

Processing Syntactic Relations in Language and Music: An Event-Related Potential Study

Aniruddh D. Patel

The Neurosciences Institute

Edward Gibson

Massachusetts Institute of Technology

Jennifer Ratner

Tufts University

Mireille Besson

CNRS-CRCN, Marseille, France

Phillip J. Holcomb

Tufts University

Abstract

■ In order to test the language-specificity of a known neural correlate of syntactic processing [the P600 event-related brain potential (ERP) component], this study directly compared ERPs elicited by syntactic incongruities in language and music. Using principles of phrase structure for language and principles of harmony and key-relatedness for music, sequences were constructed in which an element was either congruous, moderately incongruous, or highly incongruous with the preceding structural context. A within-subjects design using 15

musically educated adults revealed that linguistic and musical structural incongruities elicited positivities that were statistically indistinguishable in a specified latency range. In contrast, a music-specific ERP component was observed that showed antero-temporal right-hemisphere lateralization. The results argue against the language-specificity of the P600 and suggest that language and music can be studied in parallel to address questions of neural specificity in cognitive processing. ■

INTRODUCTION

The perception of both speech and music depends on the rapid processing of signals rich in acoustic detail and structural organization. For both domains, the mind converts a dynamic stream of sound into a system of discrete units that have hierarchical structure and rules or norms of combination. That is, both language and music have syntax (Lerdahl & Jackendoff, 1983; Sloboda, 1985; Swain, 1997). Although the details of syntactic structure in the two domains are quite different (Keiler, 1978), it is this very fact that makes their comparison useful for sifting between domain-general and domain-specific or "modular" cognitive processes. To illustrate this idea, we report a study that uses music to examine the language-specificity of a known neural correlate of syntactic processing, the P600 or "syntactic positive shift" brain potential (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992, 1993).

The P600 is a positive component of the event-related brain potential (ERP) elicited by words that are difficult to integrate structurally into meaningful sentences. For example, Osterhout and Holcomb (1992, 1993) found that in sentences of the type "The broker persuaded *to* sell the stock was sent to jail," a P600 was elicited by the word "to," relative to the same word in sentences such as "The broker hoped *to* sell the stock." The critical difference between these sentences lies in how easily the word *to* (and the following words) can be integrated with the verb. When a subject first encounters the verb "persuaded," a simple active-verb interpretation is possible (e.g., "The broker persuaded his client to sell"): This interpretation does not permit the attachment of a constituent beginning with "to." "Hoped," on the other hand, unambiguously requires a sentential complement, so the inflecting marker "to" is readily allowed. Thus, the P600 in the former sentence occurs at a time when the brain is processing a more complex syntactic relation than

might have been predicted given the preceding structural context. Evidence such as this has led several researchers to suggest that the P600 reflects reanalysis of structural relations by the human syntactic parser (e.g., Friederici & Mecklinger, 1996; Hagoort et al., 1993; but see Münte, Matzke, & Johannes, 1997). In the above example, reanalysis would involve changing the initial interpretation of “persuaded” from a simple active-verb to a more complex reduced-relative clause.¹

One can immediately see that the P600 is quite different from the better-known language ERP component, the N400, a negative-going wave that has been associated with semantic integration processes, and not with the computation of syntactic structure (Kutas & Hillyard, 1980, 1984). It is important to note, however, that the P600 is not the only (or the earliest) ERP component associated with syntactic parsing. Several researchers have reported an early left anterior negativity (LAN) at points where syntactic processing becomes demanding or breaks down (Friederici, Pfeifer, & Hahne, 1993; King & Kutas, 1995; Kluender & Kutas, 1993; Neville, Nicol, Barss, Forster, & Garret, 1991). Although the LAN is not the focus of this paper, we will have more to say about it later in the context of language-music comparisons.

For cognitive neuroscientists, the question of cognitive specificity of these (and other) neural correlates of language processing is of substantial interest. Do they reflect uniquely linguistic processes, or are they also generated by other kinds of mental activity? The answer to this question can help illuminate the nature of the neural operations underlying language functions and can yield information on the issue of informational encapsulation or modularity in language processing (Elman, 1990; Fodor, 1983).

There is debate about the degree of cognitive specificity of ERP components of language processing. For example, the N400 component has been elicited by nonlinguistic stimuli (Barrett & Rugg, 1990; Holcomb & McPherson, 1994). However, some researchers have argued that this component reflects specifically semantic (if not always linguistic) processes (Brown & Hagoort, 1993; Holcomb, 1988; Holcomb & Neville, 1990). With regard to the P600, proponents of the view that the component reflects grammatical processing have sought to distinguish it from an earlier positive component, the P300, which is typically elicited by an unexpected change in a structured sequence of events, such as a high tone in a series of low tones or a word in capital letters in a sentence of lowercase words (Picton, 1992). Recently, Osterhout, McKinnon, Bersick, and Corey (1996) directly compared the P300 and P600 in a study of orthographic and syntactic anomalies and concluded that the two components are in fact distinct. Münte, Heinz, Matzke, Wieringa, and Johannes (1998) conducted a similar study, and concluded that the P600 was not specifically related to syntactic processing. Given the

contradictory data and arguments, it is clear that the language-specificity of the P600 is still an open question.

The Current Study

The aim of the present study was to determine whether the P600 is language-specific or whether it can be elicited in nonlinguistic (but rule-governed) sequences. We chose Western European tonal music as our nonlinguistic stimulus. In much of this music there are enough structural norms that one can speak of a “grammar of music” involving the regulation of key changes, chord progressions, etc. (Piston, 1978).² Listeners familiar with this music are able to detect harmonic anomalies (i.e., out-of-key notes or chords) in novel sequences, analogously to the way competent speakers of a particular language can detect a syntactic incongruity in a sentence they have never heard before (Sloboda, 1985). A harmonic incongruity that does not have any psychoacoustical or gestalt oddness (i.e., mistuning, large jump in frequency) is a genuinely grammatical incongruity, resting on acquired knowledge of the norms of a particular musical style. One may postulate that if the P600 reflects the difficulty of structural integration in rule-governed sequences, harmonic anomalies in music should also elicit this waveform (note that our use of terms such as *anomalies* with reference to musical tones and chords is shorthand for “unexpected given the current harmonic context,” not a judgment of artistic value).

In fact, the ERP response to deviant notes in musical melodies has been recently explored by Besson and Faïta (1995), who found that harmonically and melodically deviant notes presented at the end of short musical phrases elicited a positive-going potential, with a maximum amplitude around 600 msec posttarget onset (see also Janata, 1995). These studies suggest possible links with the P600 but do not compare language and music processing directly. A cross-domain study requires a within-subjects design, in which the positivities elicited by linguistic and harmonic incongruities can be directly compared. To achieve this, we designed linguistic and musical stimuli that varied the structural “fit” between the context and the target item. These stimuli were presented to a single group of 15 participants. Musically trained subjects were selected because music perception research indicates that they are more likely to be sensitive to harmonic relations than their untrained counterparts (Krumhansl, 1990). Because we could reasonably expect our subjects to be sensitive to linguistic grammar, we wanted to ensure that they were sensitive to harmonic grammar as well. Also, we used connected speech rather than visual presentation of words to ensure that there would be no differences due to modality of stimulus presentation. The primary dependent variables used to assess the language-specificity of the P600 were the latency, polarity, and scalp distribution of brain

potentials to structurally incongruous targets in language and music. Scalp distribution is of particular interest because differences in scalp distribution of waveforms are generally taken as evidence that the underlying neural generators are not identical (see Rugg & Coles, 1995).

At the outset, we want to make clear that we draw no specific analogies between the syntactic categories of language and the harmonic structures of music. Attempts to do this (e.g., Bernstein, 1976) have generally been unfruitful (see Lerdahl & Jackendoff, 1983, for a discussion of this issue). Thus the linguistic and musical stimuli were not designed with *specific* structural parallels in mind. All that was required was that in each domain, there be some variation in the ease with which a target item could be integrated into the preceding structure; the particular principles that determined this “fit” were very different for language and music.

EXPERIMENT 1: LANGUAGE

Introduction

This experiment manipulated the structural context before a fixed target phrase so that the phrase was either easy, difficult, or impossible to integrate with the prior context. The three conditions are represented by the following sentences:

- A. Some of the senators had promoted *an old idea* of justice.
- B. Some of the senators endorsed promoted *an old idea* of justice.
- C. Some of the senators endorsed the promoted *an old idea* of justice.

The target noun phrase (an old idea) is identical in all three conditions and follows the same lexical item. However, the syntactic relations before the target vary significantly among the conditions. Condition A is a simple declarative sentence in which the noun phrase follows an auxiliary verb and a main verb. Condition B is grammatically correct but complex because the word *endorsed* is locally ambiguous between a main-verb interpretation (Figure 1) and a reduced-relative clause interpretation (Figure 2). It is well known that the main-verb interpretation is preferred in such ambiguities when both readings are plausible (Trueswell, Tanenhaus, & Garnsey, 1994; MacDonald, Pearlmutter, & Seidenberg, 1994; cf. Ferreira & Clifton, 1986 and Frazier & Rayner, 1982). That is, when subjects first encounter the verb *endorsed* in a sentence like B, they are likely to interpret it as a main verb, only changing their interpretation when subsequent words force a reduced-relative reading. This preference is accounted for by parsing theories based on structural representations (Frazier, 1978; Frazier & Rayner, 1982), thematic role assignment (Gibson, 1991;

