Research Report

SINGING IN THE BRAIN: Independence of Lyrics and Tunes

M. Besson,[1] F. Faïta,[2] I. Peretz,[3] A.-M. Bonnel,[1] and J. Requin[1]

[1]Center for Research in Cognitive Neuroscience, C.N.R.S., Marseille, France; [2]University of Bordeaux II, Bordeaux, France; and [3]University of Montreal, Montreal, Canada

Abstract—*Why is vocal music the oldest and still the most popular form of music? Very possibly because vocal music involves an intimate combination of speech and music, two of the most specific, high-level skills of human beings. The issue we address is whether people listening to a song treat the linguistic and musical components separately or integrate them within a single percept. Event-related potentials were recorded while musicians listened to excerpts from operas sung* a capella*. Excerpts were ended by semantically congruous or incongruous words sung either in or out of key. Results clearly demonstrated the independence of lyrics and tunes, so that an additive model of semantic- and harmonic-violations processing predicted the data extremely well. These results are consistent with a modular organization of the human cognitive system and open new perspectives in the search for the similarities and differences between language and music processing.*

The idea that the cognitive operations necessary to understand and to produce language are specific to the language domain (Chomsky, 1965; Pinker, 1995) is classic in language research but has recently been called into question. For instance, Seidenberg (1997), in a review of standard and current issues in language research, wrote: "Brain organization constrains how language is learned, but the principles that govern the acquisition, representation, and use of language are not specific to this type of knowledge" (p. 1603). Although this conclusion is based on a connectionist approach, other approaches aimed at discovering how language is processed by the brain have provided arguments both for and against this view. Specifically, recordings of brain electrical activity (event-related brain potentials, or ERPs) have demonstrated that violations of the structural aspects of language (e.g., syntax) and music (e.g., harmony) produce very similar brain responses. Syntactic incongruities (such as in verb-noun agreement and word order; Kutas & Hillyard, 1983; Osterhout & Holcomb, 1992) elicit a late posterior positivity (P600), with maximum amplitude around 500 to 600 ms. Similarly, incongruous ("wrong") notes or chords elicit late positive components (LPCs) that develop over posterior regions in the same latency band as the P600 (Besson & Faïta, 1995; Janata, 1995). Interestingly, results of a study explicitly designed to compare syntactic and musical incongruities reported no significant differences in the LPCs elicited by the two types of anomalies (Patel, Gibson, Ratner, Besson, & Holcomb, in press). Thus, the ERP markers of structural violations seem to be similar for language and music. Both the P600 and the LPC probably reflect the activation of a general-purpose violation detector that operates in a variety of domains and produces an outcome best known as the P300 component.[1] This view is in line with Seidenberg's conclusion that similar cognitive operations may be involved in processing the structure of language and other types of well-organized, rule-based knowledge, such as music.

In contrast, the brain response to semantic incongruity is very different from that produced by syntactic incongruities. Semantically incongruous words, whether in written or spoken sentences (e.g., "He takes coffee with cream and *sugar/dog*"), elicit an N400 (a negative component with maximum amplitude centro-parietally around 400 ms following word onset; Kutas & Hillyard, 1980) rather than a P300. If any brain activity is unique to language, it is more likely to be found at the semantic than at the syntactic level. In order to test this hypothesis directly, we compared the semantic aspects of language and the harmonic aspects of music. The main question was whether, while listening to songs, people process lyrics and tunes separately or integrate them into a unified percept. Participants were presented with brief excerpts from operas. The final word of each excerpt was either semantically congruous or incongruous with respect to the preceding linguistic context, and it was sung either in or out of key (i.e., outside the diatonic scale of the excerpt; see Fig. 1). Thus, there were four experimental conditions: The final word was (a) semantically congruous and sung in key, (b) semantically incongruous and sung in key, (c) semantically congruous and sung out of key, or (d) both semantically incongruous and sung out of key.

In summary, our goal was to analyze the time course and scalp distribution of the changes in the brain electrical activity associated with the processing of lyrics and tunes. Because very different electrophysiological markers are associated with semantically (N400) and musically (P300) incongruous events, the predictions were clear-cut. First, if words are processed in the same way whether sung or spoken, then the final incongruous words of opera excerpts would elicit an N400. Second, if melodic contour is processed in the same way whether or not accompanied by lyrics, then words sung out of key would elicit a P300. We were particularly interested in the results of the double violation (incongruous words sung out of key), as they provided a direct test of the integration-versus-independence hypothesis of semantic and musical processing in the perception of songs. Results (e.g., Serafine, Crowder, & Repp, 1984) that claim to show melody and text are integrated into a single percept suggest that ERPs should show an interaction between the N400 produced by semantic violations and the P300 produced by musical violations. In contrast, if lyrics and tunes are processed independently, ERPs should be additive (i.e., the response to a double violation should be equal to the sum of the responses to two individual violations; see Sternberg, 1969, for the additive-factor logic).

