Review of Psychology,
2007, Vol. 14, No. 2, 77-91 UDC 159.9

Sentence context induces lexical bias in audiovisual speech perception

SABINE WINDMANN

The present study investigated whether semantic context enhances accuracy of word perception or merely induces a bias to perceive any speech input as a contextually appropriate word. Audiovisual speech tokens that were typically perceived as coherent words were compared with dubbed comparison stimuli that were not perceived as coherent words, either because they did not allow for the fusion of the auditory and visual speech inputs (Experiment 1), or because successful fusion resulted in a lexically inappropriate phoneme (Experiment 2). These dubbed speech tokens were presented as endings of semantically congruent versus incongruent sentences as subjects were asked to rate their lexical status (i.e., the word-likeness of the tokens). Results showed that subjects rendered enhanced lexicality ratings in semantically congruent conditions relative to incongruent conditions, whether or not the evaluated token was perceived as a word, and whether or not it allowed for audiovisual fusion. This reflects an effect of sentence context on lexical bias, not sensitivity (i.e., accuracy). Results speak against a clear distinction between lexical and semantic levels of analysis and are therefore inconsistent with models locating word recognition prior to semantic activation.

Key words: lexical decision, signal-detection theory, semantic context, audiovisual integration, top-down

Semantic context effects on lexical identification

Using behavioral, electrophysiological, neuroimaging, and other techniques, many psycholinguistic studies have shown that semantic context facilitates processing of anticipated lexical target items. Particularly strong effects have been obtained with procedures that present semantically constraining incomplete sentences and ask subjects to select, attend to, or generate a final word that either matches or does not match the meaning of the sentence (Coninne, 1987; Duffy, Henderson, & Morris, 1989, Fischler & Bloom, 1985; Kleinman, 1980; Kutas & Hillyard, 1980; Jordan & Thomas, 2005; Schuberth, Spoer, & Lane, 1981; Sereno, Brewer, & O'Donnel, 2003; Stanovich & West, 1983). That is, when a sentence such as "I went out to walk the ___" is presented, recipients activate "dog" and semantically related concepts much more easily than, say, "fisherman" or "table". The Hayling test even makes use of this paradigm to diagnose executive dysfunctions, usually associated with prefrontal cortex damage (Burgess & Shallice, 1997; Nathaniel-James, Fletcher, & Frith, 1997).

The central question emerging from these studies is how semantic expectations are represented to influence lexical decisions. A variety of connectionist hierarchical network models have been designed to account for these processes. A subset of these models (e.g., Hinton & Shallice, 1991; Grossberg & Stone, 1986; Rogers & McClelland, 2004), usually called "interactive", implement context effects by feeding the activity from semantic levels back to the appropriate units at the lexical level. This activation subsequently facilitates recognition of the expected word (relative to other words) from the speech input.

Other models propose influences of semantic expectations on word identification without feedback connections (Becker, 1980; Forster, 1979; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980; Norris, 1995; Massaro, 1998; Masson & Borowsky, 1998; Swinney, 1979). In their strongest forms, these models propose a continuous forward flow of speech information from the level of phoneme and word form recognition to word meaning selection and eventual decision-making. The assumption is usually that the system awaits the actual speech input, activates all input-matching word forms, and then uses semantic context (among other evidence) to select or filter out the best-matching interpretation. This matching process could

Sabine Windmann, Institute of Psychology, Department of Cognitive Psychology II, Mertonstr. 17, 60054 Frankfurt/Main, Germany. E-mail: S.Windmann@psych.uni-frankfurt.de (the address for correspondence).

Acknowledgements: I would like to thank Anne Kohler and Daniela Mengel for their help with data acquisition, and Hans-Georg Bosshardt for helpful discussions.

be performed either by selective maintenance/facilitation of the contextually appropriate word (i.e., prevention from decay), by lateral suppression/inhibition of context-inappropriate representations, or by a combination of both these mechanisms (Fischler & Bloom, 1985; Schwanenflugel & Shoben, 1985).

Autonomous and interactive versions of hierarchical connectionist models involve a clear spatial segregation between phonemic, lexical, and semantic levels of processing (Reilly & Sharkey, 1992). Whether semantic context acts via feedback connections or via bottom-up driven filtering mechanisms, the assumption is essentially that semantic processes operate on and select between activated word units represented one level below the semantic level (Cree, McRae, & McNorgan, 1999; Hinton & Shallice, 1991). Accordingly, the many studies that have used phonemically ambiguous words to compare contextually-appropriate interpretations with contextually inappropriate interpretations (for meta-analysis, see Lucas, 1999) have analyzed the effects of semantic context only on word selection, not on nonword evaluation. This seems to reflect the widely (and often implicitly held) belief that word form selection occurs *before* semantic meaning extraction, and that only a positively identified word can be integrated with sentence context.

Accordingly, the question of how lexically specific (i.e., how word-form selective) semantic activation is has rarely been investigated. The question is important for all types of speech models as it refers to the width of the activation function employed by semantic context. In interactive terms, the question is how focused the activation fed back from semantic to lexical levels is; for autonomous models, the question relates to the steepness of the slope of the semantic filter. At the one extreme, semantic context could imply inhibition of all non-matching speech input patterns as in an all-or-none selection process, thereby facilitating activation exclusively for semantically appropriate word(s). Alternatively, semantic context could induce a graded activation centered around the expected word unit, but extending to neighboring representations (whether or not these are words). This would mean that input patterns that share perceptual features with the expected word would be (partially) co-activated by the semantic context along with the congruent word (Gaskell & Marslen-Wilson, 2001; Connine, 1987, 1990, 1994; Schmidt, 1976).

There is indeed some evidence in the literature that the effects of semantic context are lexically and perceptually unspecific when they operate late during speech perception. Connine and colleagues showed for ambiguous words that the interpretation of a word that matched the semantic context was preferred over the incongruent interpretation, although both interpretations were activated at the lexical level as evidenced by reaction times and priming effects (Connine, 1987; Connine, Blasko, & Wang, 1994). In addition, they showed that word interpretations covaried with manipulations of pay-off matrices (Connine & Clifton, 1987), from which they inferred that semantic effects induced a late-acting word selection bias without affecting lexical or perceptual sensitivity (similar interpretations are put forth by Samuel, 1981, Experiment 3). On the other hand, studies with emotional words and derived nonwords found influences of semantic word meaning prior to (and independent of) accurate word-nonword discrimination (Ortigue, Michel, Muray, Mohr, Carbonnel, & Landis, 2004; Windmann & Krüger, 1998; Windmann, Daum, & Güntürkün, 2004). This can only be accomplished when semantic levels have access to prelexical representations, so that some effects of semantic meaning occur before (or at least coincident with) lexical analysis. However, the available evidence is relatively sparse and rather inconsistent, so more studies are needed that directly address this issue.

Audiovisual integration

Due to the belief that the perceptual dimension is quite irrelevant for semantic analysis which is often thought to be amodal, previous studies on semantic context effects on lexical identification have almost exclusively used unimodal stimuli. However, many researchers have stressed the fact that natural speech is inherently audiovisual and could as such involve special codes and routes of lexical access (Chen, d'Arcais, & Cheung, 1995; Pring, 1985; Sartori & Masutto, 1982; Schwartz, Robert-Ribes, & Excudier, 1998), perhaps even involving co-activation of vocal gestures and motor commands (Fowler & Rosenblum, 1991; Liberman & Mattingly, 1985). It has been suggested that this multimodal characteristic of audiovisual speech might result in higher clarity and higher robustness against contextual influences, including semantic context effects (Brancazio, 2004; Fowler & Rosenblum, 1991; Green, Kuhl, Meltzoff & Stevens, 1991; Langenmayr, 1997; Navarra & Soto-Faraco, 2007; Sams, Manninen, Surakka, Helin, & Kättö, 1998; Sartori & Masutto, 1982; Windmann, 2004). Some authors have argued that it seems generally unclear in how far inferences derived from unimodal speech stimuli generalize to audiovisual language (Brancazio, 2004; Iverson, Bernstein, & Auer, 1998; Vroomen & de Gelder, 2000).

