruent condition). In addition, these accounts seem incompatible with the mounting evidence suggesting that primary auditory and extrastriate visual cortex seem to directly interact during audiovisual speech perception (Bavelier & Neville, 2002; Calvert, 2001; Calvert et al., 1998, 2001; Jones & Callan, 2003; Mottonen et al., 2002; Sams et al., 1991). Studies with high resolution functional magnetic imaging (e.g., Jäncke, Wüstenberg, Scheich, & Heinze, 2002) and intracranial recordings in monkeys (Ghazanfar & Logothetis, 2003) will be needed to further elucidate these neural mechanisms of the McGurk effect.

Acknowledgments

I am grateful to Onur Güntürkün for his support of this study. Many thanks also to Siggi Bach for programming the experimental procedures, and to Daniel Durstewitz and Anna Abraham for carefully reading and commenting on this article.

Appendix A. Sample stimuli used in Experiment 1

	Form A (congruent)			Form B (incongruent)		
	Illusion	Auditory	Visual	Illusion	Auditory	Visual
Incomplete sentences						
I prefer to take my coffee with milk and SUGAR	ZUCKER	ZUPPER	ZUTTER	GLOCKE	GLOPPE	GLOTTE
I put the book back on the SHELVES	REGAL	REBAL	REKAL	LAGE	LABE	LAKE
At our wedding my father gave a touching SPEECH	REDE	REBE	RETE	LADEN	LABEN	LAGEN
She was always dressed according to the newest FASHION	MODE	MOBE	MOGE	BODEN	BOBEN	BOTEN
With fruit cake I like to take a bit of whipped CREAM	SAHNE	SAHME	SAHDE	LEHNE	LEHME	LEHDE
The king left his empire to the eldest of his SONS	SÖHNE	SÖHME	SÖHKE	PLÄNE	PLÄME	PLÄKE
The huge dog growled and showed his TEETH	ZÄHNE	ZÄHME	ZÄHTE	LÖHNE	LÖHME	LÖTE
Uphill the biker stepped enormously into the PEDAL[a]	PEDAL	PEBAL	PENAL	FADEN	FABEN	FANEN
On an orbit in space you find Mars, the Earth and any other PLANET	PLANET	PLAMET	PLANET	KANAL	KAMAL	KANAL
Let's simply dump the stuff into this CONTAINER	TONNE	TOMME	TOGGE	SINNEN	SIMMEN	SIGGEN
Complete sentences						
On Sundays we went to church (BELL)	GLOCKE	GLOPPE	GLOTTE	ZUCKER	ZUPPER	ZUTTER
The difficult situation caused her much trouble (POSITION)	LAGE	LABE	LAKE	REGAL	REBAL	REKAL

Appendix A (*continued*)

	Form A (congruent)			Form B (incongruent)		
	Illusion	Auditory	Visual	Illusion	Auditory	Visual
For shopping they went into a mall (SHOP)	LADEN	LABEN	LAGEN	REDE	REBE	RETE
We put the carpet on the floor (GROUND)	BODEN	BOBEN	BOTEN	MODE	MOBE	MOGE
The couch in the living room is very comfortable (BACK)	LEHNE	LEHME	LEHDE	SAHNE	SAHME	SAHDE
I have many ideas with respect to my future (PLANS)	PLÄNE	PLÄME	PLÄKE	SÖHNE	SÖHME	SÖHKE
In many third world countries workers are being exploited (WAGES)	LÖHNE	LÖHME	LÖTE	ZÄHNE	ZÄHME	ZÄHTE
She wanted to sew the button on (THREAD)	FADEN	FABEN	FANEN	PEDAL	PEBAL	PENAL
The bank robber fled through an underground tunnel (CANAL)	KANAL	KAMAL	KANAL	PLANET	PLAMET	PLANET
You can hear and see and smell it (SENSES)	SINNEN	SIMMEN	SIGGEN	TONNE	TOMME	TOGGE

Only half of the sentences are shown. Congruent sentences for Form B see Appendix B.
[a] Indicates a German saying or idiom.