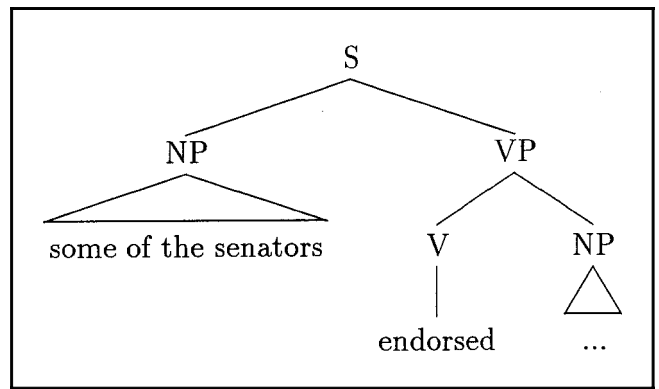


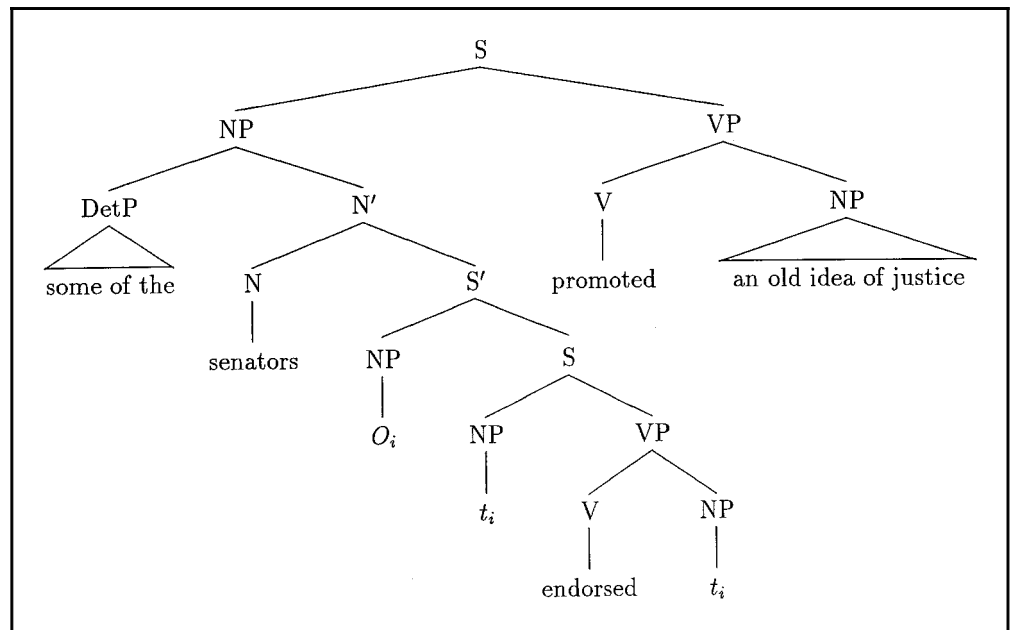
Figure 1. Syntactic structure associated with the main-verb interpretation of *endorsed* in “Some of the senators endorsed . . .” S = sentence, NP = noun phrase, VP = verb phrase, V = main verb.

Gibson, Hickok, & Schutze, 1994; Pritchett, 1988), frequency-based arguments (MacDonald et al., 1994; Trueswell & Tanenhaus, 1994), and memory and integration costs (Gibson, 1998). As a result of this parsing preference, when the target region is encountered in Condition B, subjects are engaged in syntactic reanalysis processes that were initiated upon perceiving the previous word (e.g., *promoted*). This causes syntactic integration difficulty that is not present in Condition A.

Despite the difference in structural complexity of conditions A and B, both are grammatical sentences. This stands in contrast to Condition C, where the occurrence of the target phrase renders the sentence ungrammatical. Thus the target phrase should generate the greatest integration difficulty (and thus the largest P600) in this condition.

Initially it might seem that the second verb in the sentence (e.g., *promoted*) would be the best point to designate as the target for ERP recording, because this word marks the beginning of “garden pathing” in sentence B (i.e., the point where a simple interpretation of *endorsed* as a main verb becomes untenable). However, the second verb is preceded by different classes of words in the three conditions: In A and C, it is preceded by a grammatical function word, whereas in B it is preceded by a content word. Because these word classes are known to be associated with quite different ERP signatures (Neville, Mills, & Lawson, 1992), examining the ERP at the onset of the second verb could be misleading.³ Furthermore, the point at which the second verb is recognizable *as a verb* (and thus triggers integration difficulties) is not the acoustic onset of the verb but its uniqueness point, which can be close to the word’s *end* (Marslen-Wilson, 1987). For example, in sentence B above, the word *promotions* (which would not trigger reanalysis of *endorsed*) cannot be distinguished from *promoted* till the final syllable. These factors led us to choose the onset of the word *after* the second verb as the onset of the target.

Figure 2. Syntactic structure associated with the reduced-relative clause interpretation of *endorsed* in "Some of the senators endorsed . . ."
 DetP = determiner phrase,
 N' = noun phrase projection,
 S' = clause, O_i = operator,
 t_i = trace.



Results

Subjects judged the three sentence types (main verb, reduced-relative verb, phrase-structure violation) acceptable on 95, 61, and 4% of the trials, respectively. Grand average ERP waveforms, time-locked to the first word of the target noun phrase in the three sentence types, are shown in Figure 3. Consistent with other ERP studies of connected speech (e.g., Holcomb & Neville, 1991; Osterhout & Holcomb, 1993), the early negative-positive (N1-P2) complex to the target stimuli is small in amplitude. The waveforms in the three conditions begin to diverge between 200 and 300 msec and are generally positive-going for the grammatically complex and ungrammatical sentence types. The grand averages show a hierarchy of effects: the ERP waveform to the ungrammatical sentence type is more positive than to the grammatically complex sentence type, which in turn is more positive than the grammatically simple sentence type.⁴

To assess the reliability of these differences, a repeated measures analysis of variance (ANOVA) of the mean amplitude of waveforms was conducted in three latency windows: 300 to 500 msec, 500 to 800 msec, and 800 to 1100 msec (amplitude was measured with respect to a 100-msec prestimulus baseline). Separate ANOVAs were computed for midline and lateral sites, followed by planned comparisons between pairs of conditions. Description of the results focuses on the effect of condition (sentence type): Effects involving hemisphere or electrode site are reported only if they interact with condition. Reported *p* values for all ANOVAs in this study reflect the Geisser-Greenhouse (1959) correction for nonsphericity of variance. For clarity, statistical comparisons of ERP data are organized in the following manner

within each epoch: First, results of the overall ANOVA (comparing all three conditions) are reported; this is followed by relevant pairwise comparisons. For both overall and pairwise comparisons, main effects are reported first, followed by interactions.

Effects of Sentence Type

300 to 500 Msec, Overall ANOVA. The main effect of condition was significant at midline sites ($F(2, 28) = 7.20, p < 0.01$) and marginally significant at lateral sites ($F(2, 28) = 3.03, p < 0.07$). The condition \times electrode site interactions were significant (midline: $F(4, 56) = 3.40, p < 0.04$; lateral: $F(8, 112) = 9.59, p < 0.001$), reflecting larger differences between conditions at posterior sites.

300 to 500 Msec, Comparisons. Conditions B and C were significantly more positive than A at midline and lateral sites (B versus A, midline: $F(1, 14) = 15.76, p < 0.002$; lateral: $F(1, 14) = 5.86, p < 0.03$. C versus A, midline: $F(1, 14) = 12.18, p < 0.004$; lateral: $F(1, 14) = 4.82, p < 0.05$), though B and C did not differ significantly from one another. Both conditions showed a significant condition \times electrode site interaction (B versus A, lateral: $F(4, 56) = 19.53, p < 0.001$, C versus A, midline: $F(2, 28) = 5.02, p < 0.02$; lateral: $F(4, 56) = 12.54, p < 0.001$), reflecting larger differences at posterior sites.

500 to 800 Msec, Overall ANOVA. The main effect of condition was significant at midline and lateral sites (midline: $F(2, 28) = 15.90, p < 0.001$; lateral: $F(2, 28) = 8.78, p < 0.002$). The condition \times electrode site interactions were significant (midline: $F(4, 56) = 4.26, p < 0.03$;