Address correspondence to Mireille Besson, Centre de Recherche en Neurosciences Cognitives, C.N.R.S., 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20+, France; e-mail: mbesson@lnf.cnrs-mrs.fr.

1. To clarify matters, we label the positive component elicited by harmonic incongruities a P300 component, even if its mean peak latency is around 800 ms.

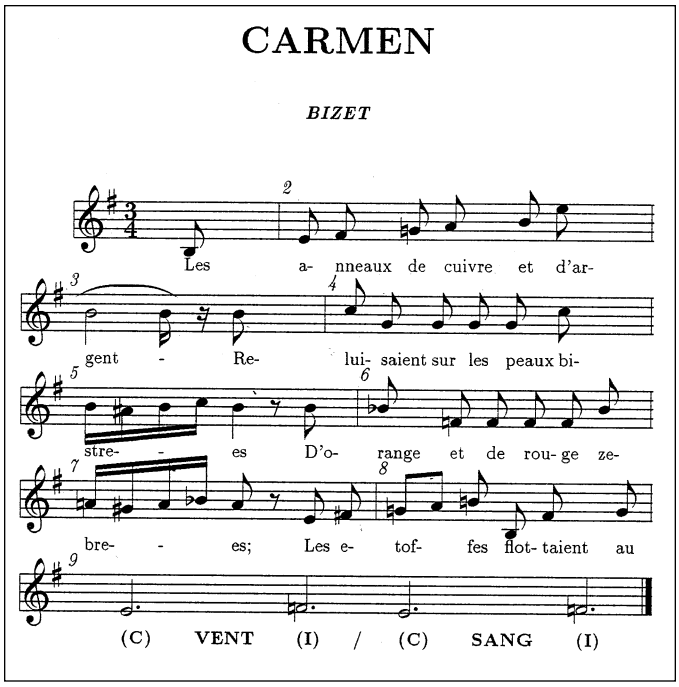

Fig. 1. Example of the materials used in the experiment. An approximate English translation of this excerpt from *Carmen* (Bizet) is, "Rings of copper and silver/ Were shining on tanned skins/Of orange and red stripes/The fabric flew in the wind." Note that in French, the incongruous word, "sang" ("blood"), rhymes with the expected completion, "vent" ("wind"). The final note is either congruous (C) or incongruous (I; out of key).

METHOD

Participants

All 16 participants (age range: 23–56 years, mean = 36 years) were professional musicians from the opera in Marseille. Participants were all right-handed, according to self-report, and had normal hearing. They were tested individually and were paid to participate in the experiment, which lasted for approximately 2 hr.[2]

Materials

Two hundred different excerpts were selected from the best known French operas. A professional singer sang each excerpt *a capella* (i.e., without instruments) in each of the four experimental conditions, so that across participants, each opera excerpt was presented in each condition (the heroic singer had to sing 800 opera excerpts). Complete musical phrases were presented so as to respect musical structure, and each excerpt lasted between 8 and 20 s.

So that ERP recordings would be time-locked to the onsets of the final words, the experimenters listened to the speech signals while viewing them on a computer screen and placed a square wave click at the onset of each of the 800 final words. These clicks were used to trigger electroencephalogram (EEG) recordings that were obtained from 19 scalp electrodes, mounted in an elastic cap and located at standard positions of the International 10/20 system.[3]

Each participant listened to 200 opera excerpts (an equal number in each experimental condition) that were presented in four blocks of 50 trials each. Excerpts were presented in pseudorandom order so that no more than 3 excerpts from the same condition were played consecutively. Congruous and incongruous words were monosyllabic and matched for frequency of occurrence and length; whenever possible, incongruous words rhymed with the expected endings.

Procedure

Participants were asked to listen carefully to each excerpt and to pay equal attention to the language and to the music in order to detect semantic and harmonic incongruities. They were not required to give a motor response, but at the end of each block, they were asked general questions regarding the difficulty of the detection and the familiarity of the excerpts to ensure that they had listened carefully. They were informed that the final word in each excerpt could be either semantically congruous or incongruous and sung in or out of key. They were also asked to avoid blinking for a 2-s period from the onset of that word until a row of four *X*s appeared on a computer screen in front of them.

