

Prosody-driven Sentence Processing: An Event-related Brain Potential Study

Ann Pannekamp¹, Ulrike Toepel¹, Kai Alter^{1,2}, Anja Hahne¹,
and Angela D. Friederici¹

Abstract

■ Four experiments systematically investigating the brain's response to the perception of sentences containing differing amounts of linguistic information are presented. Spoken language generally provides various levels of information for the interpretation of the incoming speech stream. Here, we focus on the processing of prosodic phrasing, especially on its interplay with phonemic, semantic, and syntactic information. An event-related brain potential (ERP) paradigm was chosen to record the on-line responses to the processing of sentences containing major prosodic boundaries. For the perception of these prosodic boundaries, the so-called closure positive shift (CPS) has been manifested as a reliable and replicable ERP component. It has mainly been shown to correlate to major intonational phrasing in spoken language. However, to

define this component as exclusively relying on the prosodic information in the speech stream, it is necessary to systematically reduce the linguistic content of the stimulus material. This was done by creating quasi-natural sentence material with decreasing semantic, syntactic, and phonemic information (i.e., jabberwocky sentences, in which all content words were replaced by meaningless words; pseudoword sentences, in which all function and all content words are replaced by meaningless words; and delexicalized sentences, hummed intonation contour of a sentence removing all segmental content). The finding that a CPS was identified in all sentence types in correlation to the perception of their major intonational boundaries clearly indicates that this effect is driven purely by prosody. ■

INTRODUCTION

Prosody, as a part of spoken language, is an important factor in human communication. For the last 2 decades, psycholinguistic research has increasingly focused on the role of prosody in the process of natural language understanding in man. So far, many behavioral studies (for an overview, see Cutler, Dahan, & van Donselaar, 1997) have taken into account prosodic aspects for the production and perception of speech. Most often, their aim was to prove the importance of prosody for resolving semantic and syntactic ambiguities that can frequently occur during reading. For those behavioral studies, it is thus necessary to establish processing conditions that imply a violation. Only by comparing behavior to violations with the proposed “normal” perception mechanisms, can conclusions be drawn for the process of language understanding. Psychophysiological measures such as event-related potentials (ERP) are often faced with the same methodological restrictions. However, the discovery of the closure positive shift (CPS; Steinhauer, Alter, & Friederici, 1999) has made it possible to examine at least prosodic processing at sentence level on-line and without the need of structural viola-

tions in the experimental design. This component thus reflects the perception of the intonation contour of violation-free sentences.

The overall intonation contour of a sentence is often metaphorically referred to as its “sentence melody.” It can provide details about whether an utterance is meant to be a question, declarative, or imperative (the sentence mode) and about the syntactic and informational structuring (Ladd, 1996; Selkirk, 1984; Cooper & Paccia-Cooper, 1980). At sentence level, prosody comprises various tonal and durational parameters for its interpretation. One of the most important aspects in intonational languages such as German, Dutch, and English is the fundamental frequency (F0), which roughly corresponds to the pitch contour of the utterance. The pitch contour depends on the utterance’ internal prosodic structuring. Sentences generally consist of one or more major intonational phrases (IPh; Selkirk, 1984) as means of their prosodic realization. These phrases are defined as containing at least one nuclear accent and a boundary tone at their right edges (Pierrehumbert, 1980). Furthermore, the last syllable before the edge is usually lengthened. IPhs can optionally be followed by pauses to separate them from the adjacent IPh. Within one IPh, the excursion of the pitch contour is lowered toward the end (down step phenomenon)

¹Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, ²University of Newcastle

and is then being reset at the beginning of the following phrase.

The prosodic structuring at phrase level is often but not always determined by the syntactic structure. Deriving interpretation problems in situations with ambiguous applications of one level onto the other have been the motivation for many studies (for an overview, see Cutler et al., 1997). As their results indicate, syntactic cues are essential for the parsing of speech. Hence, the role of prosody does only seem to be a supporting but not leading one. The results of Beach (1991), however, suggest that prosody (here, F0 and duration) can be used in sentence interpretation before disambiguating syntactic-semantic cues are being encountered. She presented short versus long sentence fragments prosodically agreeing with either a direct object or clause complement continuation. Independent of fragment length, she found small but significant effects when participants had to judge the congruence of sentence—initial prosody and intended continuation (but see Stirling & Wales, 1996, with the same paradigm but diminished effects). More important for the present study is the role of prosodic breaks for grouping lexical and nonlexical elements into phrases. According to Kennedy, Murray, Jennings, and Reid (1989), as well as Pynte and Prieur (1996), prosodic boundaries are used to organize words into syntactic phrases.

But does this grouping strategy of the human parser also apply in the absence of meaningful units (words) in an utterance? There is at least some evidence for the superfluity of phonemic, semantic, and syntactic information in an utterance to interpret its prosodic form. Using reiterant speech (in the form of syllable strings), de Rooij (1976) found that listeners are able to detect major prosodic boundaries. The same is true when the acoustic input does not contain reliable semantic information (de Rooij, 1975, using spectrally scrambled speech). Even manipulated speech with overall-deprived segmental content can still be analyzed by listeners for major prosodic phrasing (Kreiman, 1982, with low-pass-filtered speech; Collier & 't Hart, 1975, with hummed sentences). These results present (partial) evidence for Beckman's (1996) proposal that the prosodic structure of an utterance has to be seen as a full grammatical property also requiring its own parsing. Still, the question remains open whether deprived segmental information in the speech input leads to processing difficulties and/or delays. All of the studies cited above are not sufficient to explore the time course of the integration of segmental (phonemic, semantic, and syntactic) and suprasegmental properties during spoken language understanding due to the off-line paradigms used by them.

A pioneering neurophysiological study investigating the influence of prosody for the processing of spoken language was conducted by Steinhauer et al. (1999)

using ERP measures. They constructed sentence material with different accentuation patterns depending on the syntactic structure. Two sentence types were created, one condition comprising only one IPh boundary and, in contrast, a second condition with two IPh boundaries. These sentences were then presented auditorily. ERPs showed a characteristic positive-going waveform while listeners perceived the major prosodic phrase boundaries. Because of the off-line-proven differences in intonational phrasing, ERPs exhibited significantly different patterns for the two conditions. Whereas the condition with only one IPh boundary evoked one corresponding positive shift, the condition with two IPh boundaries induced two positive shifts. Acoustic analyses and post hoc mapping procedures of the acoustic data and the electrophysiological responses confirmed the direct linking of this novel component to the closure of prosodic phrases. It was termed the *closure positive shift*. In a subsequent experiment, they carefully removed the pause after the first IPh boundary in the condition conveying two boundaries to rule out the interpretation possibility that the CPS is present as a marker for the interrupted speech stream. Results are basically the same as in the first experiment. This finding supports the idea that the CPS reflects the processing of the prosodic boundary itself rather than the perception of an (optional) pause loosely related to it. Nonetheless, these data cannot deny a possible confound between prosodic and syntactic units in the speech data used. At that point, the conclusion that the CPS is predominantly related to prosodic structuring per se could still not be convincingly drawn.

In an additional attempt to demonstrate the prosodic nature of the CPS, Steinhauer and Friederici (2001) explored the perception of prosodic phrases in delexicalized speech. For this purpose, a particular filtering procedure, PURR (Sonntag & Portele, 1998), was applied to the speech material described above to delete all segmental information (phonemic, semantic, and syntactic) while preserving the prosodic pattern (pitch, amplitude, and rhythm). These artificially delexicalized sequences were auditorily presented. This design allowed to test whether the processes underlying the CPS are syntactic rather than prosodic in nature and whether prosodic aspects are sufficient to elicit the component. Because syntactic processing heavily relies on morpho-lexical information, the occurrence of a CPS during the perception of delexicalized stimuli can only be attributed to the prosodic nature of that component. As assumed, the auditorily presented sentences evoked a CPS at the first phrase boundary. This deflection, however, diverged in shape and amplitude from the CPS in normal sentences and did not appear for the second proposed IPh boundary. Additionally, both sentence conditions exhibited a broadly distributed negative shift over the whole sentence that was not present in normal sentences. Post hoc analyses led the authors

to specify this deflection as an expectancy-correlated contingent negative variation (CNV; Tecce & Cattanach, 1987) due to task requirements.