Researchers often employ a well-known audiovisual illusion to investigate the effects of audiovisual integration during speech perception. In 1976, McGurk and MacDonald observed that perception of auditory speech can be altered significantly by observation of the speaker's lip movements. For instance, when /ba/ is presented acoustically while a speaker is mouthing /ga/, subjects typically report /da/. This means that the visual speech cue fuses with the discrepant auditory speech information into a novel phonemic percept that corresponds with neither of the two inputs actually given (henceforth called a *fusion response*). In this form, the illusion occurs most frequently when labial auditory consonants are paired with nonlabial visual consonants

(MacDonald & McGurk, 1978). By contrast, the reversed pairing, e.g., visual /ba/ paired with auditory /ga/, typically yields a *combinatorial response* (/bga/) where the two phonemes are not fused into one novel phoneme, but are both perceived and reported.

The effect resists a variety of cognitive manipulations, much unlike unimodal ambiguous speech tokens. It occurs whether or not subjects are informed about the sensory discrepancy and told to report only one of the modalities (Massaro, 1998), are given extensive practice (Summerfield & McGrath, 1984), or are made aware of the discrepancy by hearing a female voice dubbed onto the video of a male speaker and vice versa (Green, Kuhl, Meltzoff & Stevens, 1991). A similar, though somewhat less sophisticated effect of audiovisual integration has also been observed in preverbal infants (Kuhl & Meltzoff, 1982; Rosenblum, Schmuckler & Johnson, 1997) and even monkeys (Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Ghazanfar, Maier, Hoffman, & Logothetis, 2005). Such findings have led to the belief that audiovisual integration is a highly elementary process performed at very early sensory levels, prior to lexical analysis (Calvert et al., 1997; Dekle, Fowler, & Funnell, 1992; Ghazanfar et al., 2005; Hietanen, Leppänen, Illi, & Surakka, 2004; Green et al., 1991; Langenmayr, 1997; Sams et al., 1998).

Two recent studies have shown that audiovisually integration in the form of the "McGurk illusion" is nonetheless subject to context effects. Both studies embedded the effect in spoken sequences that could potentially be perceived as correct words, depending on the interpretation of the audiovisually discrepant phoneme. Brancazio (2004) showed that the likelihood to choose either the visual or the auditory version of such dubbed spoken sequences depends on which of the two interpretations reflects a legitimate word as opposed to a pseudoword. Interestingly, this effect was larger for slow responses compared to fast responses, presumably due to the fact that lexical effects need to unfold over time during on-line speech perception. Second, Windmann (2004, Experiments 1 and 2) presented the McGurk illusion embedded in lexical contexts such that the typical audiovisual fusion yielded a legitimate word (e.g., auditory /laben/ and visual /lagen/ yield /laden/, the German word for "shop"). These 'fusion words' were then presented as sentence-final words in highly constrained semantic contexts with which they were either semantically congruent ('We went to buy some chocolate at the corner in the little coffee ....') or not ('I prefer to take my coffee with milk and ...'). Results showed that the fusion response occurred more frequently (Experiment 1), and was rated as perceptually clearer (Experiments 1 and 2) in the congruent condition compared to the incongruent condition.

Both studies concluded that despite previous reports stressing the autonomy and cognitive inaccessibility of the phenomenon, the McGurk illusion, and perhaps audiovisual speech in general, is still subject to higher-order speech context, and may therefore provide unexplored means for investigating the mechanisms of speech perception, in particular with respect to the role of audiovisual integration. What remains unclear from both of these studies is whether the context effects are specific to words containing the McGurk illusion or generalize to other audiovisual speech tokens as well that are either lexically illegitimate or do not allow for audiovisual fusion.

Aims and design of the present study

The present study investigated semantic context effects in audiovisual speech perception, and, conversely, the contribution of audiovisual fusion to semantic context integration. Specifically, the study addressed the question of how specifically semantic context facilitates the expected audiovisual word as opposed to perceptually incoherent words and nonwords. Highly expected audiovisual words were compared with perceptually similar audiovisual nonwords to see whether context effects would generalize to items that are not part of the mental lexicon but are perceptually similar to the expected word. To specifically examine the role of audiovisual fusion in this process, audiovisual stimuli that were perceived as lexically correct words due to intact audiovisual fusion were compared with audiovisual stimuli that were perceived as nonwords either because audiovisual fusion failed (Experiment 1), or because audiovisual fusion rendered a lexically incorrect phoneme (Experiment 2).

Target stimuli were spoken two-syllable sequences containing dubbed audiovisual consonant information in the central position, as in the studies of Windmann (2004) and Connine (1990). Target words in the experimental condition were designed such that the dubbed phonemes typically prompt audiovisual fusion according to MacDonald and McGurk (1978), thereby yielding perception of a lexically correct German word (henceforth called "fusion words" or "experimental stimuli"). That is, auditory /laben/ and visual /lagen/ were presented, but fused to /laden/, the German word for "shop". These illusory fusion words were then presented as sentence-final words of highly constrained incomplete sentences with which they were either semantically congruent or not (as in Connine, 1987). Effects of the congruency manipulation on identification and evaluation responses were examined and compared to those obtained with comparison stimuli that i) contained the same amount of audiovisual conflict but did not allow for audiovisual fusion (Experiment 1), or ii) did allow for audiovisual fusion but were nonetheless lexically inappropriate (Experiment 2). Notably, both types of comparison stimuli were objectively comparable to the fusion words with regards to their lexical difficulty and syntax as they both contained one incorrect phoneme at the central position. The difference was only that this flaw remained subjectively unnoticed in the case of the fusion words but not in the case of the comparison stimuli (which were therefore nonwords). Specifically, in Experiment 1, the comparison stimuli were nonwords be-

cause audiovisual fusion was not possible; in Experiment 2, the comparison stimuli were nonwords (despite successful audiovisual fusion) because the fused phoneme was still lexically incorrect (beyond ambiguity). The question was whether any of these manipulations would reduce the effects of semantic context, that is, the difference between semantically congruent and incongruent tokens.

Following the logic of Connine and others (Connine, 1987; Connine & Clifton, 1987; Connine et al., 1994; Gaskell & Marslen-Wilson, 2001), the following general hypotheses were derived: If word meaning selection requires prior word identification (i.e., only identified words can be subject to semantic activation), then significant semantic context effects should be observed only for the fusion words (i.e., words that are perceived as lexically intact). The comparison stimuli, by contrast, should not show any semantic context effects (in both experiments) as they are clearly identifiable nonwords that would not be subject to semantic analyses.