Appendix B. Sample stimuli from Experiment 2 (form A not shown)

Cond.	Sentences Form B	Illusion	Auditory	Visual
E	Among the staff they did not hang the case on the large CLOCK[a]	GLOCKE	GLOPPE	GLOTTE
E	For driving a car he was not anymore in the right POSITION[a]	LAGE	LABE	LAKE
E	When I was a child I often went to purchase chocolate in the little Tante Emma SHOP[a]	LADEN	LABEN	LAGEN
E	He shamed himself into ground and BOTTOM[a]	BODEN	BOBEN	BOTEN
E	When he sat back, there was a cracking sound from the chair's BACK	LEHNE	LEHME	LEHDE
E	For the future we had no further PLANS	PLÄNE	PLÄME	PLÄKE
E	The workers were on strike for an increase of their WAGES	LÖHNE	LÖHME	LÖTE
E	His life hung on a silk THREAD[a]	FADEN	FABEN	FANEN
E	From today on the program is aired on a different CHANNEL	KANAL	KAMAL	KANAL
E	In her rage she was completely out of her MIND[a]	SINNEN	SIMMEN	SIGGEN
U	For baking one needs SUGAR	ZUCKER	ZUPPER	ZUTTER
U	Last week I bought a new SHELF	REGAL	REBAL	REKAL
U	Ashamed he showed me his SPEECH	REDE	REBE	RETE
U	My husband was always interested in FASHION	MODE	MOBE	MOGE
U	We filled the glass with CREAM	SAHNE	SAHME	SAHDE
U	The boss of the company talked to one of his SONS	SÖHNE	SÖHME	SÖHKE
U	The children of the director always had good TEETH	ZÄHNE	ZÄHME	ZÄHTE
U	Joyfully, my mother looked for the PEDAL	PEDAL	PEBAL	PENAL
U	During sunrise there appeared a PLANET	PLANET	PLAMET	PLANET
U	In the front garden there stood a high CONTAINER	TONNE	TOMME	TOGGE

"Cond." refers to "Condition"; E, expected/congruent; U, unexpected/incongruent.
[a] Refers to German sayings or idioms.

References

Bavelier, D., & Neville, H. J. (2002). Cross modal plasticity: Where and how? *Nature Reviews Neuroscience, 3*, 43–452.

Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., & Hallett M (2003). Neural correlates of cross-modal binding. *Nature Neuroscience, 6*, 190–195.

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex, 11*, 1110–1123.

Calvert, A. G., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal identification. *Trends in Cognitive Sciences, 2*, 247–253.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., Iversen, S. D., & David, A. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*, 593–596.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage, 14*, 427–438.

Churchland, P. S., & Churchland, P. M. (2002). Neural worlds and real worlds. *Nature Reviews Neuroscience, 3*, 903–907.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language, 26*, 527–538.

Connine, C. M., Blasko, D. G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual-differences, and the role of sentence context. *Perception & Psychophysics, 56*, 624–636.

Connine, C. M., & Clifton, C., Jr. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 291–299.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*, 201–215.

Dekle, D. J., Fowler, C. A., & Funnell, M. G. (1992). Audiovisual integration of real words. *Perception and Psychop hysics, 51*, 355–362.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London, Series B (Biological Science), 353*, 1245–1255.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Reviews of Neuroscience, 18*, 193–222.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology, 1*, 121–144.

Eagleman, D. M. (2001). Visual illusions and neurobiology. *Nature Neuroscience Reviews, 2*, 920–926.

Easton, R. D., & Basala, M. (1982). Perceptual dominance during lip-reading. *Perception & Psychophysics, 32*, 562–570.

Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex, 10*, 308–317.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language, 27*, 143–165.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience, 2*, 704–716.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*, 1–47.

Feng, A. S., & Ratnam, R. (2000). Neural basis of hearing in real-world situations. *Annual Review of Psychology, 51*, 699–725.

Frith, C., & Dolan, R. (1997). Brain mechanisms associated with top-down processes in perception. *Philosophical Transactions of the Royal Society (London), 352*, 1221–1230.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*, 3–28.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 816–828.

Friston, K. (2002). Functional integration and inference in the brain. *Progress in Neurobiology, 68*, 113–143.

Fuster, J. M. (2000). Executive frontal functions. *Experimental Brain Research, 133*, 66–70.

Fuster, J. M. (2001). The prefrontal cortex—an update: Time is of the essence. *Neuron, 30*, 319–333.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 110–125.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*, 613–656.

Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language, 44*, 325–349.

Ghazanfar, A. A., & Logothetis, N. K. (2003). Neuroperception: Facial expressions linked to monkey calls. *Nature, 423*, 937–938.