RESULTS

The data are shown in Figure 2. Mean amplitudes (i.e., the integration of amplitude values over time) were measured in the latency bands of interest (50–600 ms for the N400 component and 400–1,000 ms for the P300), and repeated measures analyses of variance (ANOVAs) were carried out.[4] Compared with congruous words sung in key, incongruous words sung in key elicited significantly larger N400s that developed in the 50 to 600 ms following the final word's onset (main effect of congruity: $F[1, 15] = 11.99$, $MSE = 15.96$, $p < .003$; see Fig. 2a) and showed a widespread scalp distribution. In contrast, congruous words sung out of key elicited significantly larger P300s, with a clear parietal maximum (condition-by-electrode interaction: $F[18, 270] = 3.21$, $MSE = 3.02$, epsilon = .24, $p < .01$; see Fig. 2b). Note that the onset of the P300 was clearly later than the onset of the N400.

Of particular importance are the results for incongruous words sung out of key. These words yielded both a significantly larger N400 (condition-by-electrode interaction: $F[18, 270] = 4.23$, $MSE = 0.64$, epsilon = .29, $p < .001$) and a significantly larger P300 than congruous words sung in key (main effect of congruity: $F[1, 15] = 5.10$, $MSE = 20.12$, $p < .03$; see Fig. 2c). In order to assess additivity, we added the semantic-congruity effect (incongruous word sung in key minus

2. All participants gave their informed consent.

3. EEG was recorded for 2,200 ms, beginning 200 ms before onset of the terminal word, and amplified by an SA Instrumentation amplifier with a 0.01- to 30-Hz (half-amplitude cutoff) band pass and a 250-Hz sampling rate. Each scalp electrode was referred to an electrode placed on the left mastoid. ERPs were averaged off-line once trials with eye movements or muscle artifacts had been eliminated.

4. The Greenhouse-Geiser correction for inhomogeneity of variance was applied whenever appropriate; we report the uncorrected degrees of freedom, epsilon value, and probability level following correction.