This hypothesis is equivalent to expecting sentence context to enhance *lexical sensitivity,* i.e., the ability to distinguish the appropriate word from inappropriate speech inputs, including nonwords. If sentence context facilitates only the expected word above all other word units and nonwords, then the activation difference between words and nonwords should be larger in the semantically congruent condition than in the incongruent condition. Conversely, if semantic processes do not operate exclusively on the expected word unit in an all-or-none fashion, and instead induce a graded biasing influence that extends to non-lexical, but perceptually related items, then semantic activation should generalize from lexically expected units (fusion words) to perceptually similar tokens (comparison stimuli), even if these are no coherent words. In that case, the relevant variable determining the size of the sentence context effect should only be the degree of experienced similarity between the expected word and the actual input pattern, not its lexical status or its audiovisual compatibility. This hypothesis is equivalent to expecting a sentence context effect on lexical bias, i.e., on the tendency to assume that a word has been presented, whether or not a word has actually been presented. It predicts that activation of semantically congruent tokens is higher than that of semantically incongruent tokens for both coherent words (fusion words) and incoherent nonwords (comparison words).

EXPERIMENT 1

This experiment examined whether audiovisual speech signals are more strongly influenced by semantic expectations when the audiovisual input is fused into a coherent word perception (fusion words) relative to audiovisual input that cannot be fused and therefore results in perception of a nonword (comparison stimuli). Note that by successful fusion I mean the generation of a unitary and novel phoneme as opposed to the perceived *combination* of the presented auditory and visual phonemes (e.g., /bg/).

The comparison stimuli had the same syntax as the auditory words, including the same amount of audiovisual conflict from the same consonant combinations, but with inverted pairings ("inverted McGurk effect"). That is, while auditory /mobe/ and visual /moge/ are fused into the word /mode/ (the German word for "fashion") in the experimental condition with the fusion words, the inverted pairing of auditory /moge/ and visual /mobe/ was used in the comparison condition. The latter typically yields the auditory response /mobe/ or the combinatorial response /mobge/, both of which are nonwords. I refer to this phenomenon as *combination* of audiovisual speech cues. Thus, auditory and visual information are successfully fused into a novel phoneme in the experimental condition but not in the comparison condition where one of the two discrepant phonemes is either ignored or simply "added" onto the other.

The cover story asked subjects explicitly to detect semantic, lexical, and pronunciation mistakes. The target stimuli and their associated sentence contexts were presented in pseudorandom order. Subjects were asked to identify and rate the lexical quality of the target stimuli on a Likert scale ranging from 1 through 6, equivalent to obtaining confidence ratings in signal-detection tasks. More specifically, subjects were asked to evaluate how clearly the speaker had pronounced the spoken sequence by rating how close it was to the lexically correct word form. These ratings of "word-likeness" (lexicality ratings) were later used for signal-detection-theory analyses to determine whether sentence context had impacted lexical sensitivity or bias.

METHODS

Participants

22 healthy native German speakers with a mean age of 23.5 years (range 19 to 35) participated in this study; 20 were female. All participants were undergraduate students of Psychology who received course credit for participation. None of them had participated in previous experiments with the McGurk illusion.

Materials

20 bisyllabic words were chosen for both, the experimental and the comparison condition (see Appendix). All contained two vowels flanking a medial consonant which served as the target for the experimental manipulations. These words provided the endings to incomplete, highly constrained sentences. In the congruent condition, they matched the sentences semantically; in the incongruent condition, novel sentences were created which did not match

any of these words. Assignment of words to semantic conditions was counterbalanced across participants so that each word appeared equally often in both conditions.

The fusion words were created with the ten different phoneme combinations for which MacDonald and McGurk (1978) reported a fusion response in at least 50% of the cases. These were the following bilabial auditory and nonlabial visual phoneme pairings: /b/+/t/=/d/ as in REDE, /b/+/g/=/d/ as in MODE, /b/+/k/=/g/ as in REGAL, /b/+/n/=/d/ as in PEDAL, /p/+/t/=/k/ as in ZUCKER, /m/+/d/=/n/ as in SAHNE, /m/+/t/ as in ZÄHNE, /m/+/g/=/n/ as in TONNE, /m/+/k/= /n/ as in SÖHNE; /m/+/n/=/n/ as in PLANET (see Appendix). The comparison stimuli were created by inverting the medial consonant pairings.

The stimuli were produced by a female speaker (S.W.) filmed in front of a plain white background with a high resolution digital video camera. This film was later cut into segments of approximately 2.8 sec duration. The audio tracks of these segments (recorded with a sampling rate of 44.1 kHz) were then dubbed onto the video tracks using 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 2.5 seconds duration were saved and exported into Motion Pictures Expert Group (mpeg) format with a size of 352 x 288 pixels, a bit rate of 1100000 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 x 9 cm. This was done to prevent subjects from looking at the speaker's eyes or elsewhere other than the lips (c.f. Summerfield, 1979). The written instructions, sentences, and typed-in responses were also presented in this window. The auditory stimuli were played with a loudness of ca. 63 dB via two loudspeakers placed at a distance of approximately 60 cm from the subject.

Two parallel lists of words (lists A and B) were constructed for each of the 10 types of phoneme combinations (see Appendix). To half of the participants, words of list A were presented as semantically congruent endings to highly constrained incomplete sentences, whereas words of list B were presented as incongruent endings to novel sentences with which they did not match semantically. As a means of counterbalancing, the opposite was done with the other half of the participants, thereby ensuring that the same targets occurred in both semantic conditions while neither the targets nor the sentences were repeated within a given participant.

Procedure

Procedures were similar to those described by Windmann (2004). 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 50-60 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 improperly synchronized so that their quality was poor. Specifically, they were told that the spoken words might be phonetically unclear, syntactically incorrect, or mispronounced; sound and picture might be poorly synchronized, the words might not match the context in which they appeared, and any mixture of all these flaws might occur. Their task was to rate the speech quality of the tokens.

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 asked to enunciate the sentence–final word they would expect to follow. This was done, first, to assess cloze probability (Taylor, 1953), and second, to encourage specific prediction of the correct word, including its perceptual form. Subjects were then presented the spoken target word on the video and were asked to type in immediately what they had understood, even if it was a nonword. If they were unsure, they were encouraged to type in most closely what they had understood. Specific reference to either the auditory or the visual modality was avoided in these instructions.

Due to the ambiguous and/or audiovisually inconsistent nature of all speech stimuli used, this task seemed natural to the participants. Even reporting the words exactly the way they had understood them (including their flaws) provided no problem. Thereafter, subjects were asked to rate the lexical clarity of spoken sequences on a 6-point rating scale; i.e., to indicate how close the token was to the correctly pronounced word (this word was written in its lexically correct form on the screen in case it had not been correctly identified).

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; $F(1,21) = 176.55$, $p < .0001$, $Eta^2 = .98$. This was true for the illusion condition (5.91 vs. 1.64) as well as for the comparison condition (5.82 vs. 1.55); with a congruency x condition interaction that was far from significance; $F(1,21) = .023$. The cloze probability measure further suggested that the semantic constraint of the sentences was comparable in the four conditions: Cloze probability of all target words was above 88%, with no significant differences between conditions.

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

RESULTS AND DISCUSSION

Data analysis

Identifications were compared between conditions using *Chi*^{2} (χ^2) tests. 2 x 2 Analyses of variance with repeated measures for the two factors semantic congruency (congruent versus incongruent) and experimental condition (experimental versus comparison) were performed on the rating measure as well as on lexical sensitivity and bias measures as determined by two-high-threshold analysis (a nonparametric variant of signal-detection theory used for small numbers of observations; Snodgrass & Corwin, 1988). To this end, ratings equal and above 4 were considered "word" responses whereas ratings of 3 and below were considered "nonword" responses, a division that resulted in approximately equal proportions of "word" and "nonword" responses. [Note that despite the labeling, this does not necessarily imply that tokens rated 4 or higher are perceived as absolute words, while those classified as 3 or lower are perceived as absolute nonwords. What is relevant here is only that tokens classified as 4 or more are perceived as more "word-like" than tokens classified as 3 or less.] *Partial Eta*^{2} (η_p^2) is reported for all ANOVAs as a measure of effect size.