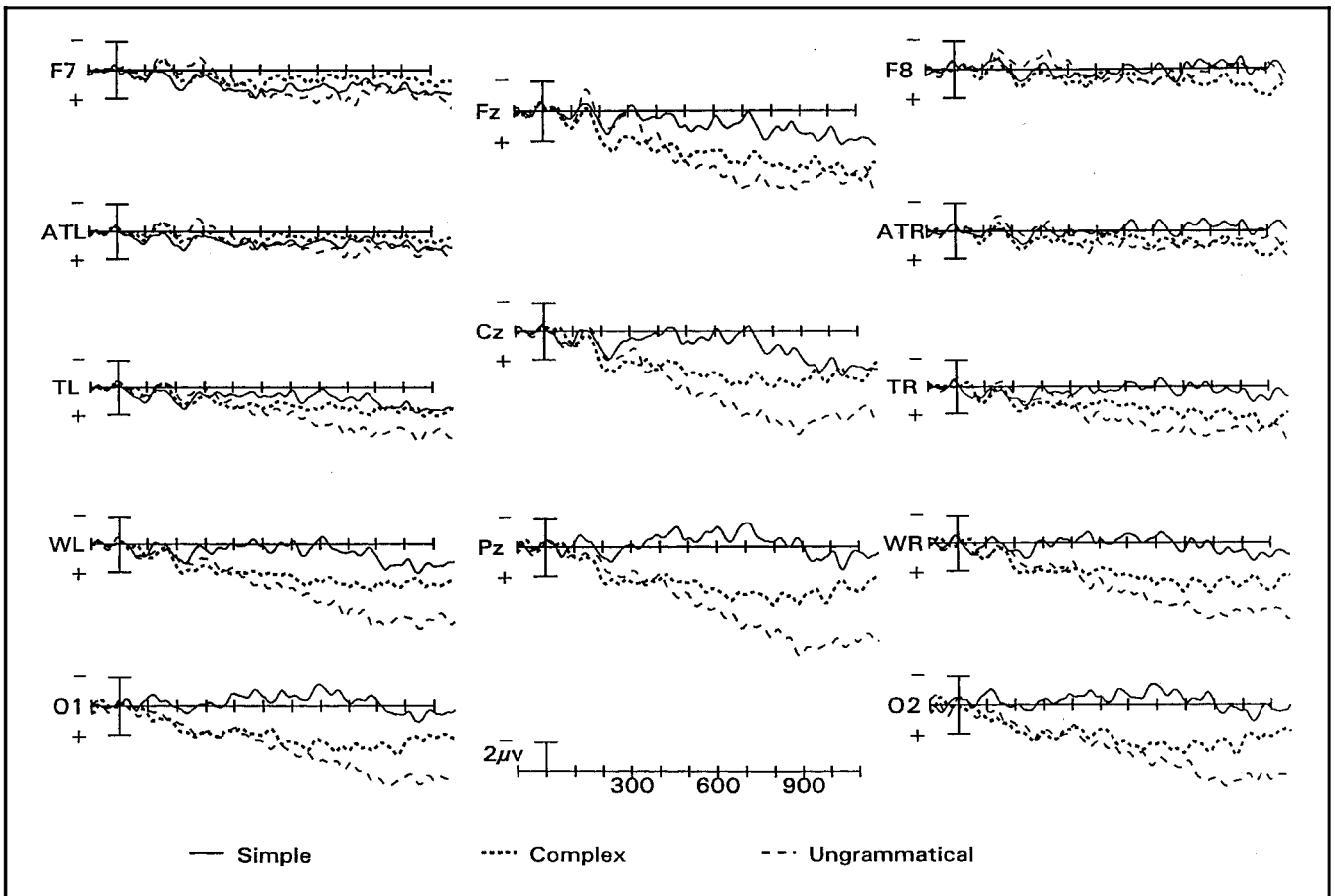


Figure 3. Grand average ERPs from 13 scalp sites time-locked to target phrases in grammatically simple, complex, and ungrammatical sentences. Each plot represents averages made over approximately 1300 trials. Recording site labels are described under ERP Recording in the "Methods" section and are shown schematically in Figure 8.

lateral: $F(8, 112) = 12.84, p < 0.001$), reflecting the greater difference between conditions at posterior sites.

500 to 800 Msec, Comparisons. Conditions B and C were significantly more positive than A at midline and lateral sites (B versus A, midline: $F(1, 14) = 19.96, p < 0.001$; lateral: $F(1, 14) = 7.07, p < 0.02$. C versus A, midline: $F(1, 14) = 41.38, p < 0.001$; lateral: $F(1, 14) = 20.66, p < 0.001$). Again, the conditions did not differ significantly from each other, although both showed significant condition \times electrode site interactions (B versus A, midline: $F(2, 28) = 4.47, p < 0.03$; lateral: $F(4, 56) = 20.42, p < 0.001$. C versus A, midline: $F(2, 28) = 11.62, p < 0.001$; lateral: $F(4, 56) = 22.94, p < 0.001$), reflecting larger differences at posterior sites generally.

800 to 1100 Msec, Overall ANOVA. The main effect of condition was significant at midline and lateral sites (midline: $F(2, 28) = 12.92, p < 0.001$; lateral: $F(2, 28) = 8.17, p < 0.004$). The condition \times electrode site interactions were significant (midline: $F(4, 56) = 7.99, p < 0.001$; lateral: $F(8, 112) = 15.56, p < 0.001$), reflecting larger differences posteriorly.

800 to 1100 Msec, Comparisons. At the midline, condition B was significantly more positive than A ($F(1, 14) = 4.90, p < 0.05$); and Condition C was significantly more positive than B ($F(1, 14) = 5.19, p < 0.04$), thus showing a significant hierarchy of effects. Laterally, these differences were marginally significant (B versus A: $F(1, 14) = 3.39, p < 0.09$; C versus B: $F(1, 14) = 3.30, p < 0.09$). For both Conditions B versus A and C versus B, the condition \times electrode site interaction was significant in both midline and lateral analyses (B versus A: midline: $F(2, 28) = 3.68, p < 0.05$; lateral: $F(4, 56) = 6.17, p < 0.007$. C versus B: midline: $F(2, 28) = 8.55, p < 0.002$; lateral: $F(4, 56) = 8.39, p < 0.002$), reflecting greater differences between the three conditions at posterior sites.

Hemispheric Asymmetries

The $C > B > A$ hierarchy is posterior in nature and symmetrical across the hemispheres. Left and right anterior sites show different reorderings of this hierarchy in the 800 to 1100-msec range, but these reorderings are not significant ($F7, F8: F(2, 28) = 0.78, p = 0.47$; $ATL, ATR: F(2, 28) = 0.15, p = 0.84$).

Summary and Discussion

In grammatically complex or ungrammatical sentences, target phrases were associated with a positive-going ERP component with a maximum around 900 msec posttarget onset. In the range of greatest component amplitude (800 to 1100 msec), a significant hierarchy of effects was observed between conditions. In sentences with a simple syntactic context before the target, no P600 was observed, whereas a small but significant P600 was observed when the preceding context was grammatically complex. This suggests that complex syntax has a cost in terms of the process of structural integration. The largest P600 was observed when the task of structural integration was impossible, due to the occurrence of a phrase-structure violation. Thus overall, the amplitude of the P600 appears (inversely) related to how easily a linguistic element fits into an existing set of syntactic relations.

We note that the P600 in our study reached maximum amplitude between 800 and 900 msec posttarget onset, a latency that is consistent with an earlier study of parsing using connected speech (Osterhout & Holcomb, 1993). The P600 was originally reported in studies using

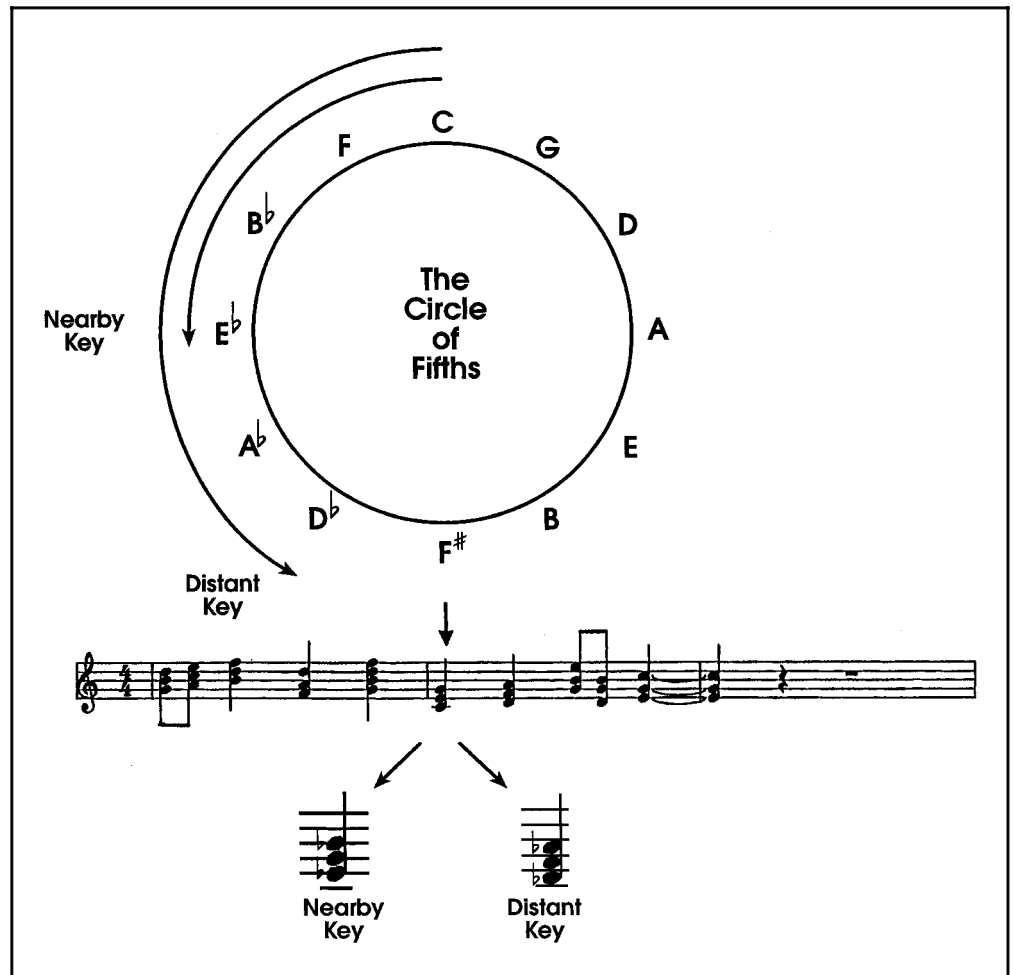
visual presentation of individual words: This component's longer peak latency in auditory language experiments may occur because in connected speech, the words are not separated by intervening silence and/or because the rapidity of spoken language caused a brief lag between the reception of the physical sounds of words and their grammatical analysis in these complex sentences.

EXPERIMENT 2: MUSIC

Introduction

This experiment manipulated a target chord in a musical phrase so that the target was either within the key of the phrase or out of key. If out of key, the chord could come from a "nearby" key or a "distant" key. In no case was an out-of-key target chord inherently deviant (e.g., mistuned) or physically distant in frequency from the rest of the phrase: rather, it was deviant only in a harmonic sense, based on the structural norms of Western European tonal music. A typical musical phrase used in this experiment is shown in Figure 4. The phrase is in the key of C major, and the target chord, which is the chord at the beginning of bar 2, is in the key of the

Figure 4. A musical phrase used in this experiment is shown in the middle of the figure. The phrase consists of a series of chords in a given key; the phrase shown is in C major. The target chord (shown by the downward pointing vertical arrow) is in the key of the phrase. The circle of fifths, shown at the top of the figure, is used to select nearby-key and distant-key target chords for a phrase in a particular key. In the case of a phrase in C major, the nearby-key chord is E-flat, and the distant-key chord is D-flat.



phrase. Across the entire set of phrases used in this experiment, all 12 major keys were represented.