Using functional magnetic resonance imaging, a study by Meyer, Alter, Friederici, Lohmann, and von Cramon (2002) then aimed to localize the underlying neural substrates of prosodic processing at sentence level. Their stimuli comprised quite similar conditions as our current studies (see Methods). Comparisons were computed for the brain activation patterns while listeners perceived speech conveying all linguistic levels (phonemic, syntactic, semantic, and prosodic) versus syntactic speech (replacement of all content words with phonotactically legal pseudowords) versus artificially delexicalized speech as introduced by Steinhauer and Friederici (2001). Meyer et al. (2002) showed the differing involvement of the frontal and temporal cortex areas in the processing conditions as a function of the presence versus absence of linguistic information. The perception of semantic and syntactic information predominantly involves anterior and posterior portions of the left hemisphere's superior temporal region. In contrast, processing pure prosodic parameters in delexicalized speech appears to strongly involve contralateral superior temporal cortical regions. Additionally, both deviant speech conditions (pseudoword and delexicalized sentences) lead to significant bilateral activations in fronto-opercular sites. Most important, the brain's activation to speech melody was overall stronger in the right than in the left hemisphere.

These results regarding the lateralization of intonational processing in the brain are consistent with several other neuroimaging studies conducted with healthy and brain-damaged participants. Zatorre, Evans, Meyer, and Gjedde (1992) investigating the processing of pitch changes by means of positron emission tomography showed an activation of the right prefrontal cortex while changes in pitch were perceived. In accordance with these data, Tzourio et al. (1997) found a reliable rightward asymmetry for the supratemporal region for passive listening to tones too. To sum up here, there is strong evidence that pitch processing in the absence of additional linguistic information such as syntax and/or semantics takes place in right hemisphere regions (i.e., Zatorre & Belin, 2001; Johnsrude, Penhune, & Zatorre, 2000; Baum & Pell, 1999).

The current series of experiments focused on several issues arising from former findings. First, Steinhauer and Friederici (2001) only partially replicated the CPS pattern gained by the perception of IPhs in normal speech when exploring delexicalized material. The question arose whether this could be because of the missing semantic-syntactic support for the interpretation of IPh boundaries as such (see Selkirk, 1984, for the "sense unit condition" of major prosodic phrases) or attributable to the artificiality of the stimulus material. Therefore, we exclusively created natural manipulations in

our material resulting in four experimental variations (normal sentences, jabberwocky sentences, pseudoword sentences, and hummed speech). Furthermore, the sentence material in each experimental variation comprises two conditions. Whereas one condition (Condition A) only contains one major intonational boundary, the other condition (Condition B) has two boundaries. By using the differing prosodic structures to elicit the CPS, it should be possible to directly relate its occurrence in time with the underlying prosodic boundaries responsible for it.

Second, we reasoned that the distribution of the CPS might show a shift to the right hemisphere with decreasing segmental information in the speech input. This expectation was based on results from the recent functional imaging studies on pitch perception.

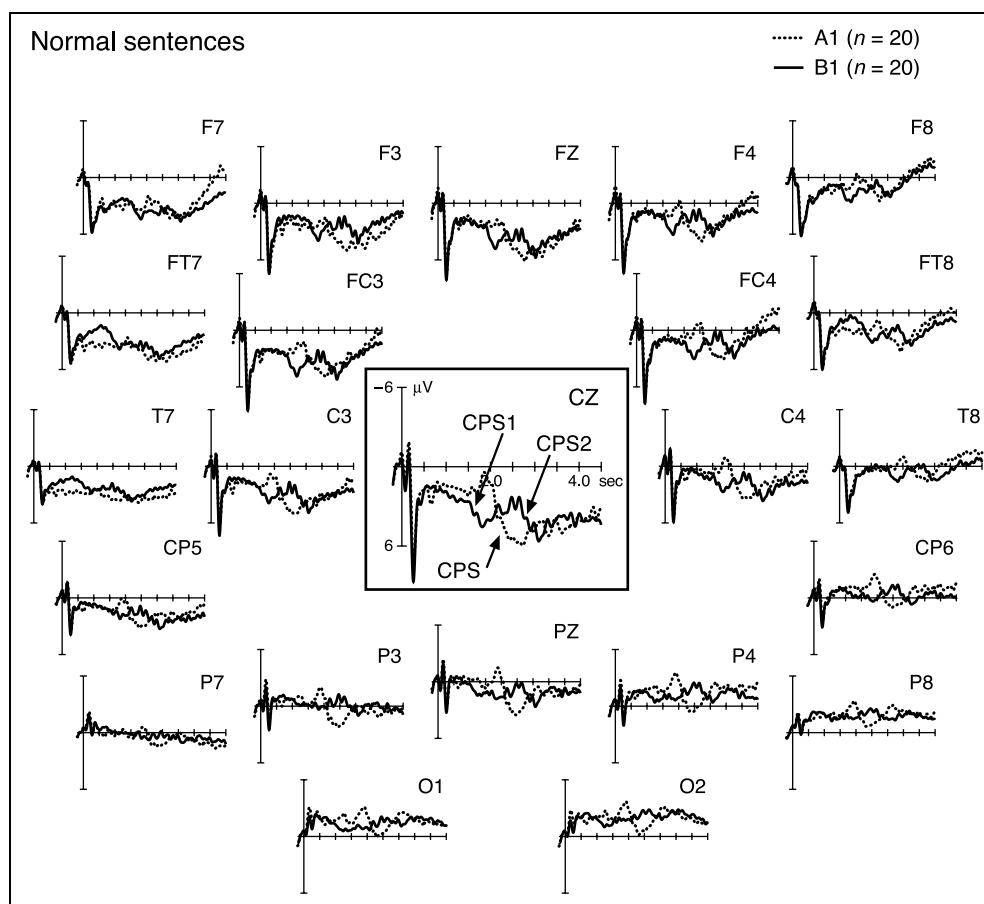
RESULTS

In the following, we present evidence for positive ERP deflections to IPh boundaries. The positivity in correlation to the first IPh boundary appears to depend on the segmental content of the stimuli. This very component moves anteriorly and partly rightward as a function of reduced segmental information. The second positivity, however, is found to be relatively independent of the amount of segmental information provided in the speech input.

Experiment 1: Normal Sentences

In Figure 1, ERPs to the normal sentence material separated into two conditions (A1 and B1) are illustrated. It can be seen that Condition B1 (the condition with two IPh boundaries) elicits a first positive shift starting at about 1500 msec after sentence onset. Statistical analysis on the ERP data was performed by analyses of variance (ANOVAs) with the factors condition, region, and hemisphere (for detailed description, see Methods). The statistical analysis reveals a main effect condition in the time window (TW) 1500–2000 msec at the midline, $F(1,19) = 7.84$, $p < .01$, as well as in the lateral regions of interest (ROIs), $F(1,19) = 6.34$, $p < .05$. A second positive-going waveform for this condition starts at about 2700 msec. Hence, the condition with two IPh boundaries evokes two positive deflections in the ERPs. In contrast, Condition A1 (comprising just one IPh boundary) induces one positive shift starting at about 2000 msec after sentence onset. A main effect condition was ascertained in the TW between 2300 and 2800 msec at the midline, $F(1,19) = 7.09$, $p < .05$, and at the lateral ROIs, $F(1,19) = 5.83$, $p < .05$. All effects are distributed bilaterally across hemispheres. No interactions are provable between the factors Condition \times Hemisphere and Condition \times Region. In summary, both conditions display signifi-

Figure 1. Grand average ERPs for normal sentences in Conditions A1 (dotted line) and B1 (solid line). B1 displays a first positive shift starting at about 1500 msec and a second one at about 2700 msec. A1 shows only a positive deflection starting around 2000 msec.



cantly different ERP patterns in correlation to the different IPh boundaries.

Experiment 2: Jabberwocky Sentences

Figure 2 shows the processing of the jabberwocky sentences in both conditions (A2 and B2). Condition B2 (with two IPh boundaries) evokes a first positive shift starting at about 1500 msec after sentence onset. This shift in Condition B2 is significantly different to Condition A2 in a TW between 1800 and 2300 msec at the midline electrodes, $F(1,21) = 6.26$, $p < .05$, and at the lateral ROIs, $F(1,21) = 5.31$, $p < .05$. In addition, an interaction between the factors Condition \times Region can be seen at the midline electrodes, $F(2,42) = 6.58$, $p < .01$, and at the lateral ROIs, $F(2,42) = 20.67$, $p < .01$. To clear up the interaction, separate analyses for each region were computed. There is a main effect condition at the frontal, $F(1,21) = 10.99$, $p < .01$, and central electrodes, $F(1,21) = 7.86$, $p < .01$. A second positive shift in condition B2 starts at approximately 2800 msec. Thus, Condition B2 elicits two positive-going waveforms. In contrast, Condition A2 (with only one IPh boundary) evokes only one positive shift. The shift starts at approximately 2200 msec after sentence onset. The statistical analysis reveals a