Hits were defined as "word" responses to fusion words and false alarms were defined as "word" responses to comparison stimuli. Semantic context effects on lexical sensitivity (i.e., the ability to discriminate between words and nonwords) and lexical bias (i.e., the likelihood to render a "word" response, whether or not a word had actually been presented) were determined. Note that these two variables are statistically independent (Snodgrass & Corwin, 1988).

Identification responses. Percentages of typical fusion responses, combinatorial responses, auditory responses, visual responses, and atypical responses ("other") were significantly different for the fusion words compared to comparison stimuli containing the inverted McGurk effect, as expected (χ^2= 223.43, *df*= 4, *p*< .00001, see Table 1). For the fusion words (experimental condition), the typical fusion response was the most frequent with 59%, while combinatorial responses occurred in less than 1%, comparable to previous studies (MacDonald & McGurk, 1978). For the stimuli with the inverted McGurk effect (comparison condition), the reversed pattern was found: auditory and combinatorial responses were the most frequent (together 62.5%), while fusion responses occurred in less than 1%.

These results indicate that the stimuli we had designed were appropriate for the present purposes. Subjects fused the visual information with the auditory information to a lexically legitimate word in the illusion condition, but not in the comparison condition. Effects of semantic congruency on this response pattern were not significant. Although congruent tokens made subjects' responses shift in the comparison condition from auditory responses to more visual, combinatorial, and 'other' responses (see Table 1), this is not relevant for the present purposes because all these response types are nonwords.

Lexicality Ratings. Central for the present purposes were the effects of semantic congruency on the lexicality (word-

Table 1

Proportion of responses (in %) given in Experiment 1 to Fusion words (experimental condition) and comparison stimuli ("inverted McGurk effect") in the congruent condition compared with the incongruent condition.

Experimental	Fusion	Auditory	Visual	Combinatory	Other
Congruent	56.4 (17.1)	6.4 (14.3)	14.5 (9.1)	0 (0)	22.7 (12.8)
Incongruent	61 (16.9)	7.3 (13.2)	6.4 (9.5)	1.8 (5.9)	23.6 (15.9)
F(1,21)	0.92	0.056	6.83*	n/a	.04
Comparison	Fusion	Auditory	Visual	Combinatory	Other
Congruent	0 (0)	15.9 (17.1)	19.1 (4.3)	34.1 (18.2)	30.9 (14.8)
Incongruent	0.9 (4.3)	49.1 (20.2)	7.3 (9.8)	26.4 (17.9)	16.4 (11.8)
F(1,21)	n/a	64.56**	30.33**	4.07 (*)	13.31**

Note. N = 220 observations in each condition. The F-values refer to the results of a univariate ANOVA comparing congruent vs. incongruent conditions; * indicates significant at p< .05, ** significant at *p*< .005, (*) marginally significant (*p*< .06). Note that the "fusion" response would reflect the lexically expected phoneme. Standard deviations are given in parenthesis.

likeness) ratings. The raw data showed a significant main effect of semantic congruency ($F(1,21) = 12.96$, $p < .005$, $\eta_p^2 = .38$), indicating that semantically congruent stimuli were rated as more word-like than semantically incongruent stimuli (see Figure 1A). Crucially, the effect was present for the fusion words ($t(21) = 2.50$, $p < .05$) as well as the comparison stimuli with the inverted McGurk effect ($t(21) = 4.33$, $p < .001$) with a far from significant Congruency x Condition interaction; $F(1,21) = 0.78$. Importantly, however, the fusion words were generally rated as more word-like than the comparison stimuli (main effect of Condition, $F(1,21) = 199.94$, $p < .0001$, $\eta_p^2 = .91$), indicating that they subjectively resembled real words more than the comparison stimuli, as intended, despite identical syntax.

The significance pattern of the rating measure was unchanged when only typical fusion responses were included in the analysis of the experimental condition. They were also unchanged when only combinatorial responses were considered for the comparison condition (i.e., when auditory responses were excluded).

Sensitivity and Bias. The two-high-threshold analysis indicated that semantic congruency had no significant effect on lexical sensitivity; $F(1,21) = .38$, but did increase lexical bias significantly; $F(1,21) = 11.18$, $p < .005$; $\eta_p^2 = .35$. This means that subjects had a bias to designate the stimuli more as word-like in the semantically congruent context compared to the incongruent context (see Figure 1B), regardless of whether they had identified those very same stimuli as correct words (illusion condition) or not (comparison condition).

In summary, results of this experiment indicate that semantic context effects are independent of successful audiovisual fusion and the resulting word-nonword status. Semantic context effects were significant in both conditions, with no significant interaction ($F<1$). This indicates that subjects rated the target stimuli as more word-like when they occurred in a semantically congruent as opposed to an incongruent context, but they did so whether or not the stimuli involved audiovisual fusion.

This unspecific facilitation of word responses seems to conform with what Connine and Clifton (1987) describe as a decision-bias account of sentence context effects. This interpretation is supported by the present two-high-threshold analysis which showed that lexical bias, but not lexical sensitivity (i.e., the ability to respond differentially to words and nonwords), is increased by congruent semantic context. Subjects considered a spoken sequence as more word-like when *parts* of the token corresponded to a semantically expected word, even if the stimulus as a whole was explicitly identified as a nonword. Whether this finding is specific to stimuli involving phonemic ambiguity due to audiovisual conflict or generalizes to other spoken sequences that do not contain any audiovisual conflict will be investigated in Experiment 2. If the effect is indeed a genuine lexical bias, then it should be independent of audiovisual fusion, and depend only on the degree of perceptual similarity between expected word and pseudoword.

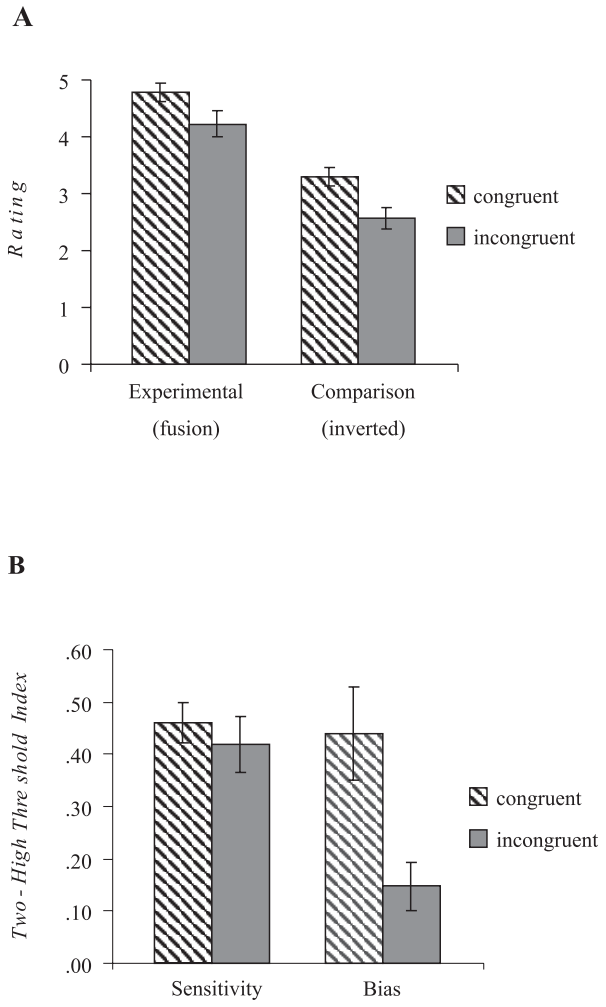

Figure 1. Lexicality ratings in Experiment 1. A: Fusion words (experimental condition) and comparison stimuli (containing the inverted McGurk effect) in semantically congruent as compared to incongruent sentence contexts. B: Results of the two-high-threshold analysis: Semantic congruency increases lexical bias, but not lexical sensitivity.