Any given phrase was used with one in-key target chord and two out-of-key target chords. These targets were selected based on a music-theoretic device called the “circle of fifths.” The circle of fifths represents the distance between keys as distance around a circle upon which the 12 musical key names are arranged like the hour markings on a clock (Figure 4, top). Adjacent keys differ by one sharp or flat note (i.e., by a single black key on a piano): Distant keys differ by many sharp and flat notes. Empirical work in music perception has shown that musically trained listeners perceive differences between keys in a manner analogous to their distances on the circle of fifths: chords from keys further apart on the circle of fifths are perceived as more different than chords from nearby keys (Bharucha & Krumhansl 1983; Bharucha & Stoeckig 1986, 1987). Bharucha and Stoeckig (1986, 1987) have demonstrated that these perceived differences reflect harmonic, rather than psychoacoustic, differences. This provided a systematic way to create a hierarchy of harmonic incongruity for target chords in a phrase of a particular key. The most congru-

ous target chord was always the “tonic” chord (principal chord in the key of the phrase). Out-of-key targets were chosen using the following rules: Choose the “nearby” out-of-key chord by moving three steps counterclockwise on the circle of fifths and taking the principal chord of that key; choose the “distant” out-of-key chord by moving five steps counterclockwise. For example, for a phrase in the key of C major the in-key chord would be the C-major chord (c-e-g), the nearby-key chord would be the E-flat major chord (e-flat, g, b-flat), and the distant-key chord would be the D-flat major chord (d-flat, f, a-flat). (For more details on stimulus construction, see the “Methods” section.)

Results

Subjects judged musical phrases in the three conditions (same-key, nearby-key, and distant-key target chord) acceptable on 80, 49, and 28% of the trials, respectively. Grand average ERP waveforms, time-locked to the onset of the target chord, are shown in Figure 5. A salient negative-positive complex appears in the first 300 msec after the onset of the target chord across conditions.

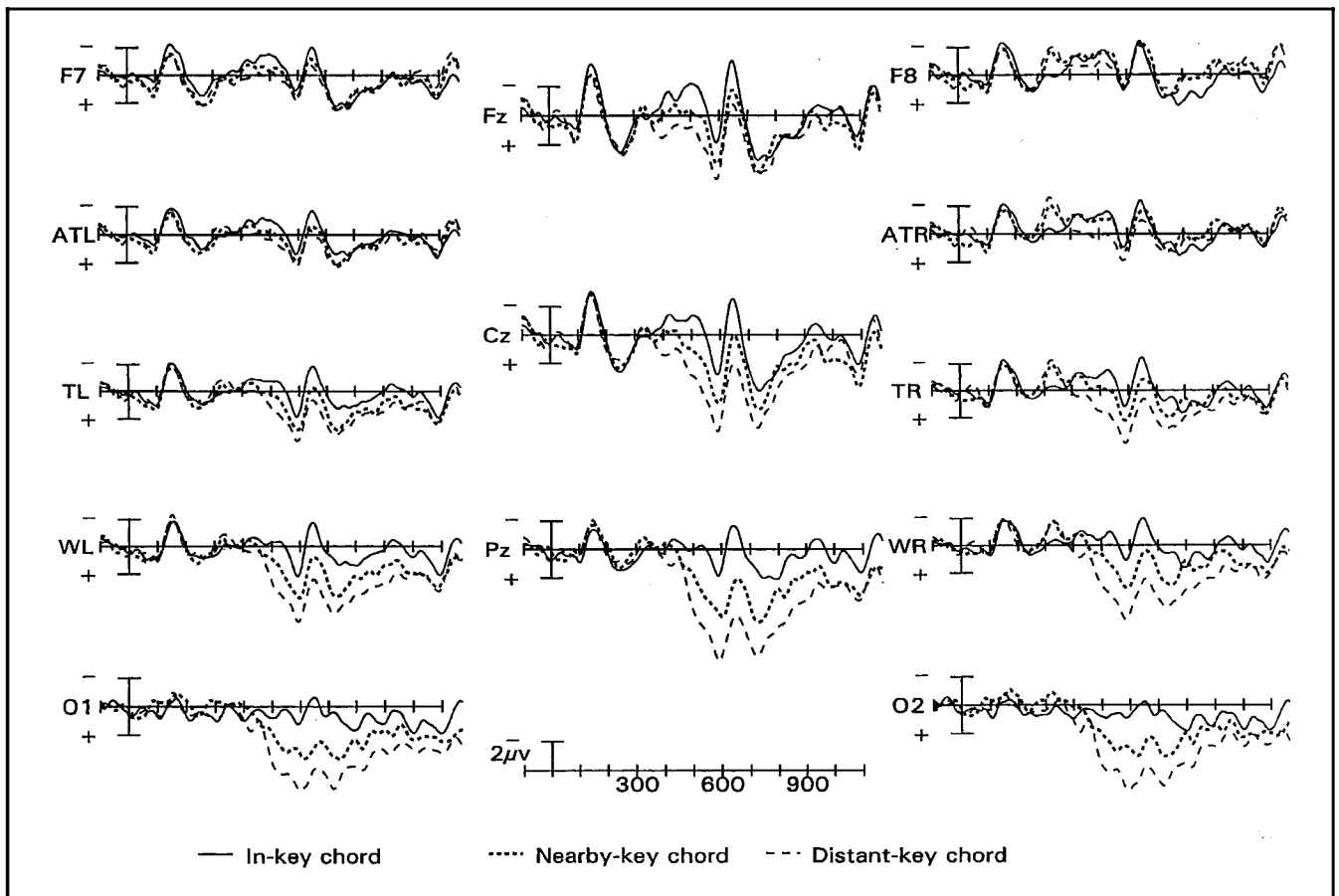


Figure 5. Grand average ERPs time-locked to the three target chord types in a phrase of a given musical key. Each plot represents averages made over approximately 1600 trials. Note that the temporal epoch of the figure also includes the chord following the target chord (onset of second chord is 500 msec after target onset).

These early “exogenous” components reflect the physical onset of the target stimulus and show little variation across stimulus type. Both the negative (N1) and positive (P2) portions of this early complex tend to have maximum amplitude anteriorly. A similar complex occurs at about 600 msec posttarget, due to the onset of the following chord. The effects of this second N1-P2 complex are superimposed on the late “endogenous” components elicited by the target chord, which begin to diverge between 300 and 400 msec and are generally positive-going for the two out-of-key targets. These late positive components show a hierarchy of effects: ERPs to the distant-key target are more positive than ERPs to the nearby-key targets, and this difference appears maximal around 600 msec. Although these positivities are largely symmetric across the hemispheres, a notably asymmetric ERP component is seen between 300 and 400 msec in the right hemisphere for the out-of-key chords.

To assess the reliability of the differences between the late positivities, a repeated measures ANOVA of the mean amplitude of waveforms was conducted in three latency windows: 300 to 500 msec, 500 to 800 msec, and 800 to 1100 msec (amplitude was measured with respect to a 100-msec prestimulus baseline). The same analyses were performed as in Experiment 1, and the results are reported in the same manner. Condition A refers to in-key target chords, and Conditions B and C refer to nearby-key and distant-key target chords, respectively.

Effects of Target Type

300 to 500 Msec, Overall ANOVA. There was no main effect of condition at midline or lateral sites, but lateral sites showed a significant interaction of condition \times hemisphere \times electrode site ($F(8, 112) = 4.91, p < 0.004$). Inspection of the waveform suggested that this three-way interaction was related to a brief negative peak at right frontal and temporal sites between 300 and 400 msec for the two out-of-key target chords (Conditions B and C). This peak was further investigated in follow-up analyses (see “Hemispheric Asymmetries,” below).

500 to 800 Msec, Overall ANOVA. The main effect of condition was significant at midline and lateral sites (midline: $F(2, 28) = 14.78, p < 0.001$; lateral: $F(2, 28) = 14.60, p < 0.001$). The condition \times electrode site interactions were significant (midline: $F(4, 56) = 5.23, p < 0.02$; lateral: $F(8, 112) = 8.09, p < 0.005$), reflecting the greater difference between conditions at posterior sites.

500 to 800 Msec, Comparisons. Condition B was significantly more positive than A at both midline and lateral sites (midline: $F(1, 14) = 12.69, p < 0.005$; lateral: $F(1, 14) = 8.69, p < 0.02$), and Condition C was significantly more positive than B at both midline and lateral sites (midline: $F(1, 14) = 6.26, p < 0.03$; lateral: $F(1, 14) = 7.98, p < 0.02$), thus showing a significant hierarchy of effects.

For B versus A, the condition \times electrode interaction was marginally significant in both midline and lateral analyses (midline: $F(2, 28) = 2.65, p < 0.09$; lateral: $F(4, 56) = 4.12, p < 0.06$), due to greater differences between Conditions B and A at posterior sites. For C versus B the condition \times electrode interaction was significant in both midline and lateral analyses (midline: $F(2, 28) = 5.45, p < 0.02$; lateral: $F(4, 56) = 6.68, p < 0.01$), reflecting the greater difference between Conditions C and B at posterior sites.

800 to 1100 Msec, Overall ANOVA. The main effect of condition was marginally significant at midline sites only ($F(2, 28) = 2.96, p < 0.07$). The condition \times electrode site interactions were significant (midline: $F(4, 56) = 4.69, p < 0.02$; lateral: $F(8, 112) = 5.93, p < 0.005$), due to greater differences between conditions posteriorly.

800 to 1100 Msec, Comparisons. C was significantly more positive than A at midline sites ($F(1, 14) = 5.30, p < 0.04$), and a significant condition \times electrode site interaction reflected the greater difference between Conditions C and A at posterior sites generally (midline: $F(2, 28) = 10.21, p < 0.03$; lateral: $F(4, 56) = 16.78, p < 0.01$).