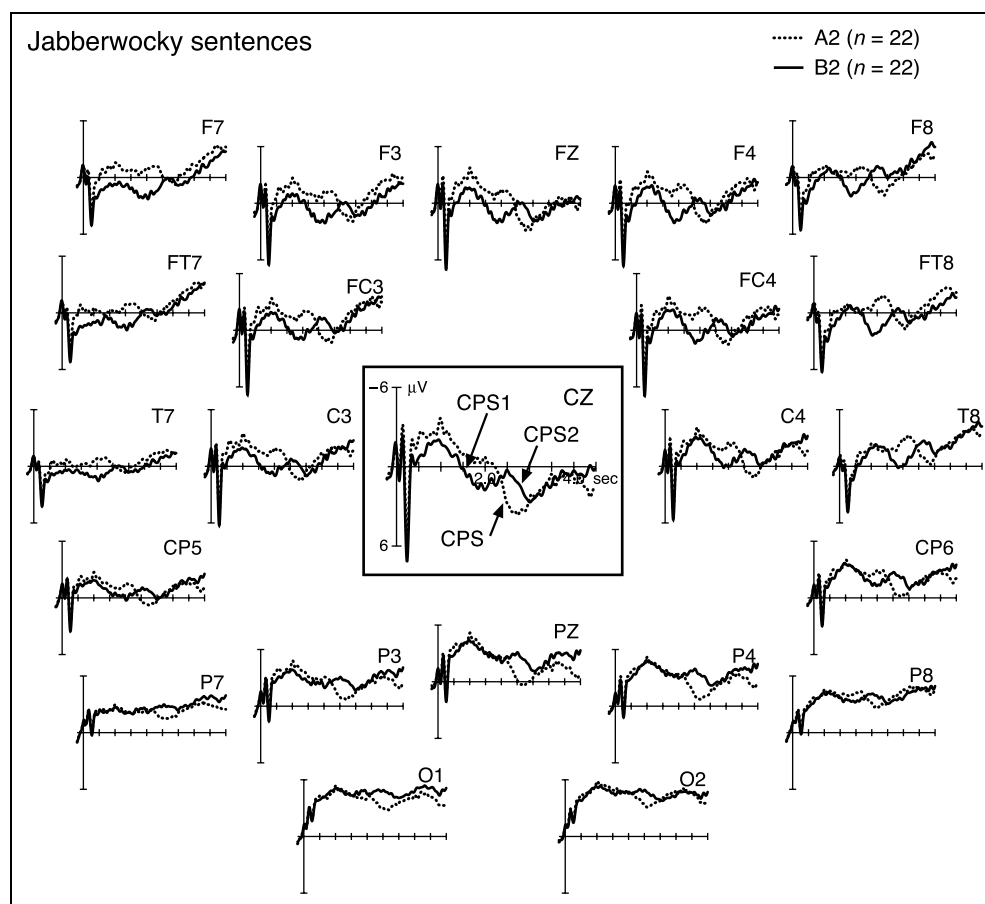
main effect condition in the TW 2500–3000 msec at midline electrodes, $F(1,21) = 6.31$, $p < .05$. Therefore, Condition A2 displays one positive shift whereas Condition B2 shows two.

Experiment 3: Pseudosentences

The ERPs for the pseudosentences in both conditions (A3 and B3) are illustrated in Figure 3. Condition B3 with two IPh boundaries again elicits two positive shifts, whereas Condition A3 (comprising only one boundary) shows only one positive shift. The first positive shift in Condition B3 starts at about 1500 msec and is confirmed statistically by two significant interactions at lateral electrode sites. A Condition \times Hemisphere interaction, $F(1,21) = 7.69$, $p < .01$, and a Condition \times Region interaction, $F(2,42) = 7.12$, $p < .01$, are shown in the TW from 1500 to 2000 msec. Further analyses reveal a main effect condition at right frontal electrodes, $F(1,21) = 5.94$, $p < .05$.

For the second positive-going waveform in the TW between 2500 and 3000 msec, there is a main effect condition at lateral, $F(1,21) = 7.21$, $p < .05$, as well as at midline electrode sites, $F(1,21) = 5.68$, $p < .05$. Thus, Condition A3 displays one positive shift in correlation to the underlying major prosodic boundary whereas

Figure 2. Grand average ERPs for jabberwocky sentences in Conditions A2 (dotted line) and B2 (solid line). B2 displays a first positive shift starting at about 1500 msec and a second one at about 2800 msec. A2 shows only a positive deflection starting around 2200 msec.



Condition B3 with two IPh boundaries shows two positive shifts.

Experiment 4: Hummed Sentences

In Figure 4, ERPs to the hummed material in the two conditions (A4 and B4) are illustrated. Descriptively, in Condition B4 with two IPh boundaries, two positive shifts are accordingly induced. In contrast, Condition A4 with only one IPh boundary only reveals one of these shifts. Furthermore, an additional negative peak is evoked by Condition B4 preceding the positive deflections. This peak reaches significance in the TW from 500 to 1000 msec, respectively, and at lateral, $F(1,21) = 7.88$, $p < .01$ and midline electrode sites, $F(1,21) = 5.48$, $p < .05$. Following this negativity, the adjacent positive shift reveals a significant Condition \times Region interaction, $F(2,42) = 14.75$, $p < .01$, in the TW from 1000 to 1500 msec. The second positive shift in the TW from 2000 to 2500 msec is confirmed by a main effect condition, $F(1,21) = 7.52$, $p < .01$, at the midline electrodes and a significant Condition \times Hemisphere interaction, $F(1,21) = 7.78$, $p < .01$. Further analyses can locate a main effect condition at right hemisphere electrodes, $F(1,21) = 6.04$, $p < .05$.

Post hoc hypotheses about the influence of the early negativity on the localization of the first positive shift

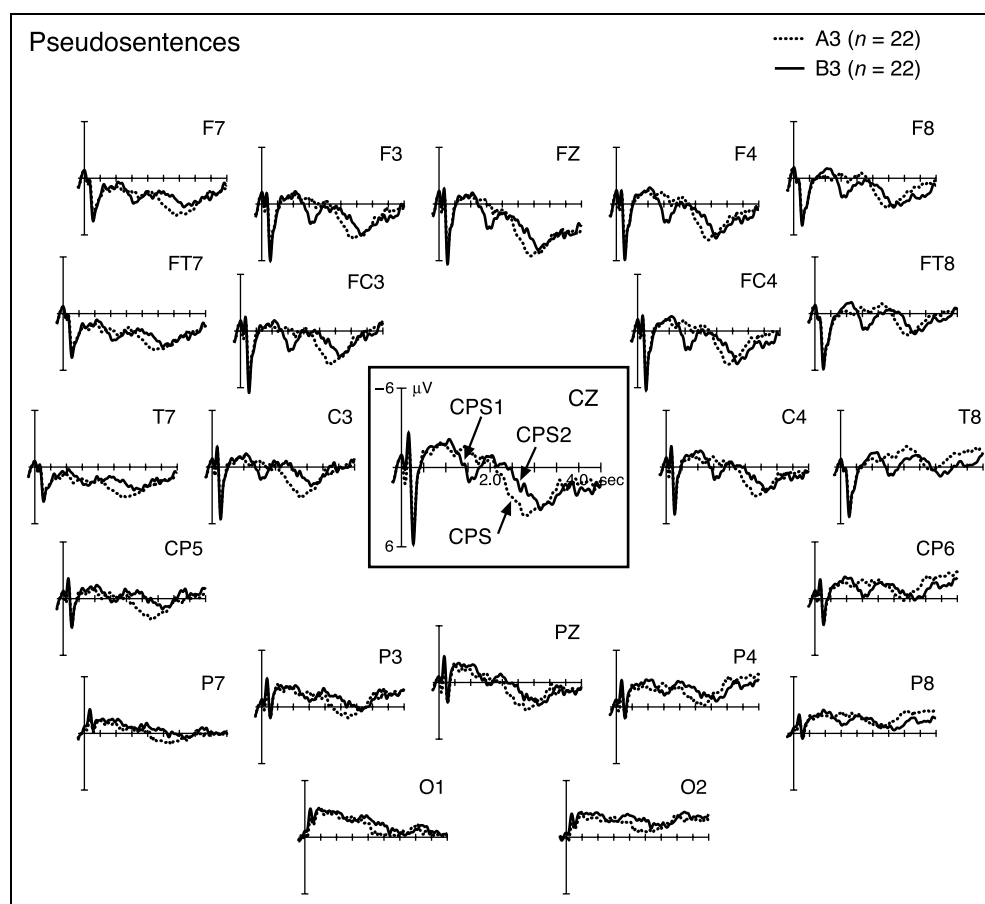
led us to conduct a new analysis of the relevant TW. For this reason, the onset was shifted to the individual end of the first sentence fragment of each item (see Figures 5 and 9). This is where we expected the actual occurrence of the component. The chosen TW for the statistical analysis was again 500 msec to reveal comparable results. Descriptively, a strong positive shift is revealed in this TW (see Figure 5). A main effect condition is demonstrated in midline, $F(1,21) = 20.39$, $p < .01$, as well as in lateral ROIs, $F(1,21) = 22.88$, $p < .01$. Additionally, a significant interaction Condition \times Region was proven for midline, $F(2,42) = 4.17$, $p < .05$, and lateral sites, $F(2,42) = 19.59$, $p < .01$. Further analyses reveal a main effect condition at frontal, $F(1,21) = 38.48$, $p < .01$, and central electrodes, $F(1,21) = 25.92$, $p < .01$.