EXPERIMENT 2

This experiment followed exactly the same procedures as Experiment 1 but used comparison stimuli that contained a lexically inappropriate phoneme instead of audiovisually conflicting information. It was designed to investigate whether the audiovisual fusion or the lexical status of the comparison stimuli used in Experiment 1 were responsible for the observed sentence context effects.

Fusion words were compared with stimuli that contained the auditory version of the fusion words in both modalities, auditory and visual. The target phonemes of these comparison stimuli were thus audiovisually concordant, but lexically inappropriate to the same degree as were the comparison stimuli in Experiment 1 (e.g., auditory /laben/ and visual /laben/ yielded audiovisual /laben/ instead of the German word /laden/), with one incorrect phoneme in the central position. The question was whether this manipulation would alter the semantic context effects that had been observed for the comparison stimuli in Experiment 1. If audiovisual conflict was a significant source of these effects, then the comparison stimuli used in this second experiment should show significantly less semantic context effects than the fusion words, reflecting significant variations of lexical sensitivity. Otherwise, if the source of the effects on the comparison words in Experiment 1 was merely their perceptual similarity to the fusion words, then the comparison stimuli in this second experiment should yield the same data pattern. Hence, this second experiment would either replicate or specify the source of the semantic context effects found for the comparison stimuli in Experiment 1.

METHODS

Participants

37 healthy subjects (15 male) with a mean age of 30.6 years (range 19 to 50), all native German speakers, participated in this study. All of them were family members or personal acquaintances of the experimenter (Anne Kohler). We preferred this pool of naive participants to the usual undergraduate student pool to ensure high motivation and to rule out the possibility that any theoretical knowledge or expectations about the aims of the experiment could influence subjects' decisions. The participants volunteered to participate without receiving any payment or course credits. None of them had participated in Experiment 1.

Materials and Procedures

Materials were the same as in Experiment 1 except for the comparison stimuli. These stimuli were created as follows: The auditory component of the fusion words was spoken and recorded on digital video twice. The auditory track of one of these two film segments was then dubbed onto the video track of the other film segment. This was done to obtain dubbed videos in both, the experimental and the control condition, for reasons of matching: Although both types of stimuli were meant to allow for audiovisual fusion, only the experimental words contained audiovisual conflict.

The procedures were the same as in Experiment 1. Average cloze probability of the sentences was 91.5% with no significant differences between conditions. As in Experiment 1, the ratings of the participants after each trial indicating how well the lexically correct words matched the sentences semantically showed a highly significant effect of semantic congruency ($F(1,36) = 4558$, $p < .0001$, $\eta_p^2 = .99$) with no significant differences between the two conditions (Condition x Congruency interaction: $F(1,36) = 0.26$, n.s.). The same statistical analyses were performed as in Experiment 1.

RESULTS AND DISCUSSION

Identification Responses. Proportions of typical fusion, auditory, visual, combinatorial, and atypical/other responses were significantly different for the fusion words compared to the "auditory" comparison stimuli ($\chi^2 = 520.5$, $df = 4$, $p < .00001$). Perception of auditory phonemes was changed by visual information (leading to fusion responses) in about 61% of the trials in the experimental condition, while auditory responses were correctly given on almost 95% of the trials in the comparison condition (the "auditory condition"), despite the fact that these latter responses were nonwords (see Table 1). As in Experiment 1, the probability of the fusion responses in the experimental condition did not show any significant effects of semantic congruency. Although some of the response portions varied with semantic congruency, they did so only between the various nonword response alternatives (auditory and visual). Similarly, fusion responses in the comparison condition did not show any significant effects of semantic congruency (see Table 2).

Lexicality Ratings. There was a significant main effect of semantic congruency on the lexicality rating; $F(1,36) = 20.78$, $p < .001$, $\eta_p^2 = .37$ (see Figure 2A). This effect was significantly stronger for the comparison stimuli than for the fusion words (Congruency x Condition interaction; $F(1,36) = 9.11$, $p < .01$, $\eta_p^2 = .20$); where only marginal significance was reached ($t(36) = 1.72$, $p < .10$), contrary to the comparison words ($t(36) = 5.39$, $p < .001$). In addition, the fusion words were generally rated as more word-like than the comparison stimuli (main effect of Condition; $F(1,36) = 126$, $p < .0001$, $\eta_p^2 = .78$) which reflected the fact that they subjectively resembled real words while the comparison stimuli were clearly nonwords.

Sensitivity and Bias. The two-high threshold analysis indicated no effect of semantic congruency on the sensitivity measure ($F(1,36) = 1.42$); but did show a significant effect on the bias measure; $F(1,35) = 16.64$, $p < .001$, $\eta_p^2 = .32$. As in Experiment 1, subjects rated the stimuli more word-like when they partly corresponded to semantically congruent as opposed to incongruent words, regardless of whether they had previously been identified as nonwords. Specifically, although 95% of the stimuli in the comparison condition were identified as nonwords, this condition nevertheless showed marked effects of semantic congruency in the rating measure (see Figure 2B).

Table 2
Proportion of responses (in %) given in Experiment 2 to Fusion words (experimental condition) and comparison stimuli ("auditory words") in the congruent condition compared with the incongruent condition.

Experimental	Fusion	Auditory	Visual	Combinatory	Other
Congruent	57.8 (19.3)	7.6 (12.8)	15.1 (9.4)	2.2 (6.3)	17.3 (15.0)
Incongruent	65 (20.8)	15.1 (18.5)	6.5 (9.5)	0 (0)	13.5 (13.4)
F(1,36)	2.81	6.21*	22.5**	n/a	1.51
Comparison	Lexically Expected	Auditory	Visual	Combinatory	Other
Congruent	2.7 (8.4)	95.7 (10.7)	0 (0)	0 (0)	1.6 (5.5)
Incongruent	0 (0)	93.0 (9.7)	n/a	n/a	7.0 (9.7)
F(1,36)	n/a	1.33	n/a	n/a	8.61*

Note. N = 370 observations in each condition. Note that the "fusion" response in the experimental condition would reflect the lexically expected phoneme. Standard deviations are given in parenthesis. For indices see Table 1.