Hemispheric Asymmetries

A hemispheric asymmetry developed in the 300 to 500-msec window, as revealed by the significant condition \times hemisphere \times electrode site interaction in the overall ANOVA for this time range. Follow-up comparisons showed that this interaction was also present in the 300 to 400-msec range ($F(8, 112) = 5.29, p < 0.01$), where a negative peak (N350) can be seen at frontal and temporal sites in the right hemisphere to the two out-of-key target chords. Pairwise analyses between conditions during this time window revealed that the three-way interaction was significant for A versus C ($F(4, 56) = 8.79, p < 0.01$) and marginally significant for A versus B ($F(4, 56) = 3.30, p < 0.07$) and B versus C ($F(4, 56) = 3.01, p < 0.07$). These data suggest a specifically right-hemisphere effect for the two out-of-key chords. In contrast to this right hemisphere N350, later positive components were quite symmetrical and showed no significant interactions of condition \times hemisphere in any of the latency windows.

Summary and Discussion

Musical sequences with out-of-key target chords elicited a positive ERP component with a maximum around 600 msec posttarget onset. This result is consistent with Besson and Faïta’s study (1995) of ERP components elicited by harmonically incongruous tones in musical melodies. In the current study, a hierarchy of ERP effects is evident between 500 and 800 msec at both midline and lateral sites: here the ERP waveform was significantly different in positivity for the three types of target chords,

in the order: distant-key > nearby-key > same-key. These differences were greatest at posterior sites and decreased anteriorly and were symmetrical across the two hemispheres.

One notable feature of the nearby-key and distant-key target chords is that they did not differ in the number of out-of-key notes they introduced into the phrase. For example, for a phrase in the key of C-major, the nearby-key target was an E-flat major chord (e-flat, g, b-flat), whereas the distant-key target was a D-flat major chord (d-flat, f, a-flat). This provided the desirable quality that the two out-of-key chords had the same number of “accidental” (flat or sharp) notes, ensuring that any difference in positivities elicited by these chords was not due to a difference in *number* of out-of-key notes but rather to the harmonic *distance* of the chord as a whole from the native key of the phrase. The fact that the two out-of-key chords elicited late positivities of significantly different amplitude, depending on their harmonic distance from the prevailing tonality, suggests that the positivity indexed the difficulty of fitting a given chord into the established context of harmonic relations.

A new and interesting result is the observation of a brief right-hemisphere negativity in response to out-of-key target chords (N350). Previous studies of music perception using ERPs have often commented on the difference between negativities produced by violations of semantic expectancy in language (e.g., the N400) and positivities produced by violations of musical expectancy (e.g., Besson & Macar, 1987; Paller, McCarthy, & Wood, 1992; see Besson, 1997 for a review). Although the N350 does not resemble the semantic N400 (which differs in symmetry, duration, and scalp distribution), it is interestingly reminiscent of another ERP component recently associated with syntactic processing, the left anterior negativity, or LAN. The LAN, whose amplitude tends to be largest in the vicinity of Broca’s area in the left hemisphere, has been associated with violations of grammatical rules (Friederici & Mecklinger, 1996; Neville et al., 1991) and with an increase in working memory associated with the processing of disjunct syntactic dependencies (King & Kutas, 1995; Kluender & Kutas, 1993). Like the LAN, the N350 shows a significant condition \times hemisphere \times electrode site interaction in statistical analyses, reflecting an anterior-posterior asymmetry in its distribution. Unlike, the LAN, however, the N350 has an antero-temporal distribution and should thus perhaps be called the “right antero-temporal negativity,” or RATN. It is tempting to speculate that the RATN reflects the application of music-specific syntactic rules or music-specific working memory resources, especially because right fronto-temporal circuits have been implicated in working memory for tonal material (Zattore, Evans, & Meyer, 1994). However, more accurate characterization of this component will have to await future investigation. Here we simply note that the elicitation of this component in our study, in contrast to other recent

studies of music perception using ERPs (e.g., Besson & Faïta, 1995, Janata, 1995), may be due to our use of musical phrases with chordal harmony and sequence-internal (versus sequence-final) targets.

COMPARISON OF LANGUAGE AND MUSIC EXPERIMENTS

A fundamental motivation for conducting the language and music experiments was to compare the positive brain potentials elicited by structural incongruities in the two domains. This provides a direct test of the language-specificity of the P600 component. We statistically compared the amplitude and scalp distribution of ERP components in language and music twice: once at moderate levels of structural incongruity and once at high levels.

The time-varying shape of the waveforms to linguistic and musical targets appears quite different because musical targets elicit a negative-positive (N1-P2) complex 100 msec after onset, and once again 500 msec later, due to the onset of the following chord. Target words, on the other hand, show no clear N1-P2 complexes. The salient N1-P2 complexes in the musical stimuli are explained by the fact that each chord in the sequence was temporally separated from the next chord by approximately 20 msec of silence, whereas in speech, the sounds of the words are run together due to coarticulation. This temporal separation of chords may also explain why the positive component peaked earlier for the music than for the language stimuli (approximately 600 versus 900 msec).

To compare the effect of structural incongruity in language and music, difference waves were calculated in each domain for Conditions B-A and C-A. By using difference waves, modality-specific effects (such as the large N1-P2 complexes to musical chords) are removed, leaving only those effects due to differences between conditions. These waveforms are shown in Figures 6 and 7.

Linguistic and musical difference waves were compared in the latency range of the P600: a repeated-measures analysis of variance was conducted between 450 and 750 msec for both B-A and C-A comparisons, with domain as a main variable. Of particular interest were interactions of domain and electrode site because differences in scalp distribution would suggest different source generators for the positive waveforms in the two domains.

For the B-A comparison, there was no main effect of domain at either midline or lateral sites (midline: $F(1, 14) = 2.37, p = 0.15$; lateral: $F(1, 14) = 0.08, p = 0.78$), nor were there any significant domain \times electrode site interactions (midline: $F(2, 28) = 0.02, p = 0.93$; lateral: $F(4, 56) = 0.72, p = 0.45$). For the C-A comparison, there was no main effect of domain at either midline or lateral sites (midline: $F(1, 14) = 0.36, p = 0.56$; lateral: $F(1, 14) = 0.02, p = 0.89$), and no interactions of domain with

electrode site (midline: $F(2, 28) = 0.61, p = 0.50$; lateral: $F(4, 56) = 0.05, p = 0.91$).⁵ Thus in the latency range of the P600, the positivities to structurally incongruous elements in language and music do not appear to be distinguishable.

GENERAL DISCUSSION

A primary goal of this study was to determine if the positive-going ERP waveform observed to syntactic incongruities in language (the P600, or “syntactic positive shift”) reflected uniquely linguistic processes or indexed more general cognitive operations involved in processing structural relations in rule-governed sequences. To test this idea, we examined ERP waveforms to linguistic and musical elements that were either easy, difficult, or very difficult to integrate with a prior structural context. We chose music because like language, it involves richly structured sequences that unfold over time. In Western European tonal music, a musically experienced listener can detect a harmonic incongruity in a previously unheard sequence, implying that the listener has some grammatical knowledge about tonal structure (Krumhansl, 1990). We took advantage of this fact to examine the language specificity of the P600: if the same group of listeners, grammatically competent in both language and music, showed similar waveforms to structurally incongruous elements in both domains, the language-specificity of the P600 would be called into question.

The principal finding was that the late positivities elicited by syntactically incongruous words in language and harmonically incongruous chords in music were statistically indistinguishable in amplitude and scalp distribution in the P600 latency range (i.e., in a time window centered about 600 msec posttarget onset). This was true at both moderate and high degrees of structural anomaly, which differed in the amplitude of elicited positivity. This strongly suggests that whatever process gives rise to the P600 is unlikely to be language-specific.

It is notable that a similar effect of structural incongruity is found in the two domains despite the fact that very different principles were used to create the incongruities. The syntactic principles of language used to construct stimuli in this study had no relationship to the harmonic principles of key-relatedness used in designing the musical stimuli. Furthermore, the linguistic incongruities were generated by manipulating the structure of a context before a fixed target, whereas the musical incongruities were based on manipulating the identity of a target in a fixed context. Thus, in both domains sequences varied in structural “fit” between context and target, but this variation was based on very different rules and achieved in quite different ways. Despite these differences, a P600 effect was obtained in both conditions, suggesting that this ERP component reflects the operation of a mechanism shared by both linguistic and musical processes.

An interesting and unexpected subsidiary finding of this study was that harmonically unexpected chords in music elicited a right antero-temporal negativity, or RATN, between 300 and 400 msec posttarget onset. Although different from the better known negativities elicited by language stimuli (e.g., the semantic N400), the RATN is somewhat reminiscent of the left anterior negativity, or LAN, a hemispherically asymmetric ERP component associated with linguistic grammatical processing (cf. “Discussion” section for music experiment, above). One notable difference between the RATN and the LAN is that the RATN is quite transient, whereas the LAN is relatively long lasting. Some have suggested that the LAN may actually be two components (which do not always co-occur): an early phasic negativity signaling the detection of syntactic violations and a later, longer negativity associated with increased working memory load. If this is so, the RATN would be more comparable to the early component of the LAN. In any case, the fact that linguistic and musical syntactic incongruities elicit negativities of opposite laterality suggests that these brain potentials reflect cognitively distinct (yet perhaps analogous) operations. Future studies comparing the two components should use a larger number of trials to increase the signal-to-noise ratio: one advantage of this would be the ability to determine if the amplitude of these components can be modulated by sequence structure.

Returning to the P600, we may ask, if the same mechanism accounts for the observed positivities in language and music, how does one account for the earlier onset and slower decay of positivities to linguistic versus musical targets (cf. Figures 3 and 5)? The earlier onset in language is perhaps due to the commencement of integration difficulties slightly before the onset of the target. In Condition B, integration difficulties should begin once the uniqueness point of the pretarget word was reached (see “Introduction” to language experiment). In Condition C, the target phrase was preceded by a verb that might be perceived as ungrammatical given the preceding context, which could trigger the start of structural integration problems. The musical targets, in contrast, were always preceded by chords that fit perfectly in the tonality of the phrase up to that point, and thus there is no reason to expect lingering positivities from preceding chords at target onset. This may explain the longer latencies of the onset of positivity to musical targets.