Green, K. (1998). The use of auditory and visual information in phonetic processing: Implications for theories of speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech* (pp. 3–26). Sussex, UK: Psychology Press.

Green, K. P., & Gerdeman, A. (1995). Cross-modal discrepancies in coarticulation and the integration of speech formation: The McGurk effect with mismatched vowels. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1409–1426.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception & Psychophysics, 50*, 524–536.

Grossberg, S., & Stone, G. O. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review, 93*, 46–74.

Goldman-Rakic, P. (1995). Cellular basis of working memory. *Neuron, 14*, 477–485.

Hietanen, J. K., Manninen, P., Sams, M., & Surakka, V. (2001). Does audiovisual speech perception use information about facial configuration? *European Journal of Cognitive Psychology, 13*, 395–407.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience, 3*, 284–291.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science, 279*, 1213–1216.

Jäncke, L., Wüstenberg, T., Scheich, H., & Heinze, H.-J. (2002). Phonetic perception and the temporal cortex. *Neuroimage, 15*, 733–746.

Jones, J. A., & Munhall, K. G. (1997). The effects of separating auditory and visual sources on audiovisual integration of speech. *Canadian Acoustics, 25*, 13–19.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *Neuroreport, 14*, 1129–1133.

Kolb, F. C., & Braun, J. (1999). Blindsight in human observers. *Nature, 377*, 336–338.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138–1141.

Kuhl, P. K., & Meltzoff, A. N. (1988). Speech as an intermodal object of perception. In A. Yonas (Ed.), *Perceptual development in infancy: The Minnesota symposia on child psychology* (pp. 235–266). Hillsdale: Erlbaum.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustic Society of America, 100*, 2425–2438.

Langenmayr, A. (1997). *Sprachpsychologie: Ein Lehrbuch. [Linguistic Psychology: A Textbook.]*. Göttingen Germany: Hogrefe.

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Science, 3*, 254–264.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revisited. *Cognition, 21*, 1–36.

Lucas, M. (1999). Context effects in lexical access: A meta-analysis. *Memory & Cognition, 27*, 385–398.

MacDonald, J., Andersen, S., & Bachman, T. (2000). Hearing by eye: How much spatial degradation can be tolerated? *Perception, 29*, 1155–1168.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24*, 253–257.

Marslen-Wilson, W. D. (2000). What phonemic decision making does not tell us about lexical architecture. *Behavioral and Brain Sciences, 23*, 337–338.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1996). Integration of multiple sources of information in language processing. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 397–432). Cambridge, MA: MIT Press.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Massaro, D. W., & Cohen, M. M. (1983a). Categorical or continuous speech perception: A new test. *Speech Communication, 2*, 15–35.

Massaro, D. W., & Cohen, M. M. (1983b). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 753–771.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology, 23*, 558–614.

Massaro, D. W., & Cohen, M. M. (1993). Perceiving asynchronous bimodal speech perception by ear and eye: A paradigm for psychological inquiry. *Speech Communication, 13*, 127–134.

Massaro, D. W., & Cohen, M. M. (1995). Continuous versus discrete information processing in pattern recognition. *Acta Psychologica, 90*, 193–209.

Massaro, D. W., & Stork, D. G. (1998). Speech recognition and sensory integration. *American Scientist, 86*, 236–244.

Massaro, D. W., & Warner, D. S. (1977). Dividing attention between auditory and visual perception. *Perception & Psychophysics, 21*, 569–574.

McClelland, J. L. (1991). Stochastic interactive processes and the effects of context on perception. *Cognitive Psychology, 23*, 1–44.

McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive Psychology, 18*, 1–86.

McGrath, M., & Summerfield, A. Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America, 77*, 678–685.

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

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 433–443.

Mercado III, E., Myers, C. E., & Gluck, M. A. (2002). A computational model of mechanisms controlling experience-dependent reorganization of representational maps in auditory cortex. *Cognitive, Affective and Behavioral Neuroscience, 1*, 37–55.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Reviews of Neuroscience, 24*, 167–202.

Mottaghy, F. M., Gangitano, M., Krause, B. J., & Pascual-Leone, A. (2003). Chronometry of parietal and prefrontal activations in verbal working memory revealed by transcranial magnetic stimulation. *Neuroimage, 18*, 565–575.