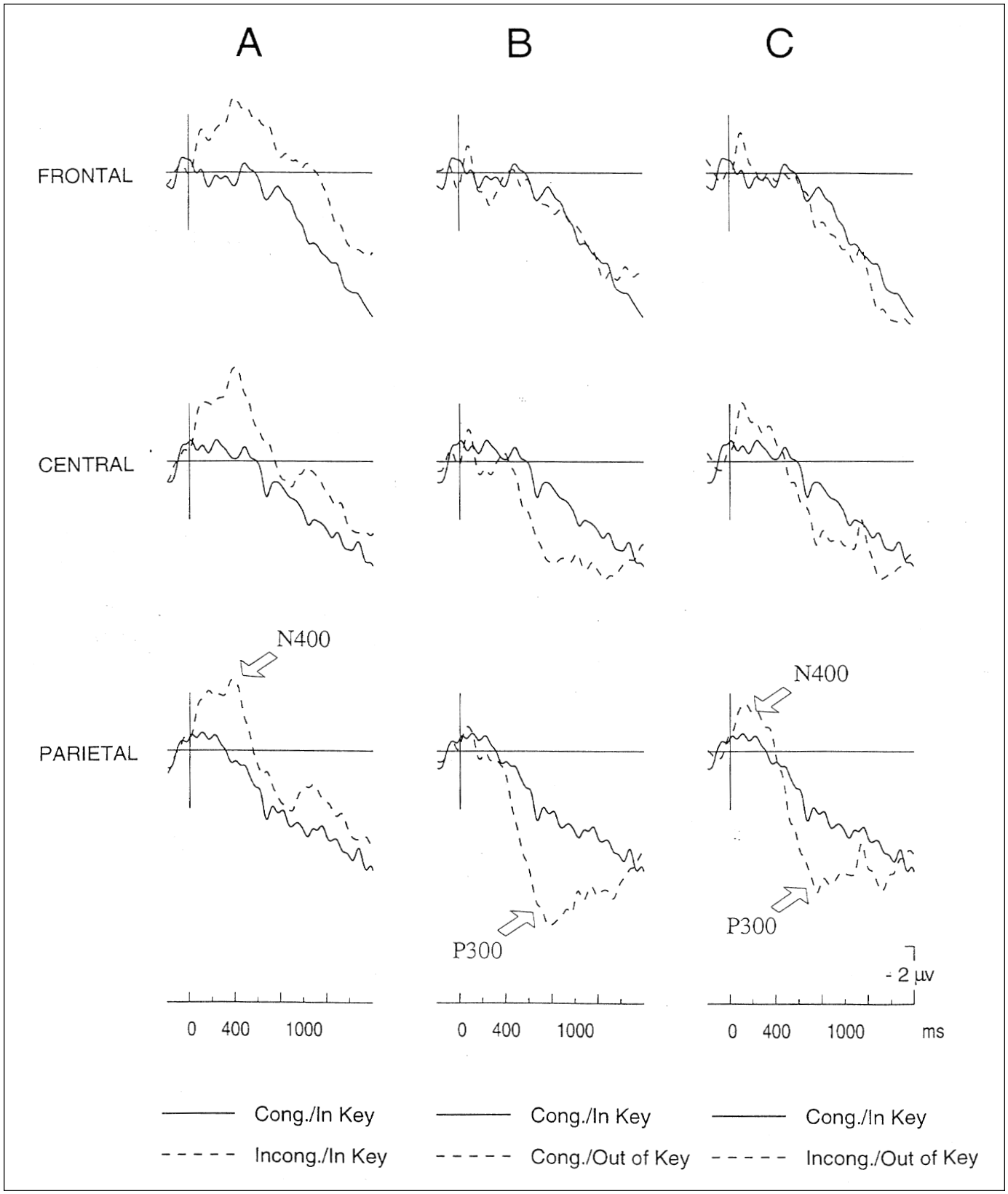

Fig. 2. Event-related potentials (ERPs; averaged over 16 professional musicians) associated with the processing of semantic, harmonic, and double incongruities. ERPs elicited by congruous ("Cong.") words sung in key (solid lines) are compared with those elicited by (a) incongruous ("Incong.") words sung in key, (b) congruous words sung out of key, and (c) incongruous words sung out of key. Data presented in this figure are from midline frontal, central, and parietal electrodes, and negative is up. The vertical bars represent the onset of the final word in the excerpts.

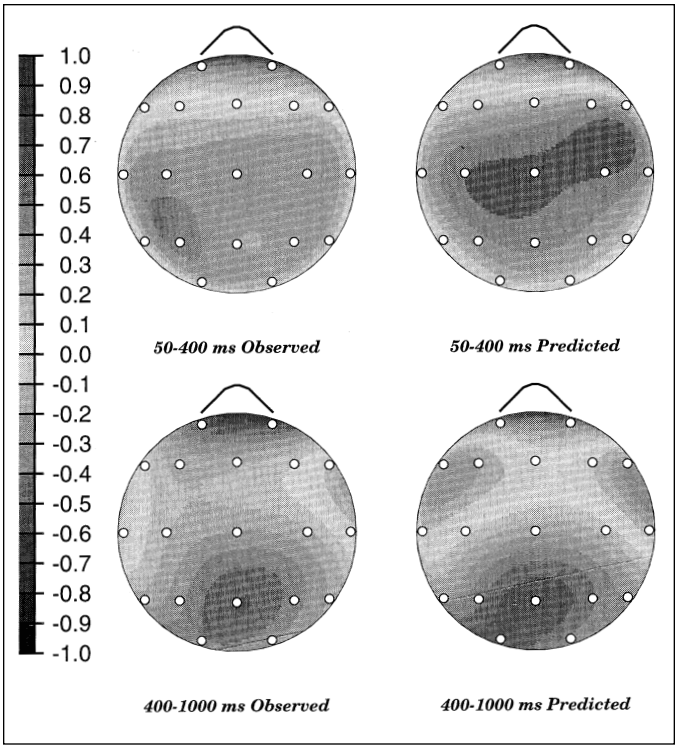

Fig. 3. Topographic maps computed as an integration of mean amplitude values across time, in specific time windows. These maps illustrate the comparison between the data recorded in the experiment (observed) and the predictions of an additive model of language and music processing (see the text for details).

congruous word sung in key) and the harmonic-congruity effect (congruous word sung out of key minus congruous word sung in key) offline and compared the results with the double-incongruity effect (incongruous word sung out of key minus congruous word sung in key). As can be seen in Figure 3, the observed double-incongruity effect and the predictions of the additive model are very similar. Statistical comparisons between the predicted and observed ERPs, in both the N400 and the P300 latency bands, did not reveal any significant differences between the two (for all results, $p > .10$).