There still remains the issue of the slower decay of positivities to linguistic targets, which can be clearly seen in the difference waves of Conditions B-A and C-A in language versus music (Figures 6 and 7). A reason for this difference is suggested by comparing the B-A difference with the C-A difference in language. Inspection of these waveforms reveals that the B-A difference wave is returning to baseline at the end of the recording epoch, (around 1100 msec posttarget onset), whereas the C-A difference is still quite large at many electrode sites. If the positivities observed in this study are an index of

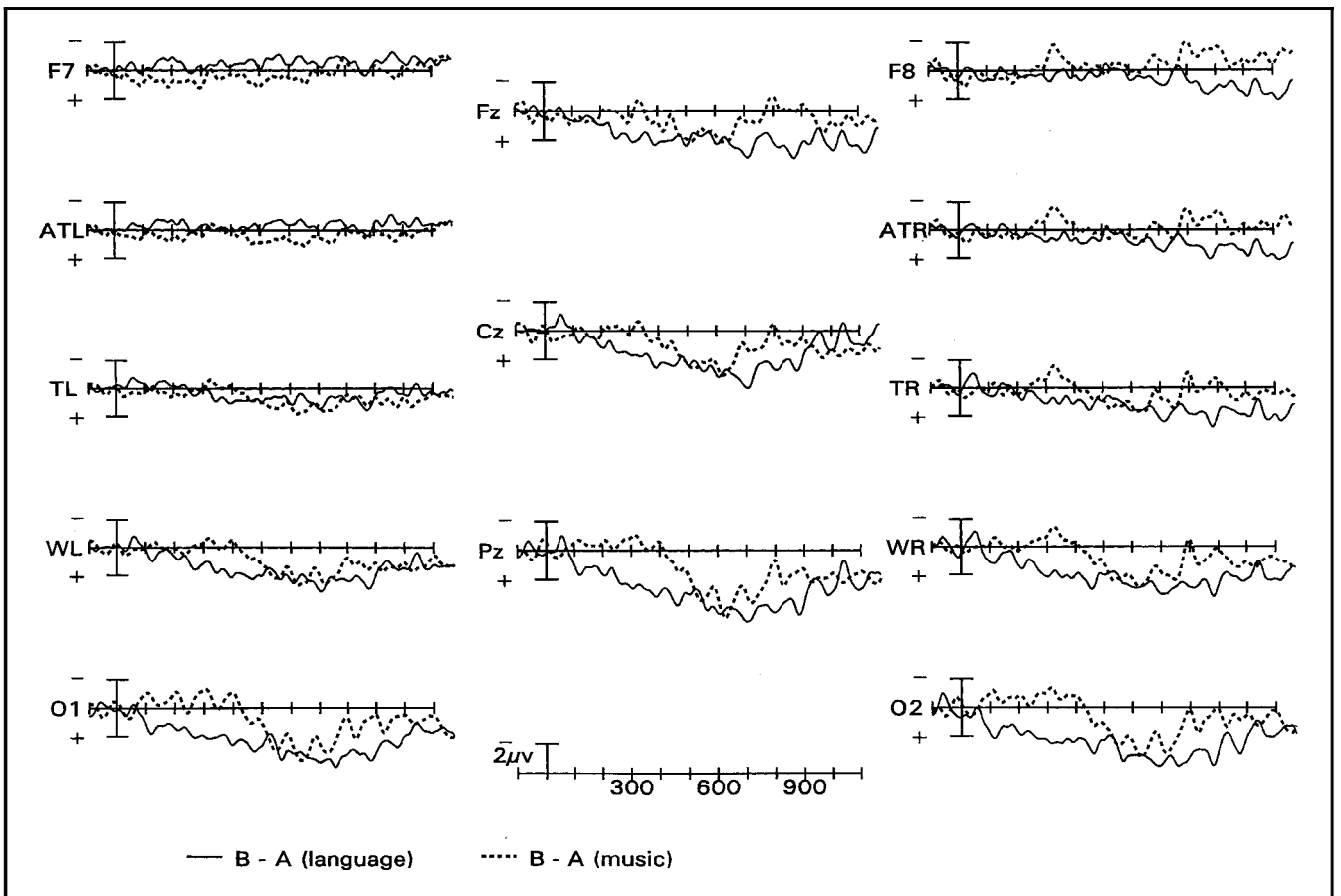


Figure 6. Difference waves for Condition B-A in language and music. The solid line represents the target phrase in a grammatically complex sentence minus the target phrase in a grammatically simple sentence. The dashed line represents the nearby-key target chord minus the in-key target chord.

structural “fit” between context and target, then the diminishing positivity to B relative to A at the end of the recording epoch may reflect successful integration of grammatically complex information into the sentence, while the continuing positivity of C relative to A reflects the impossibility of grammatically integrating the target phrase *and anything following it* with the preceding context. In contrast, in musical phrases with incongruous targets there is a return to a sensible harmonic structure immediately following the target: the structural rupture is brief, and thus the positivities decay relatively quickly. This illustrates an interesting difference between language and music: a stray structural element in language (e.g., “the” in sentence type C) can throw off the syntactic relations of preceding and following elements, whereas in music a stray element can be perceived as an isolated event without making subsequent events impossible to integrate with preceding ones.

At least two key questions remain. If the P600 is not the signature of a uniquely linguistic process, what underlying process(es) *does* it represent? Our data suggest that the P600 reflects processes of knowledge-based structural integration. We do not know, however, whether the component *uniquely* reflects knowledge-

based processes. Presumably structural integration processes are also involved in the P300 phenomenon, in which physically odd elements (such as a rare high tone in a series of low tones) elicit a positive-going waveform of shorter latency (but similar scalp distribution) to the P600. The P600 could be a type of P300 whose latency is increased by the time needed to access stored structural knowledge. Only further studies can resolve this issue (see Osterhout et al., 1996, and Münte et al., (1998) for recent studies with opposite conclusions). Whatever the relation of P600 and P300, we may conclude from our results that the P600 does not reflect the activity of a specifically linguistic (or musical) processing mechanism.

A second key question is the following: If structural integration processes in language and music engage similar neural resources, how is it possible that the perception of musical harmonic relations can be selectively impaired after brain lesion without concomitant syntactic deficits (Peretz, 1993; Peretz et al., 1994)? In our view, such individuals have suffered damage to a domain-specific knowledge base of harmonic relations and not to structural integration processes per se. The deficit results from an inability to access musical harmonic

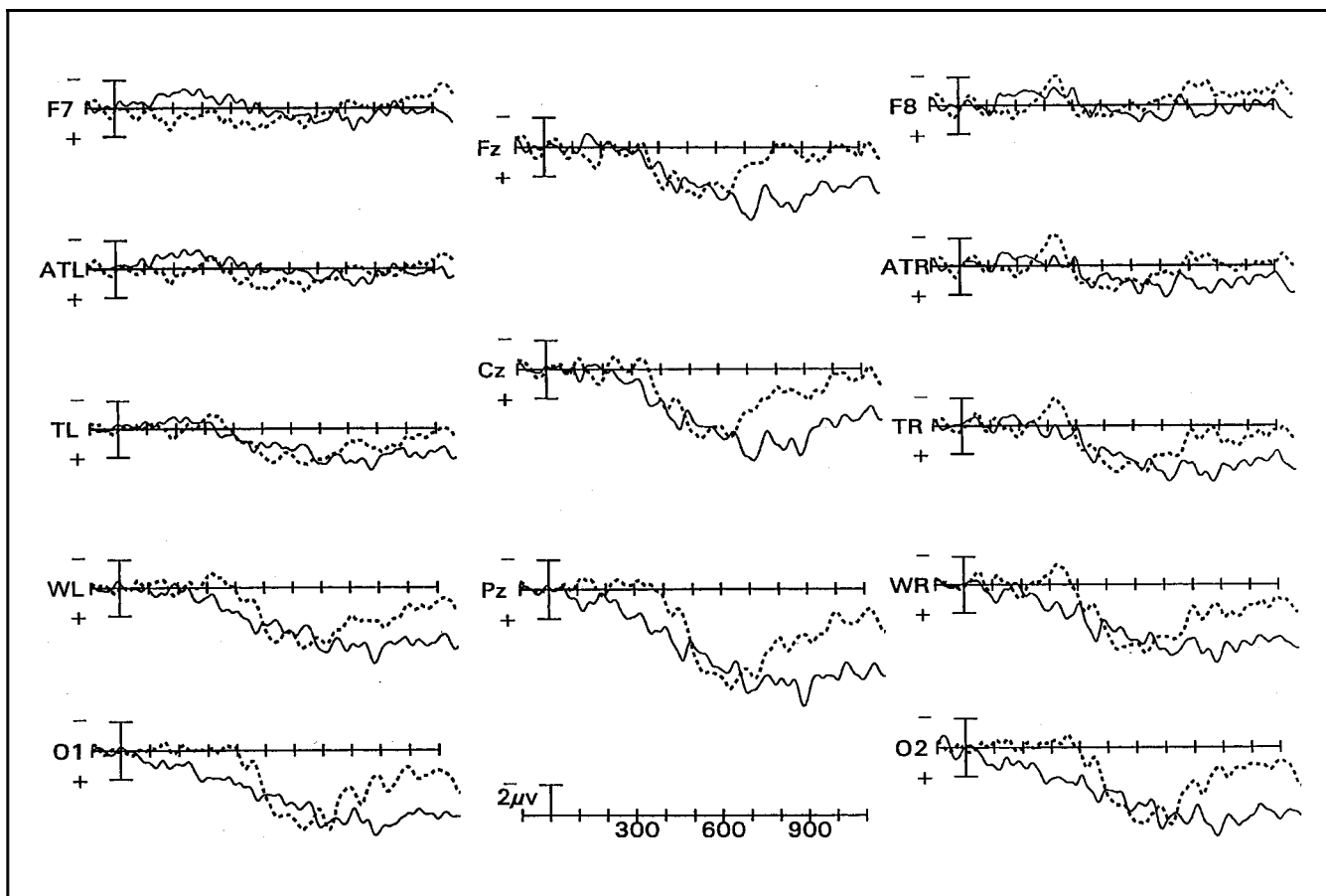


Figure 7. Difference waves for Condition C-A in language and music. The solid line represents the target phrase in an ungrammatical sentence minus the target phrase in a grammatically simple sentence. The dashed line represents the distant-key target chord minus the in-key target chord.

knowledge rather than a problem with integration itself. Consistent with this idea, selective deficits of harmonic perception have been associated with bilateral damage to temporal association cortices (Peretz, 1993; Peretz et al., 1994), which are likely to be important in the long-term neural representation of harmonic knowledge.