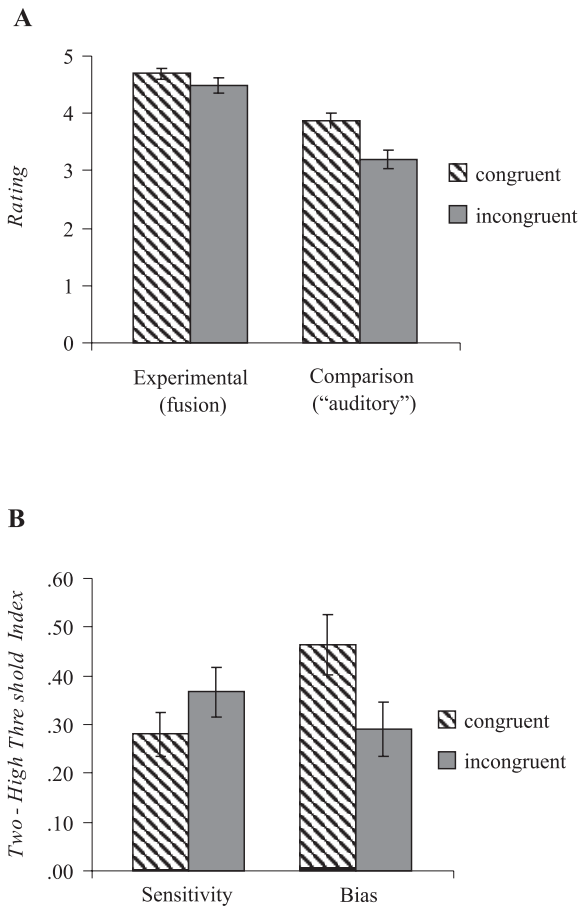

Figure 2. Lexicality ratings in Experiment 2. A: Fusion words and comparison stimuli ("auditory words") in semantically congruent as compared to incongruent sentence contexts. B: Results of the two-high-threshold analysis: Semantic congruency increases lexical bias, but not lexical sensitivity.

GENERAL DISCUSSION

This study investigated how far highly constrained incomplete sentences aid extraction and identification of semantically congruent words from the audiovisual input stream. To specifically examine the role of audiovisual integration in this process, words containing audiovisually fused phonemes were compared with tokens containing phonemes that could not be fused (Experiment 1). These two types of stimuli were designed such that their syntactical structure was objectively the same despite the fact that they elicited subjectively different lexical experiences depending on whether or not they evoked the fusion response.

To further specify the role of audiovisual fusion as opposed to mere perceptual similarity in this process, another comparison was performed with pseudowords that also allowed for audiovisual fusion (Experiment 2). The main question for both experiments was how close the subjective match between audiovisual expectations and actual audiovisual speech inputs had to be to allow for facilitatory semantic context effects to take place. If semantic context acted solely on expected words (not on nonwords), then lexically correct and audiovisually coherent words would profit more from congruent semantic context than lexically incorrect or audiovisually conflicting stimuli. As a result, hit rates (correct word identifications) would increase, and false alarms (incorrect word identifications) would decrease. This is synonymous with expecting an effect of semantic context on lexical sensitivity.

Results showed, first, that the experimental manipulation of the used audiovisual stimuli was successful. When stimuli contained conflicting audiovisual information that could be resolved by phoneme fusion, they were identified as words on the majority of trials. When stimuli contained audiovisual conflict that could not be resolved by fusion but

resulted in phoneme combination (Experiment 1), or that could be resolved by fusion but nevertheless contained a lexically inappropriate phoneme (Experiment 2), stimuli were reliably identified as nonwords. The stimuli were therefore appropriate for examining whether semantic context effects depended on audiovisual fusion and/or perceived lexical status.

Second, and more interestingly, results showed that the semantic context effects were not specific to audiovisual stimuli previously identified as coherent words. The context effects were not larger for the fusion words than they were for the comparison stimuli in either of the two experiments (if anything, they were smaller in Experiment 2). This means that semantic context altered the *bias* to evaluate spoken sequences as word-like (even if these sequences were perceptually incoherent or lexically flawed), but not the sensitivity specifically for words. In other words, congruent relative to incongruent sentence context enhanced lexicality ratings regardless of whether the target stimuli allowed for audiovisual fusion (Experiment 1), and regardless of whether they were identified as words (Experiments 1 and 2). Although subjects were sensitive to the lexical and audiovisual inconsistencies of the stimuli, as evidenced by the significantly reduced lexicality ratings they rendered for the comparison stimuli relative to the fusion words, audiovisually implausible or lexically flawed pseudowords were not immune to and not even significantly less susceptible to semantic activation than were the audiovisually coherent fusion words. Instead, the sentence context effects were statistically reliable across all conditions examined, reflecting a significant impact on lexical bias.

According to these results, semantic activation by sentence context does not require prior identification of a word, nor does it require the subjective identification of a coherent audiovisual input. Instead, the gross perceptual similarity between the expected words and the presented comparison stimuli seems to have been sufficient to increase the lexicality judgments. This result parallels earlier findings with unimodal stimuli showing that derived nonwords (though probably not maximal nonwords) can bear semantic meaning (Connine, Blasko, & Titone, 1993; Samuel, 1981; Windmann et al., 2002).

The findings provide further support for vertical similarity mapping accounts (Connine, 1987, 1994) and probabilistic matching rules of semantic-lexical interactions (Jurafsky, 1996; Marslen-Wilson, 1987; Massaro, 1998; McClelland, 1991) according to which semantic context does not exclusively operate on positively identified words, but varies gradually with the goodness-of-fit of the actual speech input to the expected speech input. If semantic activation had specifically affected the expected word over and above any other speech input, then it would have increased activation of the expected word more than that of the comparison stimuli, thereby increasing word-nonword discrimination (c.f., Samuel, 1981; Connine et al., 1994). What was observed instead were highly significant context effects on the lexical bias measure that were indifferent with regards to word-nonword status and perceptual coherence (specifically, audiovisual fusion). Although the underlying process is less selective and therefore less accurate than a word-specific activation process would have been, it has the advantage of being more robust against noisy variations in the speech input that are common in real-life audiovisual communications.

Earlier reports suggest that such biasing effects of semantic context occur primarily under delayed conditions that allow for the perceptual input to be re-interpreted in accordance with context (Borsky, Shapiro, & Tuller, 2000; Tyler, 1990). Subjects in the present study were given as much time as they liked to type in their responses, so they may indeed have relied on post-perceptual processes more than on perceptual processes as they made their judgments. It is possible that online measures of speech processing, such as electrophysiological recordings, would have been more successful in detecting traces of more word-specific effects (c.f., Sereno et al., 2003). Nevertheless, the present data show that even if such word-specific effects exist during early perceptual processing, they are either too weak to determine subjects' decisions or are inhibited/overridden during later processing in favor of a perceptually and lexically indifferent response. This finding is particularly remarkable considering that in the present task design, subjects had to enunciate the expected word before they observed and evaluated the target video, and hence their expectations were very specific and explicit.

On the other hand, there is some reason to believe that subjects did indeed use perceptual information, at least in part, when they made their lexical judgments, despite the fact that they showed lexically undifferentiated semantic effects. First, although identification and rating responses were both delayed, identification responses differed between the experimental condition and the comparison conditions, while the semantic context effects on the rating measure did not. Second, although lexicality ratings differed markedly between the experimental condition and the comparison conditions in both experiments, semantic context effects on that same measure were comparable. To reconcile both of these dissociations with a post-perceptual account, one would have to claim that the differences between conditions in the lexical judgments were indicative of perceptual processes while the lacking differences in semantic context effects were indicative of post-perceptual influences, even though both of these effects refer to the same kinds of judgments rendered at the same delay after stimulus presentation. Essentially, this explanation would imply that subjects are unable to differentiate between perceptual and semantic features as they make lexical decisions; that is, they let semantic context influence judgments about lexical status (in contrast with task instructions). This observation argues against clearly segregated levels of processing as in hierarchic connectionist networks,

in favor of alternative models, e.g. with overlapping lexical-semantic feature representations and/or re-entrant circuits (e.g., Gaskell & Marslen-Wilson, 1997; Elman, 1990; Plaut & Shallice, 1993).