DISCUSSION

Several important results emerge from these analyses of on-line variations in the brain's electrical activity as musicians listened to various opera excerpts. First, our results demonstrate that N400 components are elicited by semantic incongruities in sung passages in the same manner as they are elicited in speech. Thus, the processing of the semantics of the sentence is not affected by the musical structure imposed on it. Second, P300 components are associated with words sung out of key, in the same manner that they are associated with notes or chords played outside the diatonic scale. Thus, harmonic processing is not affected by the semantics of the sentence. These two findings clearly support the idea that semantic and harmonic violations are processed independently even when presented in stimuli in which the lyrics and the tunes are strongly intertwined. Such a conclusion is further supported by the results of the additivity tests, which showed that the data recorded in the double-incongruity condition were not statistically different (in both the N400 and the P300 latency bands) from the predictions of an additive model of semantic and harmonic processing. Note also that although no behavioral response was required in this experiment, the present results are consistent with those of a study in which the same materials were used and the percentage of correct responses was analyzed (Bonnel, Faïta, Besson, & Peretz, 1998). Taken together, these results do not support an interactive view of semantic and harmonic processing. Rather, they support the view that these two forms of processing are strongly independent.[5]

The ERP markers of the semantic and harmonic violations differed markedly in polarity, time course, and scalp distribution. The response to semantic violations was predominantly negative, was widespread across scalp sites, and developed in the 50- to 600-ms range. The response to harmonic violations was positive, was maximum over parietal regions, and developed in the 400- to 1,000-ms range. Note that such an early onset of the N400 in the auditory modality has previously been reported (Connolly & Phillips, 1994; Holcomb & Neville, 1991); in our experiment, it probably reflects both (a) strong coarticulation effects linked with the fact that the words were sung and (b) the fact that the words were monosyllabic and the expected and incongruous sentence completions differed in the first phoneme. In any event, it is striking that for musicians who play music 6 hr a day on average, the electrophysiological marker of semantic processing (the N400) nevertheless shows a clear temporal precedence over the electrophysiological marker of harmonic processing (the P300). However, a quick survey of the music-perception literature shows that reaction times to decide whether a note or a chord is musically appropriate within a musical sequence are, on average, much longer (around 1,200 ms; Janata, 1995) than are the reaction times to decide whether a word is semantically appropriate within a sentence context (around 800 ms).

In conclusion, although previous results have shown that lyrics and tunes in vocal music may be integrated in memory (Serafine et al., 1984), the present data clearly show evidence of their independent on-line processing, at least when the semantic and harmonic aspects are considered. It may well be that different results would obtain if syntax and prosody were contrasted with harmony, melody, or rhythm. Such comparisons will be tested in further experiments. Finally, our results are in agreement with recent neuropsychological evidence from amusic patients who were able to recognize the lyrics of familiar songs quickly and accurately but unable to recognize the corresponding melodies (Peretz et al., 1994). As is the case for so many other cognitive skills, the exquisite unity of vocal music (Risset, 1991) emerges from the concerted activity of separate processors.

Acknowledgments—We thank Emanuel Donchin, Sylvan Kornblum, Marta Kutas, Guillaume Masson, Allen Osman, and Bertram Scharf for helpful comments on earlier versions of this article. The work reported here was supported by a grant from the Groupement d'Interêt Scientifique, "Science de la Cognition" (1995–1997), to M. Besson, E. Bigand, and M. Habib.

5. Our results, as well as many others in the literature, such as the fact that every word (not only incongruous ones) elicits an N400, lead us to believe that the independence of semantic and harmonic processing is not restricted to violations but extends to congruous events as well.

REFERENCES

Besson, M., & Faïta, F. (1995). An Event-Related Potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1278–1296.

Bonnel, A.-M., Faïta, F., Besson, M., & Peretz, I. (1998). *Independent processing of lyrics and tunes in opera: Comparison of single and dual-tasks performances.* Manuscript submitted for publication.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Connolly, J.F., & Phillips, N.A. (1994). Event-Related Potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*, 256–266.

Holcomb, P., & Neville, H. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology, 19*, 286–300.

Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience, 7*, 153–164.

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

Kutas, M., & Hillyard, S.A. (1983). Event-Related brain Potentials to grammatical errors and semantic anomalies. *Memory & Cognition, 11*, 539–550.

Osterhout, L., & Holcomb, P. (1992). Event-Related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806.

Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (in press). Processing grammatical relations in language and music: An Event-Related Potential study. *Journal of Cognitive Neuroscience*.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain, 117*, 344–356.

Pinker, S. (1995). *The language instinct*. New York: Harper Perennial.

Risset, J.C. (1991). Speech and music combined: An overview. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 368–379). Cambridge, England: Cambridge University Press.

Seidenberg, M.S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599–1603.

Serafine, M.L., Crowder, R.G., & Repp, B.H. (1984). Integration of melody and text in memory for song. *Cognition, 16*, 285–303.

Sternberg, S. (1969). The discovery of processing stages: Extension of Donder's method. *Acta Psychologica, 30*, 276–315.

(RECEIVED 9/10/97; ACCEPTED 3/4/98)