Identifying neural correlates of syntactic processing in language is important for the study of neurolinguistic mechanisms and for a better understanding of the treatment of language disorders, particularly aphasia. As neural studies of syntactic processing proceed, the question of specificity to language should be kept in mind because language is not the only domain with syntax. Appropriate control studies using musical stimuli can help illuminate the boundaries of specifically neurolinguistic and neuromusical processes.

EXPERIMENT 1: METHOD

Subjects

Fifteen adults between 18 and 35 years of age (mean: 24.1) served as subjects. All were right-handed native

speakers of English and all were college or graduate students recruited from the Boston area.

Stimuli

The stimuli for this experiment consisted of 30 sets of spoken sentences. Each set contained three sentences (grammatically simple, grammatically complex, and ungrammatical), which differed by the preceding structural context before a fixed target phrase (for details of syntactic structure, see the introduction to Experiment 1). Each sentence set used different verbs and target noun-phrases. Within a set, the first noun of each sentence was always varied. In addition to these 90 sentences, 60 filler sentences were included. These were designed to prevent subjects from predicting the acceptability of a sentence based on its pretarget context (for example, the filler "Some of the dignitaries had departed an old idea of justice" is nonsensical despite the use of the verb "had," a verb always used in Condition A). The final stimulus list consisted of 150 spoken sentences (randomly ordered), divided into five blocks of 30 sentences.⁶

The sentences were recorded from an adult female speaker at a rate of approximately five syllables per second and digitized at 12,500 Hz (12-bit resolution, 5000-Hz low-pass filter) for acoustic segmentation based on phonetic cues in a broad-band spectrographic display (Real Time Spectrogram, Engineering Design, Belmont, MA). Sentences had a duration of about 3.5 sec.⁷ Sentences were segmented at points of interest, such as the onset of the target phrase, allowing the placement of event codes necessary for ERP averaging. In the experiment, segments were seamlessly reassembled during playback by the stimulus presentation program.

Procedure

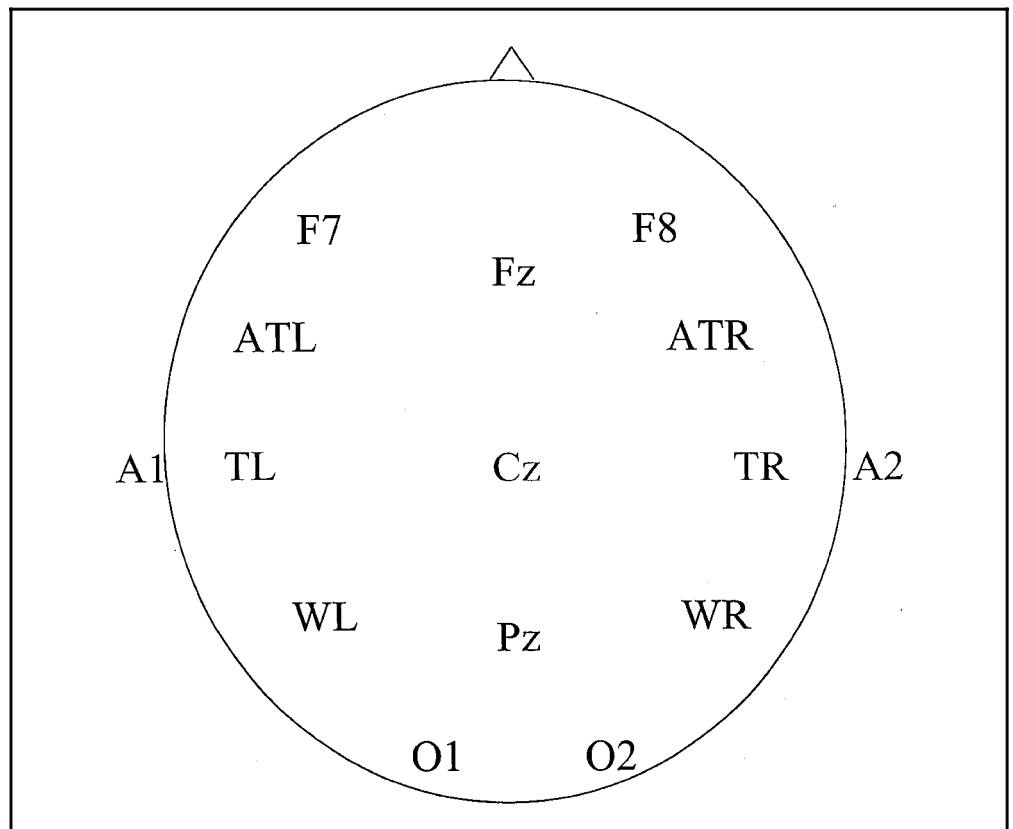
Subjects were seated comfortably in a quiet room approximately 5 feet from a computer screen. Each trial consisted of the following events. A fixation cross appeared in the center of the screen and remained for the duration of the sentence. The sentence was sent to a digital-to-analog converter and binaurally presented to the subject over headphones at a comfortable listening level (approximately 65 dB sound pressure level, or SPL). Subjects were asked not to move their eyes or blink while the fixation cross was present (as this causes electrical artifacts in the electroencephalogram, or EEG). A 1450-msec blank screen interval followed each sentence, after which a prompt appeared on the screen asking the subjects to decide if the previous sentence was an

“acceptable” or “unacceptable” sentence. Acceptable sentences were defined as sensible and grammatically correct; unacceptable sentences were defined as semantically bizarre or grammatically incorrect. Subjects indicated their choice by pressing a button on a small box held in the lap; a decision prompted the start of the next trial. The buttons used to indicate “acceptable” and “unacceptable” (left or right hand) were counterbalanced across subjects. Six example sentences were provided (none of which were used in the experiment), and the subjects were asked if they felt comfortable with the task. The examples were repeated if necessary. The experiment began with a block of 30 sentences. After this, blocks of musical and linguistic stimuli alternated until the experiment was completed (breaks were provided between blocks). Subjects were tested in one session, lasting approximately 2 hr.

ERP Recording

EEG activity was recorded from 13 scalp locations, using tin electrodes attached to an elastic cap (Electrocap International). Electrode placement included the International 10-20 system locations (Jasper, 1958) at homologous positions over the left and right occipital (O1, O2) and frontal (F7, F8) regions and from the frontal (Fz), central (Cz), and parietal (Pz) midline sites (see Figure 8). In addition, several nonstandard sites were used over posited language centers, including Wernicke’s area and

Figure 8. Schematic diagram of electrode montage used in this study.



its right hemisphere homolog (WL, WR: 30% of the interaural distance lateral to a point 13% of the nasion-inion distance posterior to Cz), posterior-temporal (PTL, PTR: 33% of the interaural distance lateral to Cz) and anterior-temporal (ATL, ATR: half the distance between F7 and T3 and between F8 and T4). Vertical eye movements and blinks were monitored by means of an electrode placed beneath the left eye, and horizontal eye movements were monitored by an electrode positioned to the right of the right eye. The above 15 channels were referenced to an electrode placed over the left mastoid bone (A1) and were amplified with a bandpass of 0.01 to 100 Hz (3-dB cutoff) by a Grass Model 12 amplifier system. Activity over the right mastoid bone was actively recorded on a sixteenth channel (A2) to determine if there were lateral asymmetries associated with the left mastoid reference.

Continuous analog-to-digital conversion of the EEG was performed by a Data Translation 2801-A board and AT-compatible computer, at a sampling rate of 200 Hz. ERPs were quantified by a computer as the mean voltage within a latency range time-locked to the onset of words of interest, relative to the 100 msec of activity preceding those words. Trials characterized by excessive eye movement (vertical or horizontal) or amplifier blocking were rejected prior to signal averaging. Less than 10% of the trials were removed due to artifact. In all analyses, ERP averaging was performed without regard to the subject's behavioral response.

EXPERIMENT 2: METHOD

Subjects

The same 15 subjects who served in the language study also served in this experiment (which took place in the same session as the language study). All subjects had significant musical experience (mean: 11 years), had studied music theory, and played a musical instrument (mean: 6.2 hr per week). None of the subjects had perfect pitch.

Stimuli

The stimuli for this experiment consisted of 36 sets of musical phrases.⁸ Each set contained three musical phrases based on a single "root" phrase (such as the phrase shown in Figure 4), which consisted of a sequence of chords within a certain key. The critical difference between phrases in a set was the harmonic identity of the target chord: This could either be the principal chord of the key of the phrase or the principal chord of a nearby or distant key as determined by a music-theoretic device known as the circle of fifths (for details of harmonic structure, see the introduction to Experiment 2).