In summary, also the differing prosodic structuring in the hummed speech material evokes different ERP patterns; whereas Condition A4 with one major intonational boundary evokes one positive shift, Condition B4 shows two positive shifts in correlation to the two IPh boundaries.

DISCUSSION

The intention of the described experiments was to show that prosodic information is processed relatively

Figure 3. Grand average ERPs for pseudosentences in Conditions A3 (dotted line) and B3 (solid line). B3 displays a first positive shift starting at about 1500 msec and a second one at about 2500 msec. A3 shows only a positive deflection starting around 2000 msec.



independent of segmental information during the perception of spoken language and to show the purely prosodic nature of the CPS. The positive deflections in the ERPs of all experiments strongly resemble this component as established by Steinhauer et al. (1999).

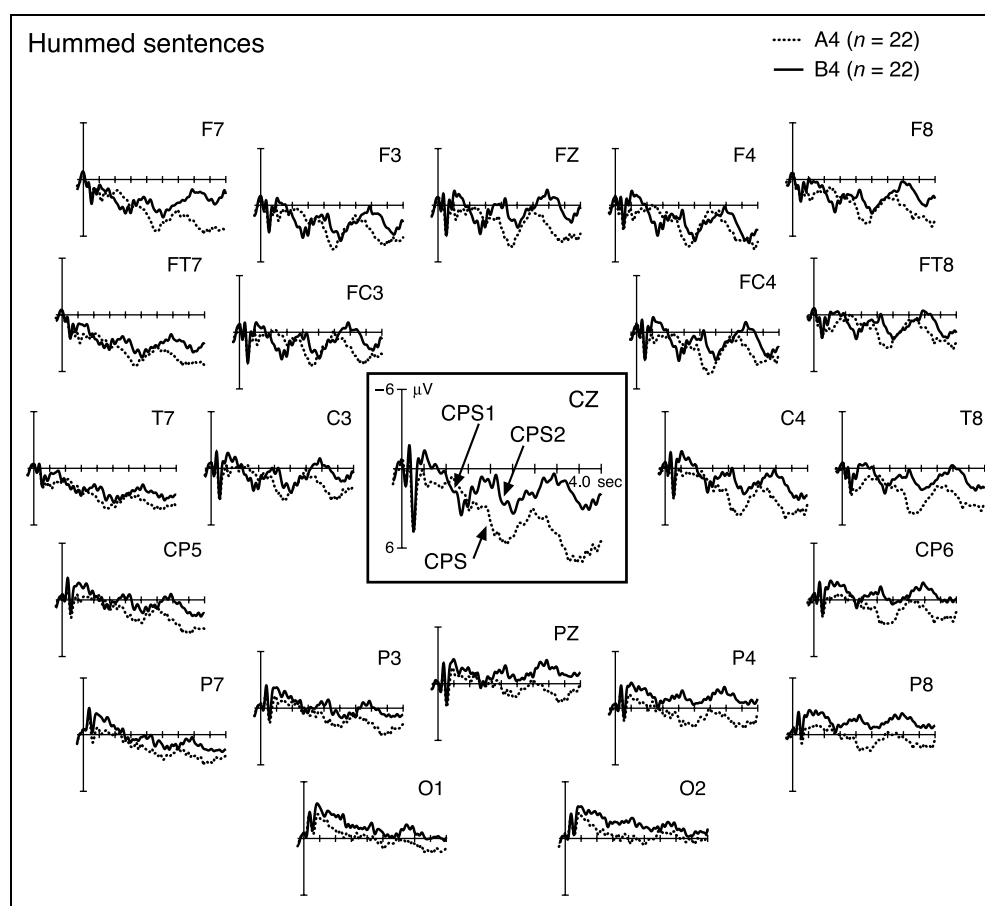
By systematically varying the phonemic, semantic, and syntactic substance, cues for the interpretation the speech input differ between every experiment. The present ERP data and the acoustic analyses of the speech material indicate that the proposed CPS reflects the processing of major intonational boundaries because it is observable in all experimental variations (normal, jabberwocky, pseudo, and hummed sentences). For Conditions A1–A4 (containing one IPh boundary in all sentence types), a CPS is observed in the position of this boundary, whereas for Conditions B1–B4 (containing two IPh boundaries in all sentence types), all experimental variations reveal two CPS, one at each boundary. This result is analogous to the data of Steinhauer et al. (1999).

However, the distribution of the CPS over the scalp varies as a function of the experimental variation and of sentence condition (for an overview, see Table 1). In the first TW, the CPS in the normal sentences (Experiment 1) shows a broad distribution over the whole scalp. In addition, sentences lacking semantic content

(jabberwocky sentences, Experiment 2) replicated a main effect condition at midline and lateral sites and an additional move of the CPS to anterior sites. In contrast, sentences lacking semantic and syntactic content (pseudosentences, Experiment 3) reveal a distribution of the CPS to right-hemispheric anterior electrodes. For sentences lacking any segmental content (hummed sentences, Experiment 4), the CPS is again broadly distributed across midline and lateral electrodes and shows a stronger effect at anterior and central sites.

For this first TW, an anterior move of the CPS could thus be established as a function of decreasing linguistic substance of the speech input. Surprisingly, a general rightward trend could only be found for the pseudosentences, but not for the hummed material. A possible explanation for the missing lateralization of the CPS in the first TW in the hummed sentences could be the exclusive interpretation of the acoustic signal by means of timing (pause and segment duration) and pitch information. Some theories about hemispheric specialization for processing time and pitch information state that the left hemisphere is more involved with the former (Dykstra, Gandour, & Stark, 1995) and the right hemisphere is more involved with the latter (Baum & Pell, 1997; Van Lancker & Sidtis, 1992). We suggest that the exclusive reliance on these properties

Figure 4. Grand average ERPs for hummed sentences in Conditions A4 (dotted line) and B4 (solid line). B4 first displays a negative shift in the TW between 500 and 1000 msec. It is followed by a positive shift starting at about 1000 msec and a second one at about 2000 msec. A4 shows only a positive deflection starting around 2500 msec.



leads to a stronger integrative processing of timing and pitch information after the beginning of the hummed utterances than in the experimental variations with varying segmental content. In the further time course of perception, the expected right-hemispheric lateralization is nevertheless being observable. The second relevant TW then also reveals a significant right hemispheric effect.

In addition, the negative deflections in the hummed material for Condition B4 (TW 500–1000 msec) propose an early difference in the processing mechanisms for this purely prosody-conveying input. According to acoustic data, Condition B4, as opposed to Condition A4, does not convey an early high accent in the beginning of the first IPh. This accentual pattern, however, does support listeners' predictions about the subsequent prosodic structure of the utterance (see also Warren, Grabe, & Nolan, 1995; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Beach, 1991). We therefore interpret the early broadly distributed negativity as a pure detection mechanism for a nonexistent early accent, which determines listeners' implicit hypotheses about the length of the prosodic phrase being currently processed.

For the CPS in the second relevant TW (i.e., the TW in which the CPS is present for both sentence condi-