In any case, the present findings are inconsistent with any strong forms of autonomous or interactive models that propose selective and exclusive facilitation of expected words all the way from perception to decision-making. Whatever mechanism drives sentential context effects does not seem to pre-require identification of a word. Instead, semantic effects can either "bypass" word identification levels or "overwrite" the results of word recognition processes at subsequent levels. As illustrated in Figure 3, where words and nonwords are represented at the same level, the behavioral consequence of this process is that semantic activation influences lexicality ratings in a graded way, not in an all-or-nothing manner, consistent with similarity mapping accounts (Connine, 1987; Connine & Clifton, 1987; Connine et al., 1994).

To the degree to which the present results obtained with audiovisual stimuli conform to earlier reports obtained with unimodal stimuli (Borsky et al., 2000; Samuel, 1981; Connine, 1987; Connine & Clifton, 1987; Connine et al., 1994), they suggest that audiovisually integrated speech underlies the same functions and variations as unimodal speech. However, it should be noted that the context effects found for the fusion words were significant only in the bias-sensitive rating measure, not in the categorical measure (identification response), unlike in previous reports with phonetically ambiguous stimuli (Samuel, 1981; Connine, 1987; Connine & Clifton, 1987; Connine et al., 1994). It is plausible that task instructions or response modalities that prompt subjects to focus more on particular perceptual details of the evaluated speech token render more differentiated results (Pring, 1985).

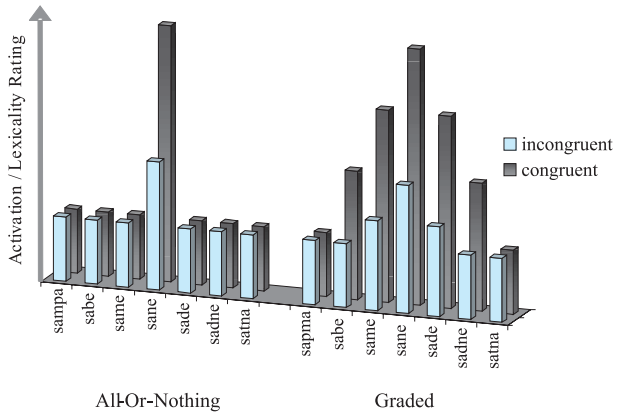

Figure 3. Illustration of the hypotheses and results.
Note. Words (here: /sane/, the German word for "cream") are represented amidst nonwords at the same level of processing. If semantic context was word-specific, then the difference between congruent and incongruent conditions should be larger for words than for nonwords (left panel, "All-or-nothing"). By contrast, if the semantic activation function followed a vertical similarity gradient (right side, "Graded"), then the difference between congruent and incongruent conditions should be comparable for words and (perceptually similar) nonwords. Evidence presented in this article speaks for the latter alternative. The illustration also takes into account the higher activation for congruent words relative to incongruent words as well as the generally higher activation of words relative to nonwords.

Furthermore, the failure to find significant effects on the categorical measure is inconsistent with the results of Experiment 1 in Windmann (2004) where the expected fusion word was given more often in the congruent condition than in the incongruent condition (though this was not the case in experiments 2 and 3 of that same study). There are two procedural differences between the present study and Experiment 1 of the earlier study that might account for this inconsistency. First, subjects in Experiment 1 of Windmann (2004) did not enunciate the expected word after reading the incomplete sentences, but they did so in Experiment 2 of Windmann (2004), which also failed to find significant effects of semantic context on the categorical response. The enunciation was chosen for the present study as to ensure perceptually specific preactivation of the expected word to encourage word-specific effects. It is plausible that this procedure countered the effects of semantic context on the categorical measure as it made subjects use a perceptually more well-defined comparison standard that was harder to be overturned. The graded lexicality ratings, by comparison, may have been less resistant to this procedure as they have a finer quantitative resolution which might make them more sensitive to postperceptual interpretations. Moreover, as a byproduct of the required enunciation, identification responses were not made immediately after reading the incomplete sentence in the present study, unlike in the Windmann (2004) study where this was the case. This subtle change in the time-line of the trials could have additionally increased the relative contribution of post-perceptual processing to the response, thereby reducing semantic context effects on the perceptually sensitive categorical measure. However, these differences do not limit the conclusions of the present study as it was not concerned with the distinction of perceptual and post-perceptual processes, but with the generalizability of semantic context effects from expected words to perceptually similar words and nonwords.

Future studies could use event-related potentials as an on-line measure of perceptual access, semantic integration and post-perceptual decision-making to verify the present interpretations and to specify the level of processing at which the effects reported here take place relative to audiovisual integration. Of particular interest is the ques-

tion of whether there is an electrophysiological signature of audiovisual fusion during spoken word perception and sentence comprehension, and if so, in what time domain. To investigate this issue, fusion words could be compared with unimodal ambiguous speech tokens and with "inverted" McGurk comparison stimuli. These stimuli would then have to be presented in semantically congruent as compared to incongruent sentence contexts to find out whether and at what stage of processing their interpretation is altered. If sentence constraint and cloze probability are carefully manipulated, this design may dissociate perceptual effects of semantic prediction from postperceptual semantic integration (Connolly, Phillips, & Forbes, 1995; Sereno et al., 2003; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).

REFERENCES

Barraclough, N.E., Xiao, D., Baker, C.I., Oram, M.W., & Perrett, D.I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience. 17,* 377-391.

Becker, C.A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8*, 493-512.

Borsky, A., Shapiro, L.P., & Tuller, B. (2000). The temporal unfolding of local acoustic information and sentence context. *Journal of Psycholinguistic Research, 29*, 155-168.

Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception & Performance, 30,* 445-463.

Burgess P.W., & Shallice, T. (1997). *The Hayling Island and Brixton Test manual.* Bury St. Edmunds: Thames Valley Test Co.

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.

Chen, H., d'Arcais, G.B., & Cheung, S. (1995). Orthographic and phonological activation in recognizing Chinese characters. *Psychological Research, 58,* 144-153.

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. (1990). Effects of sentence context and lexical knowledge in speech processing. In G. Altman (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 281-294). Cambridge, MA: MIT Press.

Connine, C.M. (1994). Vertical and horizontal similarity in spoken word recognition. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 107-120). Hillsdale, NJ: Erlbaum.

Connine, C.M., Blasko, D.M., & Titone, D.A. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32*, 193-210.

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.

Connolly, J.F., Phillips, N.A., & Forbes, K.A. (1995). The effects of phonological and semantic features of sentence-ending words on visual event-related potentials. *Electroencephalography and Clinical Neurophysiology, 94*, 276-287.

Cree, G.S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science, 23,* 371-414.

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

Duffy, S. A., Henderson, J.M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 791-801.

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

Fischler, G.S., & Bloom, P.A. (1985). Effects of constraint and validity of sentence contexts on lexical decisions. *Memory & Cognition, 13,* 128-139.

Forster, K.I. (1979). Levels of processing and the structure of of the language processor. In W.E. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic processes presented to Merill Garrett* (pp. 27-85). Hillsdale, NJ: Erlbaum.

Fowler, C.A., & Rosenblum, L.D. (1991). Perception of the phonetic gesture. In I.G. Mattingly & M. Studdert-Kennedy (Eds.)*, Modularity and the Motor Theory*. Hillsdale, NJ: Lawrence Erlbaum.