"Root" phrases ranged in length from seven to twelve chords and had between four and nine chords before the target. Thus, the target chord was always embedded within the phrase. These phrases used harmonic syntax characteristic of Western European tonal music (Piston, 1978). Voice-leading patterns (the movement of individual melodic lines) and rhythms were representative of popular rather than classical styles. The musical phrases averaged about 6 seconds in duration, with chords occurring at a rate of about 1.8/sec. However, across phrase sets there were fluctuations around these averages: to create rhythmic variety, chords could be of different durations, ranging from a sixteenth note (1/16 of a beat) to a tied half-note (over 2 beats). A beat was always 500 msec in duration (tempo fixed at 120 beats per minute). All phrases were in 4/4 time (4 beats per bar), and the target chord was always one full beat.

Within a set, all three phrases had the same rhythmic pattern and key as the "root" phrase, varying only by the use of chord inversions (an inversion of a chord rearranges its pitches on the musical staff, as when c-e-g is replaced by as e-c-g). This created some variety in the sounds of the phrases in a set while keeping their harmonic structure constant. However, two beats before and one beat after the target were always held constant in chord structure, in order to maximize the comparability of the acoustic context immediately surrounding the target. Finally, in order to avoid priming of the in-key target, the principal chord of a key was always avoided before the target position.

In addition to these 108 musical phrases, another 36 phrases without harmonic incongruities were added. This was done so that phrases with out-of-key chords would be equally common as phrases without such chords, to avoid any contribution to ERP effects introduced by a difference in probability between these two types of sequences. The phrases were produced using a computer MIDI system (Recording Session, Turtle Beach Multisound grand piano) and recorded onto a cassette analog tape. Each phrase was digitized at 11,025 Hz (16-bit resolution, 4000-Hz low-pass filter) and segmented for ERP stimulus coding (in the experiment, segments were seamlessly reassembled during playback by the stimulus presentation program). The final stimulus list consisted of 144 musical phrases, which were randomly ordered and divided into four blocks of 36 phrases.

Procedure

The procedure was identical to that described for the language experiment, except that different definitions were given for "acceptable" and "unacceptable" sequences. Acceptable sequences were defined as "sounding normal" and unacceptable sequences were defined as "sounding odd." Subjects were told that they could

use their own criteria within this very broad scheme of classification but that they should be attentive to the harmonic structure of the music. We decided to explicitly mention harmony after pilot work showed that without this instruction, subjects differed in the musical dimension most attended to (e.g., some focused primarily on rhythm). This suggested that the results for music would include variation due to differences in structural focus. Because we expected that in language subjects would be consistent in focusing on certain structural dimensions (due to their task of comprehending the sentences), we opted to instruct subjects to attend to a particular structural dimension in music: harmony. Four blocks of musical stimuli (36 each) were alternated with the five blocks of linguistic stimuli (30 each) in the same experimental session. The order of blocks was fixed across subjects; breaks were given between each block. The experiment lasted approximately 2 hr.

ERP Recording

ERP recording and averaging were performed in the same manner as in Experiment 1.

Acknowledgments

We thank Jane Andersen, Jennifer Burton, Peter Hagoort, Claus Heeschen, Edward O. Wilson, and two anonymous reviewers for their valuable comments and support. The first author was supported by a grant from the Arthur Green Fund of the Department of Organismic and Evolutionary Biology, Harvard University. This research was supported by NIH Grant HD25889 to the last author.

Reprint requests should be sent to Aniruddh D. Patel, The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121, or via e-mail: apatel@nsi.edu.

Notes

1. Although this example focuses on a closed-class word (*to*), the P600 is also elicited by open class words that are difficult to integrate with the preceding structural context (e.g., number-agreement violations, Hagoort et al., 1993).
2. In Western European tonal music, octaves are divided into 12 discrete pitches, separated by logarithmically equal steps, creating a system of 12 *pitch classes* (named a, a#/b-flat, b, c, . . . g#/a-flat). Subsets of this group of pitch classes form the keys of tonal music: Each key has eight pitch classes and differs from other keys in its number of sharp and flat notes. Chords are simultaneous soundings of three or more notes (in certain interval relations) from a given key.
3. In fact, we found baseline differences at the onset of the second verb that indicate that this was the case.
4. All ERP waveforms shown in this study are grand averages made without regard to the subject's behavioral response. Responses were made off-line (1.5 sec after the auditory sequence ended) and were collected to ensure attention to the stimuli. Although our study was not designed for response-contingent analysis, we visually examined waveforms based on response-contingent averaging and found that the intermediate

ERPs to Condition B in language and music appear to be due to a mixture of more positive waveforms (for sequences judged unacceptable) and less positive waveforms (for sequences judged acceptable). The low number of trials/condition after response-contingent reaveraging prevents us from performing meaningful statistical analyses on these patterns, but we believe they merit further study. In a future study we would also like to address the question of individual variation: although preliminary analyses suggest a relation between size of ERP effects in a given subject and the subject's tendency to reject items in Conditions B and C, a larger number of similarly rated stimuli per condition are needed to make a meaningful assessment.

5. These comparisons were conducted without normalization of the data. Had we found significant condition \times electrode site interactions, we would have normalized the data and repeated the analysis to ensure that these differences were not an artifact of nonlinear effects of changing dipole strength (McCarthy & Wood, 1985).

6. Stimuli are available from the first author upon request.

7. Based on pilot studies suggesting that rapid speech can attenuate the P600 effect, we decided to present these sentences at a slightly slower rate than they had originally been spoken (12% slower, or 4.4 syllables per second, increasing the duration of the sentences to about 4 sec). This made the voice of the female speaker lower, but prosodic aspects of the speech seemed to remain perceptually normal. It should be noted that the speaker was instructed to read grammatically complex sentences (Condition B) in a manner such that the intonation would help communicate the meaning. This stands in contrast to an earlier ERP study of connected speech (Osterhout & Holcomb, 1993), in which splicing procedures were used in order to ensure that prosodic cues *did not* differ between syntactically simple and complex sentences.

8. Originally we had planned to use 30 sets of sequences, as in the language experiment. However, to equally represent the 12 musical keys in the phrases, 36 sets were necessary (each key was represented by three sets of phrases).

REFERENCES

- Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition, 14*, 201–212.
- Bernstein, L. (1976). *The unanswered question*. Cambridge, MA: Harvard University Press.
- Besson, M. (1997). Electrophysiological studies of music processing. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 217–250). Hove, UK: Psychology Press.
- Besson, M., & Fäita, 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.
- Besson, M. and Macar, F. (1987). An event-related potential analysis of incongruity in music and other nonlinguistic contexts. *Psychophysiology, 24*, 14–25.
- Bharucha, J., & Krumhansl, C. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition, 13*, 63–102.
- Bharucha, J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 403–410.
- Bharucha, J., & Stoeckig, K. (1987). Priming of chords: Spread-

- ing activation or overlapping frequency spectra? *Perception and Psychophysics*, *41*, 519-524.
- Brown C. M., & Hagoort, P. (1993). The processing nature of the N400; Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*, 34-44.
- Elman, J. L. (1990). Representation and structure in connectionist models. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 345-382). Cambridge, MA: MIT Press.
- Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348-368.
- Fodor, J. A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut, Storrs, CT.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178-210.
- Friederici, A. D., & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second-pass parsing processes. *Journal of Psycholinguistic Research*, *25*, 157-176.
- Friederici, A., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*, 183-192.
- Geisser, S., & Greenhouse, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95-112.
- Gibson, E. (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1-76.
- Gibson, E., Hickok, G., & Schutze, C. (1994). Processing empty categories: A parallel approach. *Journal of Psycholinguistic Research*, *23*, 381-406.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439-483.
- Holcomb, P. J. (1988). Automatic and attentional processing: an event-related brain potential analysis of semantic priming. *Brain and Language*, *35*, 66-85.
- Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*, *24*, 259-276.
- Holcomb, P. J., & Neville, H. J. (1990). Semantic priming in visual and auditory lexical decision: A between modality comparison. *Language and Cognitive Processes*, *5*, 281-312.
- Holcomb, P. J., & Neville, H. J. (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.
- Jasper, H. (1958). The ten twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371-375.
- Keiler, A. (1978). Bernstein's *The unanswered question* and the problem of musical competence. *The Musical Quarterly*, *64*, 195-222.
- King, J., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*, 376-395.
- Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, *5*, 196-214.
- Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.
- Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic anomaly. *Science*, *207*, 203-205.
- Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161-163.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 71-103). Cambridge, MA: MIT Press.
- MacDonald, M. C., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676-703.
- McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203-208.
- Münter, T. F., Heinze, H. J., Matzke, M., Wiering, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, *39*, 66-72.
- Münter, T. F., Matzke, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudowords. *Journal of Cognitive Neuroscience*, *9*, 318-329.
- Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex*, *2*, 244-258.
- Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., & Garret, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*, 151-165.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785-806.
- Osterhout, L., & Holcomb, P. J. (1993). Event-related potential and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, *8*, 413-437.
- Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience*, *8*, 507-526.
- Paller, K., McCarthy, G., & Wood, C. (1992). Event-related potentials elicited by deviant endings to melodies. *Psychophysiology*, *29*, 202-206.
- Peretz, I. (1993). Auditory atonalia for melodies. *Cognitive Neuropsychology*, *10*, 21-56.
- Peretz, I., Kolinsky, R., Tramo, M., Labreque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, *117*, 1283-1301.
- Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*, 456-479.
- Piston, W. (1978). *Harmony*. 4th ed. (Revised and expanded by Mark DeVoto). New York: Norton.

- Pritchett, B. (1988). Garden path phenomena and the grammatical basis of language processing. *Language*, *64*, 539-576.
- Rugg, M., & Coles, M. G. H. (1995). The ERP and cognitive psychology: Conceptual issues. In M. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind* (pp. 27-39). Oxford: Oxford University Press.
- Sloboda, J. (1985). *The musical mind: The cognitive psychology of music*. Oxford: Clarendon Press.
- Swain, J. (1997). *Musical languages*. New York: Norton.
- Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155-180). Hillsdale, NJ: Erlbaum.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, *33*, 285-318.
- Zattore, R. J., Evans A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*, *14*, 1908-1919.