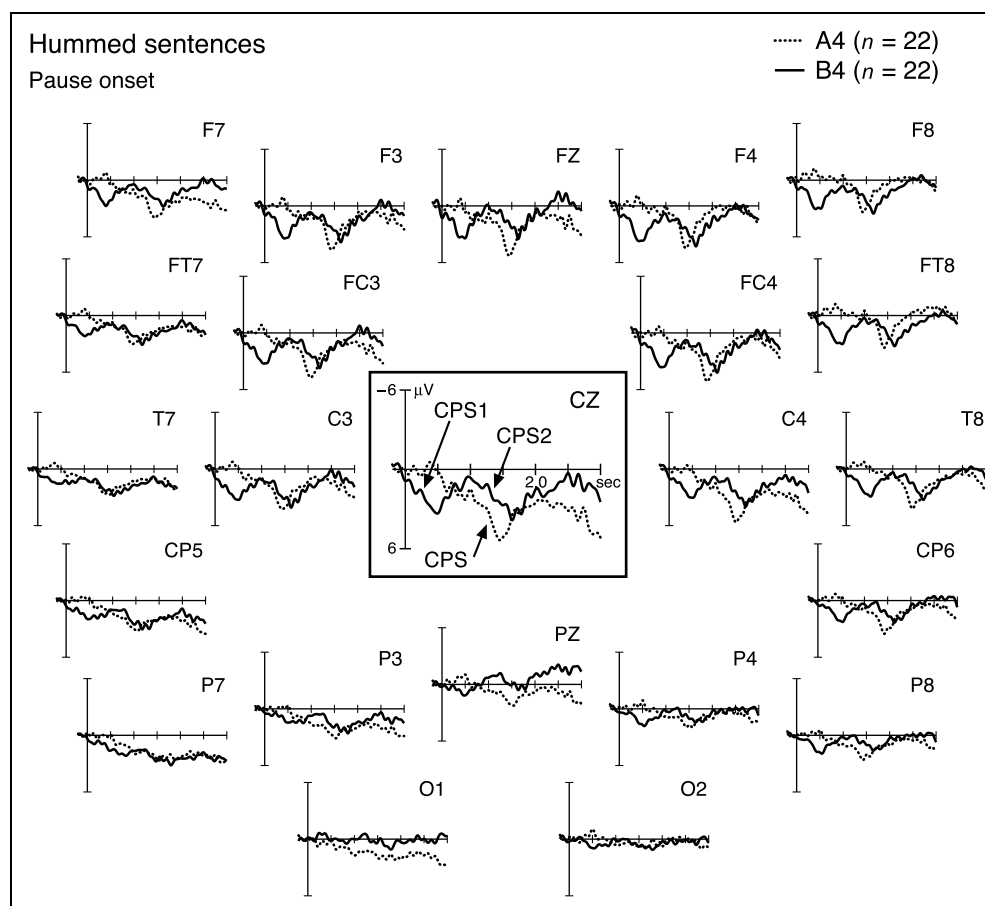
tions in all experiments), its distribution is always broad over the midline electrodes. For normal and pseudosentences, it also extends to the lateral sites. In this TW, a lateralization to the right hemisphere can only be proven for hummed speech. This latter effect stands in apparent contrast to the data of Steinhauer and Friederici (2001), where the (second) CPS could not be observed at all with artificially delexicalized material. It thus seems that the naturalness of acoustic material comprising exclusively suprasegmental information (such as the hummed material) is a better source for investigating on-line the processing of major prosodic boundaries.

Overall, it could be shown that the on-line prosodic sequencing is indicated by the CPS, independent of the segmental content of the incoming speech or speech-like stream.

Conclusions

The CPS, as an electrophysiological correlate to one aspect of prosodic processing, namely, the perception of major intonational boundaries (IPh), could be established as being independent from the four experimental variations described above. The absence of semantic,

Figure 5. Grand average ERPs for the offset of the first sentence fragment in hummed sentences for Conditions A4 (dotted line) and B4 (solid line). B4 reveals a first positive shift in the TW 0–500 msec and a second one at about 1000 msec. A1 shows only a positive deflection starting around 500 msec.



syntactic, and phonemic information in the speech stream leads to a persisting elicitation of the CPS in IPh positions as proven by the acoustic data.

Thus, these findings indicate that the component is exclusively relying on pure prosodic information. The observed differences in the scalp distribution of the CPS as a function of the segmental content of the acoustic speech stream suggest that prosodic processing interacts with other information types involving different brain systems. Surprisingly, only the CPS in earlier TWs shows an anterior and rightward movement that can only be attributed to the decreasing phonemic, lexical, and syntactic information in the speech signal.

The CPS in the later TWs, however, displays a quite steady scalp distribution across all experimental variations.

METHODS

Subjects

All participants in the experiments were strongly right-handed as measured by the Edinburgh Handedness Inventory (Oldfield, 1971). All subjects were native speakers of German with no reported hearing or neurological disorders. They were students of the Univer-

sity of Leipzig; four of them came from the linguistic department. Five subjects play an instrument. None reported any professional music activities. Most of them had prior experience in ERP recordings. In the first experiment (normal sentences), 20 students (mean age 23.4 years; 10 women) took part. Twenty-two participants (mean age 24.6 years; 11 women) were tested in the second experiment (jabberwocky sentences). The same number of participants took part in the third experiment with pseudosentences (mean age 23.6 years; 11 women) and in the fourth with hummed sentences (mean age 24.5 years; 11 women).

Acoustic Stimuli

Four different sentences types were created, namely, normal (Experiment 1), jabberwocky (no lexical content with preserved syntactic information, Experiment 2), pseudo (no lexical or syntactic information, Experiment 3), and hummed (no lexical, syntactic nor phonemic information, Experiment 4) sentences. Every sentence type additionally comprised two conditions. Condition A contains only one intonational boundary, whereas there are two intonational boundaries in Condition B. Every type of sentences was manifested in 96 utterances. Half

Table 1. Overview of the Significant Effects for the Four Experimental Variations in the Two Relevant TWs

| | <i>Normal Sentences</i> | <i>Jabberwocky Sentences</i> | <i>Pseudosentences</i> | <i>Hummed Sentences</i> |
|------------------------------------------------------------------------------|-----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| First TW (proposed position of the first IPh boundary in Cond B1–B4) | Main effect Cond at midline and lateral sites | Main effect Cond at midline and lateral sites Interaction Cond × Reg Main effect Cond at anterior and central electrodes | Interaction Cond × Hem and Cond × Reg Main effect Cond at right anterior electrode | Main effect Cond at midline and lateral sites Interaction Cond × Reg at midline and lateral sites Main effect Cond at anterior and central electrodes |
| Second TW (proposed position of the IPh boundaries common to all conditions) | Main effect Cond at midline and lateral sites | Main effect Cond at midline sites | Main effect Cond at midline and lateral sites | Main effect Cond at midline sites Interaction Cond × Hem Main effect Cond at right Hem |

Cond = Condition; Hem = Hemisphere; Reg = Region.

of the utterances per type thereby belong to Condition A, whereas the other 48 sentences per type are of Condition B.

All sentences were produced by the same trained female speaker in a soundproof chamber and then digitized (44.1 kHz/16-bit sampling rate/mono). All stimuli were normalized in amplitude to 70%.

For the analysis of the tonal and durational properties of the speech material, utterances were divided into three parts of interest. The first part of all experimental variations was defined analogous to the subject + verb fragment (“Kevin promises”) in the normal sentences. The second part included the successive object + verb fragment (“mom to kiss”) and the third the whole conjunctive phrase (“and to be a good boy for a while”). For each defined part, four F0 values were computed: the first, the maximal, the minimal, and the last per part per condition.

Experiment 1: Normal Sentences

In Experiment 1, sentence material was used, which comprises all levels of linguistic information (i.e., phonemic, syntactic, semantic, and prosodic) contained in natural speech. Because of the valence of the second verb (intransitive vs. transitive), conditions differed in syntactic structure (see Table 2).

The analysis of the tonal (Figure 6A) and durational (Figure 6B) properties shows clear differences between