Mottonen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research, 13*, 417–425.

Mull, B. R., & Seyal, M. (2001). Transcranial stimulation of left prefrontal cortex impairs working memory. *Clinical Neurophysiology, 112*, 1672–1675.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics, 58*, 351–362.

Näätänen, F., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). Primitive intelligence in the auditory cortex. *Trends in Neuroscience, 24*, 283–288.

Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 873–889.

Noesselt, T., Shah, N. J., & Jäncke, L. (2003). Top-down and bottom-up modulation of language related areas—an fMRI Study. *BMC Neuroscience, 4*, 13. Available from http://www.biomedcentral.com/1471-2202/4/13.

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299–370.

Olson, I. R., Gatenby, J. C., & Gore, J. C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cognitive Brain Research, 14*, 129–138.

Pandya, D. N. (1995). Anatomy of the auditory cortex. *Revue Neurologique (Paris), 151*, 486–494.

Pare, M., Richler, R. C., ten Hove, M., & Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception & Psychophysics, 65*, 553–567.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 699–725.

Rainer, G., Rao, S. C., & Miller, E. K. (1999). Prospective coding for objects in the primate prefrontal cortex. *Journal of Neuroscience, 19*, 5493–5505.

Rees, G., Frackowiak, R., & Frith, C. (1997). Two modulatory effects of attention that mediate object categorisation in human cortex. *Science, 275*, 835–838.

Rosenblum, L. (1989). Towards an ecological alternative to the motor theory of speech perception. *Perceiving-Action Workshop Review, 2*, 25–28.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics, 59*, 347–357.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S.-T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127*, 141–145.

Sams, M., Manninen, P., Surakka, V., Helin, P., & Kättö, R. (1998). McGurk effect in Finnish syllables, isolated words, and words in sentences: Effect of word meaning and sentence context. *Speech Communication, 26*, 75–87.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*, 474–949.

Samuel, A. G. (1991). A further examination of attentional effects in the phonemic restoration illusion. *Quaterly Journal of Experimental Psychology, 43*, 679–699.

Samuel, A. G. (1997). Lexical activation produced potent phonemic percepts. *Cognitive Psychology, 32*, 97–127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within. *Psychological Science, 12*, 348–351.

Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language, 48*, 416–434.

Schwartz, J.-L., Robert-Ribes, J., & Excudier, P. (1998). Ten years after Summerfield: A taxonomy of models for audio-visual fusion in speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech* (pp. 85–108). Sussex UK: Psychology Press.

Summerfield, A. Q. (1979). Use of visual information for phonetic perception. *Phonetics, 36*, 314–331.

Summerfield, A. Q. (1992). Lipreading and audiovisual speech perception. *Philosophical Transactions of the Royal Society of London (B), 335*, 71–78.

Summerfield, A. Q., & McGrath, M. (1984). Detection and resolution of audiovisual incompatibility in the perception of vowels. *Quaterly Journal of Experimental Psychology: Human Experimental Psychology, 36*, 51–74.

Taylor, W. L. (1953). 'Cloze procedure': A new tool for measuring readability. *Journal Quaterly, 30*, 415–433.

Von Stein, A., Chiang, C., & König, P. (2000). Top-down processing mediated by interareal synchronization. *Proceedings of the National Academy of the Sciences USA, 97*, 14748–14753.

Walker, S., Bruce, V., & O'Malley, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. *Perception & Psychophysics, 57*, 1124–1133.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392–393.

Warren, R. M., & Sherman, H. L. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics, 16*, 150–156.

Windmann, S., Daum, I., & Güntürkün, O. (2002a). Dissociating Prelexical and Postlexical Processing of Affective Information in the Two Hemispheres: Effects of the Stimulus Presentation Format. *Brain and Language, 80*, 269–286.

Windmann, S., & Krüger, T. (1998). Subconscious detection of threat as reflected by an enhanced response bias. *Consciousness and Cognition, 7*, 603–633.

Windmann, S., & Kutas, M. (2001). Electrophysiological correlates of emotion-induced recognition bias. *Journal of Cognitive Neuroscience, 13*, 577–592.

Windmann, S., Urbach, T., & Kutas, M. (2002b). Cognitive and neural processes underlying decision biases in recognition memory. *Cerebral Cortex, 12*, 808–817.

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex, 13*, 1034–1043.