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., Maier, J.X., Hoffman, K.L., & Logothetis, N.K. (2005). Multisensory integration of dynamic faces

and voices in rhesus monkey auditory cortex. *Journal of Neuroscience, 25,* 5004-5012.

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 and 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.

Hietanen, J.K., Leppänen, J.M., Illi, M., & Surakka, V. (2004). Evidence for the integration of audiovisual emotional information at the perceptual level of processing. *European Journal of Cognitive Psychology, 16,* 769-790.

Hinton, G.E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98,* 74-95.

Iverson, P., Bernstein, L.E., & Auer, E.T. (1998). Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition. *Speech Communication, 26*, 45-63.

Jordan, T.R., & Thomas, S.M. (2002). In search of perceptual influences of sentence context on word recognition. *Journal of Experimental Psychology, 28,* 34-45.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20,* 137-194.

Kleinman, G.M. (1980). Sentence frame contexts and lexical decisions: Sentence-acceptability and word-relatedness effects. *Memory & Cognition, 8*, 336-344.

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

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

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

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

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

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

Marslen-Wilson, W.D (1987). Functional parallelism in spoken word recognition. *Cognition, 25,* 71-102.

Marslen-Wilson, W.D., & Welsh, A. (1978). Processing interactions during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

Marslen-Wilson, W.D., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1-71.

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

Masson, M.E. & Borowsky, R. (1998). More than meets the eye: context effects in word identification. *Memory & Cognition, 26,* 1245-1269.

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

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

Nathaniel-James, D. A., Fletcher, P., & Frith, C. D. (1997). The functional anatomy of verbal initiation and suppression using the Hayling test. *Neuropsychologia, 35,* 559-566.

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: visual articulartory information enables the perception of second language sounds. *Psychological Research, 71,* 4-12.

Norris, D. (1995). Signal detection theory and modularity: On being sensitive to the power of bias models of semantic priming. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 935-939.

Ortigue, S., Michel, C.M., Murray, M.M., Mohr, C., Carbonnel, S., & Landis, T. (2004). Electrical neuroimaging reveals early generator modulation to emotional words. *Neuroimage, 21,* 1242-1251.

Plaut, D.C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377-500.

Pring, L. (1985). Phonological encoding and word comprehension. *Psychological Research, 47,* 211-216.

Reilly, R.G., & Sharkey, N.E. (1992) (Eds.). *Connectionist approaches to natural language processing.* Hove, UK: Lawrence Erlbaum.

Rogers, T.T., & McClelland, J.L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach.* Cambridge, MA: MIT Press.

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

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.

Sartori, G., & Masutto, S. (1982). Visual access and phonological recoding in reading Italian. *Psychological Research, 44*, 243-256.

Schmidt, R. (1976). On the spread of semantic activation. *Psychological Research, 38,* 333-353.

Schuberth, R.E., Spoehr, K.T., & Lane, D.M. (1981). Effects of stimulus and contextual information on the lexical decision process. *Memory & Cognition, 9,* 68-77.

Schwanenflugel, P.J., & Shoben, E.J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language, 24*, 232-252.

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, and D. Burnham (eds), *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech* (pp. 85-108). Sussex, UK: Psychology Press.

Sereno, S.C., Brewer, C.C., & O'Donnell, P.J. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science, 14,* 328-333.

Seidenberg, M.S., Waters, G.S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word perception. *Memory & Cognition, 12*, 315-328.

Snodgrass, J.G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34-50.

Stanovich, K.E., & West, R.F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General, 112,* 1-36.

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

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.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645-659.

Taylor, W.L. (1953). 'Cloze procedure': a new tool for measuring readability. *Journal Quarterly, 30*, 415-433.

Tyler, L.K. (1990). The relationship between sentential context and sensory input: Comments on Connine's and Samuel's chapters. In G. Altman (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 315-323). Cambridge, MA: MIT Press.

Van Berkum, J.J.A., Brown, C., Zwitserlood, P., Kooijman V., & Hagoort, VP. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443-467.

Vroomen, J., & de Gelder, B. (2000). Crossmodal integration: A good fit is no criterion. *Trends in Cognitive Science, 4,* 37-38.

Windmann, S. (2004). Effects of sentence context and expectation on the McGurk illusion. *Journal of Memory and Language, 50*, 212-230.

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

Windmann, S., Daum, I., & Güntürkün, O. (2002). Dissociating prelexical and postlexical processing of affective information in the two hemispheres: Effects of stimulus presentation format. *Brain and Language, 80*, 269-286.

Appendix: Stimulus List

	Congruent sentences			Incongruent sentences
List A	**Fusion**	**Auditory**	**Visual**	(Lists A and B)
I prefer to take my coffee with milk and SUGAR. [ZUCKER]	/zuka/	/zupa/	/zuta/	It takes more than one swallow to make a SUMMER. (SOMMER)
I put the book back on the SHELVES. [REGAL]	/regal/	/rebal/	/rekal/	Lies have short LEGS. (BEINE).
At our wedding my father gave a touching SPEECH. [REDE]	/rede/	/rebe/	/rete/	The internal revenue examines the case closely under the MAGINFIER. (LUPE)
She was always dressed according to the newest FASHION. [MODE]	/mode/	/mobe/	/moge/	If you sit in a glass house you better not throw with STONES. (STEINE)
With fruit cake I like to take a bit of whipped CREAM. [SAHNE]	/sane/	/same/	/sade/	The last ones get bit by the DOGS (HUNDE).
The king left his empire to the eldest of his SONS. [SÖHNE]	/söne/	/söme/	/söke/	On the building of the consulate waved the American FLAG. (FLAGGE)
The huge dog growled and showed his TEETH. [ZÄHNE]	/zäke/	/zäme/	/zäte/	That thing looks good but has a serious HITCH. (HAKEN)
Uphill the biker stepped heavily into the PEDAL.* [PEDAL]	/pedal/	/pebal/	/penal/	To be or not to be, that is the QUESTION. (FRAGE)
On an orbit in space you find Mars, the Earth and every other PLANET. [PLANET]	/planet/	/plamet/	/planet/	The robber shouted: “Your money or your LIFE!”. (LEBEN)
Let’s simply dump the stuff into this CONTAINER. [TONNE]	/tone/	/tome/	/toge/	The fourth commandment honors father and MOTHER. (MUTTER)
List B				
Among the staff they did not hang the case on the large CLOCK.* [GLOCKE]	/gloke/	/glope/	/glote/	
For driving a car he was not anymore in the right POSITION.* [LAGE]	/lage/	/labe/	/lake/	
When I was a child I often went to purchase chocolate in the little Tante Emma SHOP.* [LADEN]	/laden/	/laben/	/lagen/	
He shamed himself into ground and BOTTOM.* [BODEN]	/boden/	/boben/	/boten/	
When he sat back, there was a cracking sound from the chair’s BACK. [LEHNE]	/lene/	/leme/	/lede/	
For the future we had no further PLANS. [PLÄNE]	/pläne/	/pläme/	/pläke/	
The workers were on strike for an increase of their WAGES. [LÖHNE]	/löne/	/löme/	/löte/	
His life hung on a silk THREAD.* [FADEN]	/faden/	/faben/	/fanen/	
North- and South-America are connected by the Panama-CHANNEL. [CANAL]	/kanal/	/kamal/	/kanal/	
The TV show is hosted by Hella von SINNEN.* [SINNEN]	/sinen/	/simen/	/sigen/	

Note. * indicates a well-known German name, phrase or idiom.