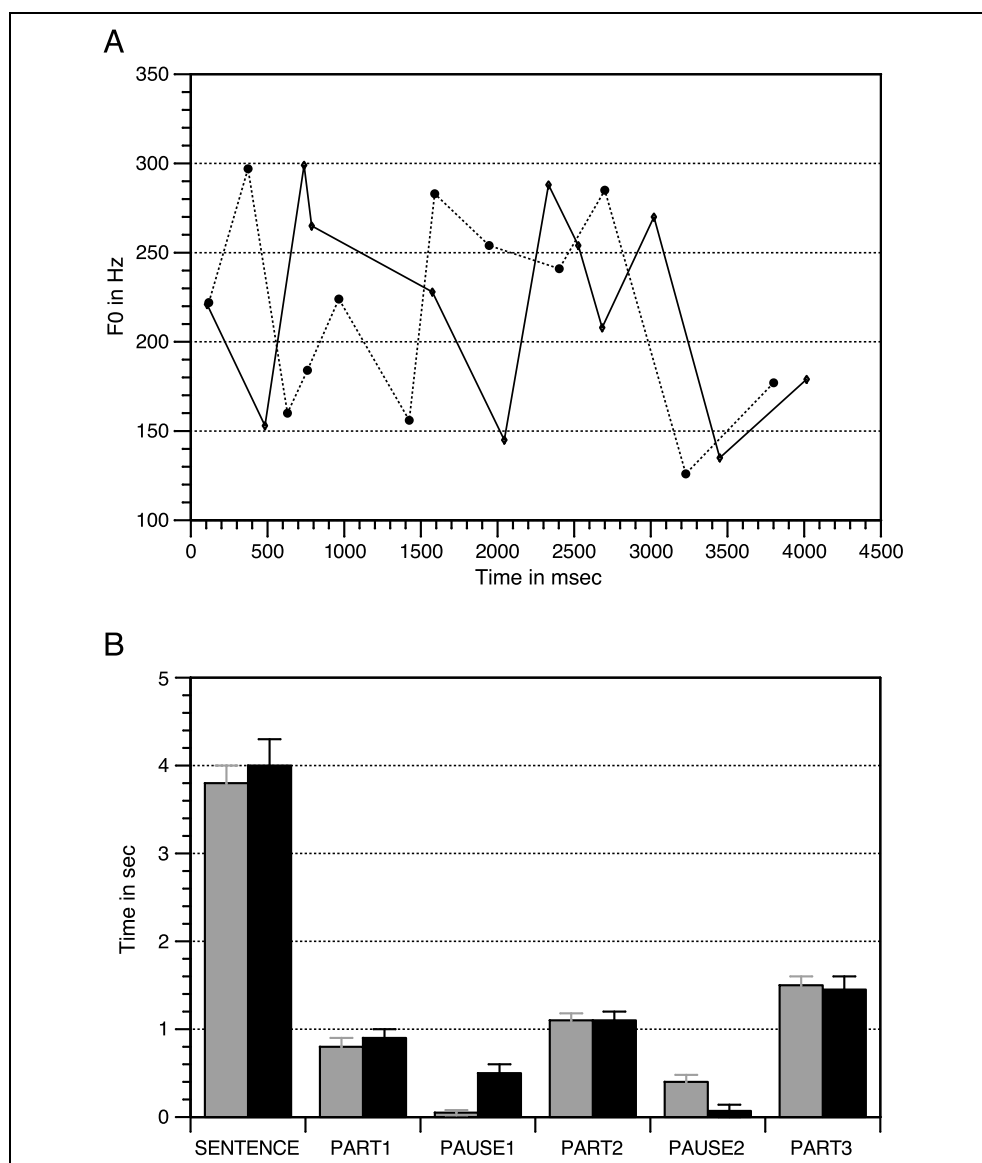
condition A1 and B1. In Condition B1, the fundamental frequency pattern exhibits a high tone at the end of the first sentence part (after the verb “promises”). This is not the case in Condition A1. A second clear difference between conditions is the duration of this first part. It is significantly longer for Condition B1 as proven by the two-tailed *t* test, $t(94) = -6.00$, $p < .01$. Furthermore, Condition B1 exhibits a significantly longer pause after the first part in comparison to Condition A1, $t(94) = -21.86$, $p < .01$. Taking these factors together, it can be concluded that Condition B1 (transitive condition) comprises an IPh boundary (IPh, Selkirk, 1984) after the first part whereas Condition A1 (intransitive condition) does not.

Additionally, after the position of the second verb, both conditions show an IPh boundary, mainly indicated

Table 2. Examples of the Material in Experiment 1 (With Literal Translations)

| | |
|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Condition A1 | [Kevin verspricht Mama zu schlafen] _{IPh1} [und ganz lange lieb zu sein.] [Kevin promises mom to sleep] _{IPh1} [and to be a good boy for a while.] |
| Condition B1 | [Kevin verspricht] _{IPh1} , [Mama zu küssen] _{IPh2} [und ganz lange lieb zu sein.] [Kevin promises] _{IPh1} , [mom to kiss] _{IPh2} [and to be a good boy for a while.] |

Figure 6. (A) Normal sentences: Stylized course of F0 (in Hz) for Conditions A1 (dotted) and B1 (solid). (B) Normal sentences: Sentence, part, and pause durations for Conditions A1 (gray) and B1 (black).



by a high boundary tone. In summary, Condition A1 displays one IPh boundary (approximately 1950 msec after sentence onset) within the sentences, whereas Condition B1 exhibits two (at 950 and 2700 msec postonset).

Experiment 2: Jabberwocky Sentences

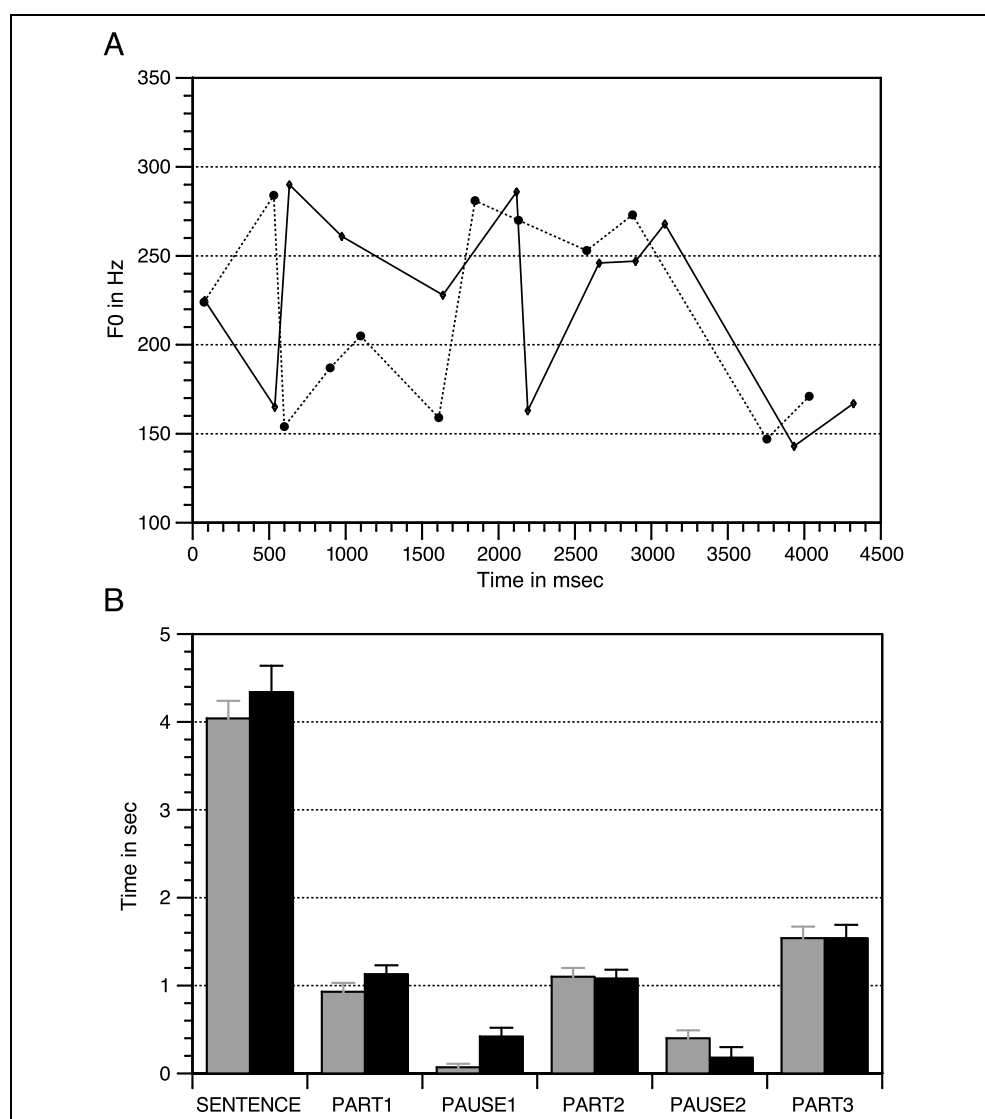
For Experiment 2, jabberwocky sentences were created analogous to the structure of the normal sentences of Experiment 1. They contained all function elements necessary to recover a syntactic structure but no content words. All content words were replaced with phonotactically legal pseudowords. Hence, the sentences comprised phonemic, syntactic, and prosodic, but no semantic information. Again, the materials differed in syntactic structure. The sentence structure was then comparable to Experiment 1 and remains grammatically correct (Table 3).

The fundamental frequency pattern (Figure 7A) displays a high tone for Condition B2 at the end of the first part, which is not measurable in Condition A2. Durational analyses (Figure 7B) reveal the existence

Table 3. Examples of the Material in Experiment 2 (With Literal Translations)

| | |
|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Condition A2 | [Der Bater verklicht Onna zu labeiken] _{IPh1} [und das Rado zu nupen.] [The bater rabels Onna to lubol] _{IPh1} [and the rado to nupe.] |
| Condition B2 | [Der Bater verklicht] _{IPh1} , [Onna zu labeiken] _{IPh2} [und das Rado zu nupen.] [The bater rabels] _{IPh1} , [Onna to lubol] _{IPh2} [and the rado to nupe.] |

Figure 7. (A) Jabberwocky sentences: Stylized course of F0 (in Hz) for Conditions A2 (dotted) and B2 (solid). (B) Jabberwocky sentences: Sentence, part, and pause durations for Conditions A2 (gray) and B2 (black).



of a significantly longer first sentence part for Condition B2, $t(94) = -8.59$, $p < .01$. The following pause is also significantly longer in Condition B2, $t(94) = -20.32$, $p < .01$. Analogous to the material in Experiment 1, both conditions comprise an IPh boundary after the second part. Thus, it can be followed that there is only one IPh boundary within Condition A2 (approximately 2100 msec), whereas there are two boundaries in Condition B2 (at 1100 and 2600 msec).

Experiment 3: Pseudosentences

In Experiment 3, linguistic levels for the interpretation of spoken language were further manipulated. Pseudosentences were generated by replacing all function and content words with meaningless phonotactically legal words so that the sentences neither contained semantic

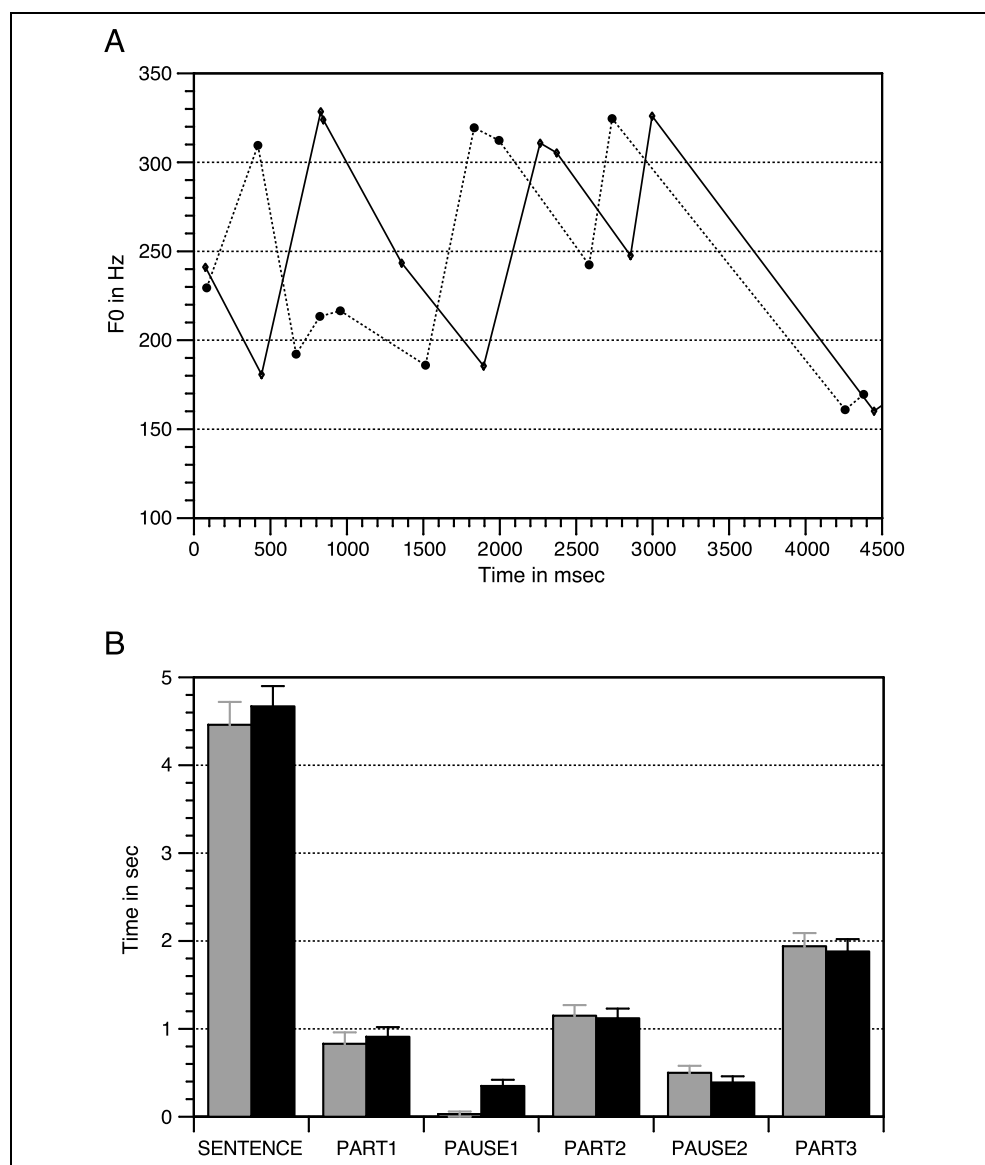
nor syntactic information. Phonemic and prosodic properties remained intact (Table 4).

The inspection of the fundamental frequency (Figure 8A) shows a high tone at the end of the first part of Condition B3. There is no such tone in Condition A3. Durational properties (Figure 8B) reveal a significant difference, $t(94) = -3.41$, $p < .01$, between conditions for the first part. The adjoining pause is also significantly prolonged for Condition B3 as opposed to Condition A3, $t(94) = -27.61$, $p < .01$. This leads to the assumption of an IPh boundary after the first part in

Table 4. Examples of the Material in Experiment 3

| | |
|--------------|----------------------------------------------------------------------------------------------------|
| Condition A3 | [Bater saklimm Onna ko labei keg] _{IPh1} [nug som Rado lie nupes.] |
| Condition B3 | [Bater saklimm] _{IPh1} , [Onna ko labei keg] _{IPh2} [nug som Rado lie nupes.] |

Figure 8. (A) Pseudosentences: Stylized course of F0 (in Hz) for Conditions A3 (dotted) and B3 (solid). (B) Pseudosentences: Sentence, part, and pause durations for Conditions A3 (gray) and B3 (black).



Condition B3 in contrast to Condition A3. At the end of the second phrase, IPh boundaries are detectable in both conditions. Thus, the prosodic structuring of the pseudosentence material resembles that of normal and jabberwocky sentences. Condition A3 comprises one IPh boundary (at about 2000 msec) and Condition B3 comprises two IPh boundaries (at about 920 and 2400 msec).

Experiment 4: Hummed Sentences

In Experiment 4, the processing of isolated prosody was investigated. For this purpose, sentences with the same intonation contour as the normal sentences were hummed by the speaker. Humming the material assured that it neither contained phonemic variation nor syntactic or semantic information, whereas the overall intona-

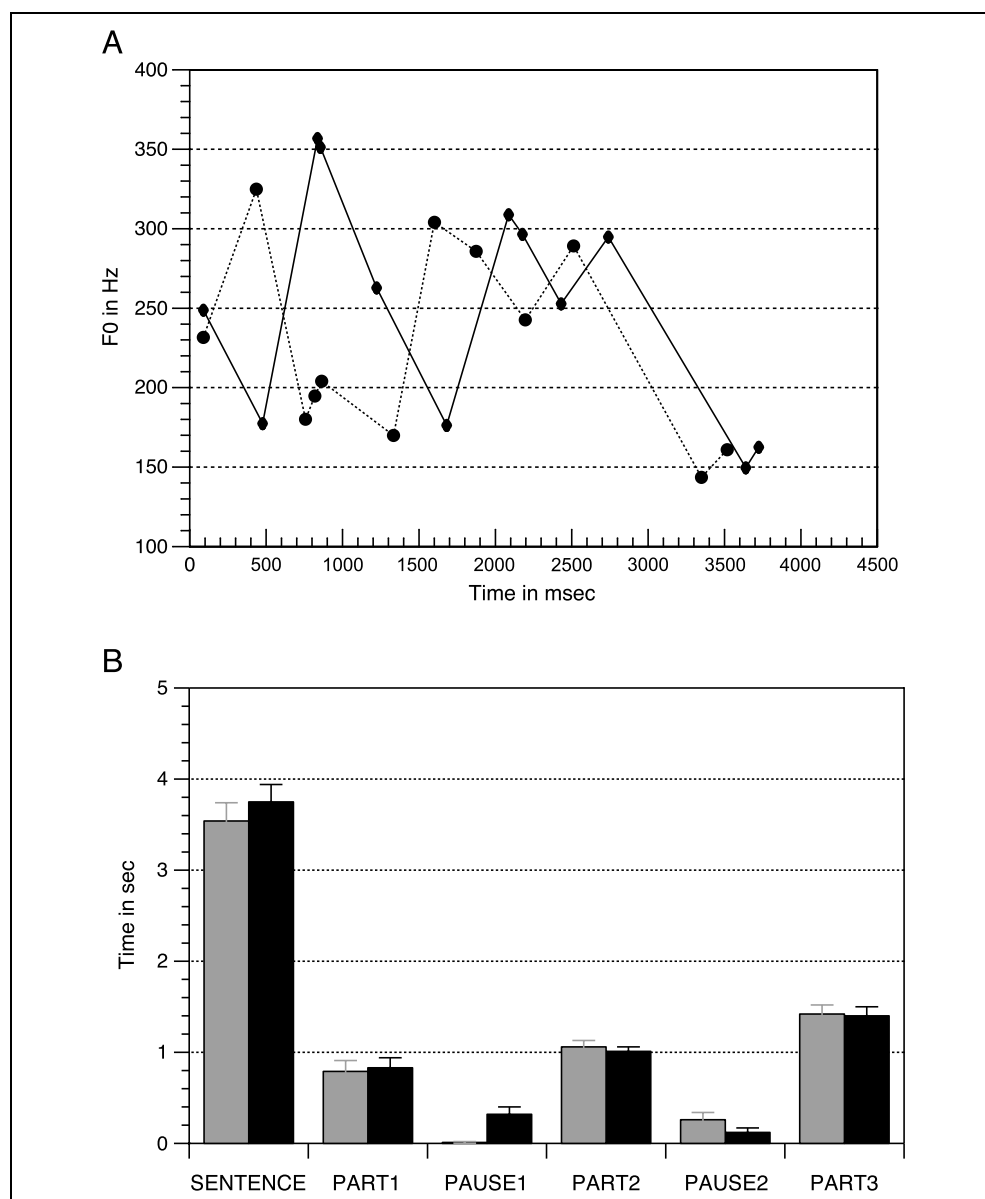
tion contour was similar to the one of the normal sentences (Table 5).

Figure 9A shows that the end of the defined first part is marked by a high tone in Condition B4, whereas Condition A4 does not exhibit this tonal pattern. The length of the first part does not differ significantly between conditions. However, the pause following the fragment differs significantly in length [Condition A4 < Condition B4, $t(94) = -25.30$, $p < .01$, between the two hummed conditions]. In the position after

Table 5. Examples of the Hummed Material in Experiment 4

| | |
|--------------|----------------------------------------------------------------------------------------|
| Condition A4 | [mm mmm mmmm mm mmmm] _{IPh1} [mmmm mmm mmm mmm mmmmm.] |
| Condition B4 | [mm mmm] _{IPh1} , [mmmm mm mmmm] _{IPh2} [mmmm mmm mmm mmm mmmmm.] |

Figure 9. (A) Hummed sentences: Stylized course of F0 (in Hz) for Conditions A4 (dotted) and B4 (solid). (B) Hummed sentences: Sentence, part, and pause durations for Conditions A4 (gray) and B4 (black).



the second part, both conditions comprise an IPh boundary as reflected by tonal and durational parameters. Hence, it is indicated that Condition A4 only contains one IPh boundary (at approximately 1850 msec) and Condition B4 two boundaries (at about 850 and 2150 msec).

Procedure

For each sentence type (normal, jabberwocky, pseudo, and hummed), separate experiments were conducted. The procedure in all four experiments was the same. Acoustic single sentence presentation with a recognition memory task was used for all different variations. The two conditions were presented in a randomized order within every experiment. For the memory task in the hummed variation, filler sentences were constructed by

splicing real words into the existing hummed versions. The position of these “real words” within the hummed sentences was random to convey the impression that they belong to an entire spoken phrase. These fillers did not enter the data analysis. Participants were seated in a soundproof chamber in front of a PC monitor. Every kind of sentence was presented via loudspeakers. Participants were instructed to look at a fixation point in the middle of the monitor throughout the whole experiment and to blink only during the presentation of the probe words.

Each trial in all experiments started with the auditory presentation of a sentence. After an interstimulus interval of 1500 msec, the probe followed. Participants had to decide whether the probe had been a part of the preceding sentence or not by pressing a corresponding button as soon as possible.

An experimental session lasted about 1.5 hours, including electrode application.

Event-related Brain Potential Recording

Electroencephalogram (EEG) was recorded from 23 cap-mounted Ag/AgCl electrodes. Electrodes were placed at F7, F3, Fz, F4, F8, Ft7, Fc3, Fc4, Ft8, T7, C3, Cz, C4, T8, Cp5, Cp6, P7, P3, Pz, P4, P8, O1, and O2. The electro-oculogram (EOG) was recorded from electrodes placed at the outer canthus of each eye and from sites above and below the right eye. On-line, the system was referenced to the left mastoid, and off-line, the system was re-referenced against linked mastoid. A ground electrode was placed on the participants' sternum. The EEG and EOG were acquired with XREFA amplifiers at a sampling frequency of 250 Hz.

Data Analysis

The EEG data were processed with the software packages EEP 3.2 (Max Planck Institute of Cognitive Neuroscience). EEG epochs containing eye blinks or movement artifacts were rejected from the data and did not enter the ERP averages. After the artifact rejection, at least 38 trials per condition entered the following analyses. Averages were computed for 4500 msec across the whole sentences using a 200-msec prestimulus baseline. In a first step, averages per condition per subject were calculated, and in a second step, means were estimated across subjects.

For the statistical analysis, ANOVAs were computed for the midline and lateral electrodes separately for each experimental stimulus variation. At the midline electrodes (Fz, Cz, and Pz), a two-way ANOVA was chosen with the factors condition (A vs. B) and Region (frontal vs. central vs. parietal). For the lateral electrodes, six ROIs were defined by crossing the two factors region (frontal vs. central vs. parietal) and hemisphere (left vs. right). Each ROI included three electrodes: frontal left (F7, F3, Fc3), frontal right (F8, F4, Fc4), central left (T7, C3, Cp5), central right (T8, C4, Cp6), parietal left (P7, P3, O1), and parietal right (P8, P4, O2). A three-way ANOVA design with the factors condition (A vs. B), region (frontal vs. central vs. parietal), and hemisphere (left vs. right) was applied. In cases in which interactions of the factor condition with topological factors could be observable, separate analyses were computed for hemispheres, regions, or ROIs, respectively. ANOVAs were computed in TWs of 500 msec.

Acknowledgments

This work was supported by the German Language Development Study (German Research Foundation, DFG; FR-519/18-1) awarded to A. D. Friederici, a grant from the German Research Foundation (FR 519/17-3) awarded to A. D. Friederici and K.

Alter, and a grant from the Human Frontier Science Program (HFSP RGP5300/2002-C102) awarded to K. Alter.

Reprint request should be sent to Angela D. Friederici, Max Planck Institute for Human Cognitive and Brain Sciences, P.O. Box 500 355, 04303 Leipzig, Germany, or via e-mail: angelafr@cbs.mpg.de.

REFERENCES

- Baum, S. R., & Pell, M. D. (1997). Production of affective and linguistic prosody by brain-damaged patients. *Aphasiology*, 11, 177–198.
- Baum, S. R., & Pell, M. D. (1999). The neural basis of speech prosody: Insights from lesion studies and neuroimaging. *Aphasiology*, 13, 581–608.
- Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, 30, 644–663.
- Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11, 17–67.
- Collier, R., & 't Hart, J. (1975). The role of intonation in speech perception. In A. Cohen, & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 107–121). Heidelberg: Springer-Verlag.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge: Harvard University Press.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201.
- Dykstra, K., Gandour, J., & Stark, R. E. (1995). Disruption of prosody after frontal lobe seizures in the nondominant hemisphere. *Aphasiology*, 9, 453–476.
- Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, 123, 155–163.
- Kennedy, A., Murray, W. S., Jennings, F., & Reid, C. (1989). Parsing complements: Comments on the generality of the principle of minimal attachment. *Language and Cognitive Processes*, 4, 51–76.
- Kreiman, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics*, 10, 163–175.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology*, 45A, 73–87.
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, 17, 73–88.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9, 97–113.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. MIT Linguistics PhD thesis, Indiana University Linguistics Club, Bloomington, IN.
- Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, 11, 165–192.
- Rooij, J. J. de (1975). Prosody and the perception of syntactic boundaries. *IPO Annual Progress Report*, 10, 36–39.

- Rooij, J. J. de (1976). Perception of prosodic boundaries. *IPO Annual Progress Report, 11*, 20–24.
- Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge: MIT Press.
- Sonntag, G. P., & Portele, T. (1998). PURR—A method for prosody evaluation and investigation. *Journal of Computer Speech and Language, 12*, 437–451.
- Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*, 191–196.
- Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research, 30*, 267–295.
- Stirling, L., & Wales, R. (1996). Does prosody support or direct sentence processing? *Language and Cognitive Processes, 11*, 193–212.
- Tecce, J. J., & Cattanach, L. (1987). Contingent negative variation (CNV). In E. Niedermeyer, & F. Lopes da Silva (Eds.), *Electro-encephalography: Basic principles, clinical applications and related fields* (pp. 658–679). Munich: Urban & Schwarzenberg.
- Tzourio, J. P., Massiou, F. E., Crivello, F., Joliot, M., Renault, B., & Mazoyer, B. (1997). Functional anatomy of human auditory attention studied with PET. *Neuroimage, 5*, 63–77.
- Van Lancker, D., & Sidtis, J. J. (1992). The identification of affective–prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not equal. *Journal of Speech and Hearing Research, 35*, 963–970.
- Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology and parsing in closure ambiguities. *Language and Cognitive Processes, 10*, 457–486.
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*, 946–953.
- Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*, 846–849.