PROSODIC PROPERTIES, PERCEPTION, AND BRAIN ACTIVITY

Karsten Steinhauer, Kai Alter, and Angela D. Friederici
Max-Planck-Institute of Cognitive Neuroscience, Leipzig, Germany
{steinhau, alter, angelafr}@cns.mpg.de

ABSTRACT

This paper investigates how differences in syntactic structure influence the speaker's prosodic realization of temporarily ambiguous utterances and whether the respective prosodic information guides the listener's sentence comprehension. Exhaustive acoustic analyses of the speech signals as well as behavioral and event-related brain potential (ERP) data of 56 listeners revealed the following results. 1. As predicted by certain theories of syntax-prosody mapping, syntactic differences led to early characteristic changes in the prosodic pattern. 2. Prosodic differences involved word duration, pause insertion, pitch contours, and the loudness function of the speech signals. 3. The disambiguating prosodic cues were immediately decoded by the listeners and prevented them from initial misanalyses typically observed during reading. 4. The processing of Intonational Phrase (IPh) boundaries was reflected by a specific brain response in the ERP. 5. In the presence of other prosodic cues, pause insertion was completely dispensable for the marking and perception of IPh boundaries.

1. PROSODIC PROPERTIES

Speech, in contrast to written text, provides prosodic cues in order to express both linguistic (e.g. syntactic) and non-linguistic (e.g. affective) information. In order to realize a certain prosodic effect (e.g. accentuation, prosodic phrasing, etc.), speakers can use a variety of prosodic parameters such as pause insertion, constituent lengthening, and pitch or loudness variations (Cutler, Dahan & van Donselaar, 1997; Alter, Steinhauer & Friederici, 1998). The listener, on the other hand, has to decode and to integrate these different parameters in order to achieve full comprehension.

1.1. Syntax-prosody mapping

With respect to linguistic prosody, theories of syntax-prosody mapping assume a more or less direct relationship between the hierarchical syntactic organization of a sentence and its prosodic realization. According to Jacobs (1993), accent positions in terms of their relative prominence (e.g., the weight of accents distributed across a syntactic structure) are calculable once the syntactic structure is known. This approach predicts early prosodic differences for sentences which are temporarily ambiguous due to different directions of branching. This is the case in an Object-Verb-language (OV) such as German. In structures with verb-final word order, i.e. in most subordinate clauses, the verb takes its arguments from its left (*'Anna'* in **B**), whereas the verb of the main clause takes its arguments from its right (*'Anna'* in **A**).

'Anna' as indirect object of verb_1:

(A) [**IPh1** *Peter verspricht Anna zu* ARBEITEN]
Peter promises Anna to work
[**IPh2** *und das Büro zu putzen.*]
and to clean the office

'Anna' as direct object of verb_2:

(B) [**IPh1** *Peter verspricht #*] [**IPh2** *A*NNA *zu entlasten*]
Peter promises # to support Anna
[**IPh3** *und das Büro zu putzen.*]
and to clean the office

Note that both conditions are structurally ambiguous up to the verb_2. In (A), the second verb *'arbeiten'/'to work'* is intransitive and NP2 *'Anna'* is the indirect object of the preceding verb_1*'verspricht'/'promises'*. Here, the second verb is accented. In (B), by contrast, *'Anna'* is demanded as direct object by the subsequent transitive verb_2 *'entlasten'/'to support'*. The syntactic structure in (B) requires a deeper embedding of the NP2 *'Anna'* which therefore receives the main accent (marked by small CAPITALS in A/B) when applying the algorithm proposed by Jacobs (1993).

This can be illustrated via bracketed metrical grids assigning the highest column of beats ('*') to the designated constituents (cf. **Figure 1**). The highest column indicates the position of the main accent. One advantage of using bracketed metrical grids is the possibility to translate syntactic constituents directly into prosodic domains. Brackets in the metrical grid mark boundaries of Intonational Phrases (IPh). The relevant part of (A) consists of only one IPh, whereas (B) is prosodically restructured with an additional IPh boundary after the first verb.

Peter verspricht....

```
            *                          *
            *     )                   (*              )
      *     *     ))                  (*        *     ))
... Anna zu ARBEITEN ]IPh1           [ANNA zu entlasten]IPh2
 (A)                                  (B)
```

Figure 1: The bracketed metrical grids for the relevant parts of (A) and (B). In (A), the highest column of beats is assigned to the lexically stressed syllable of verb_2 *'arbeiten'* whereas in (B),

the highest column is assigned to the lexically stressed syllable of NP2 *’Anna’*.

1.2. Association with Tonal Sequences

According to recent tonal sequence models (Reyelt, Grice, Benzmüller, Mayer, and Batliner 1996 for German), the main accent positions derived from the syntactic structure serve as anchor points for the association of tonal sequences. In intonational languages such as German, accents can be assumed to be realized preferably by tonal/pitch variations.

We refer to the German-ToBI system (Reyelt et al 1996) in order to predict the correct tonal sequences: Concerning the conditions (A) and (B), we assume the main accents to be associated with rising tonal sequences of the type L+H*. Note that the L+H*-sequence is associated via the metrical grid with the lexically stressed syllable of verb_2 in (A), and of NP2 ‘Anna’ in (B).

Furthermore, we expect the IPh boundaries to be marked by boundary tones. Sentence internal boundaries are marked by high boundary tones (H%) and/or durational parameters such as pause insertions ('#'). In condition (A), only one prosodic boundary appears after verb_2 whereas in the condition (B) an additional boundary is expected between the first verb and NP2 *’Annd*.

L+H* H%
(A) [*Peter verspricht Anna zu ARbeiten #*]

H% L+H* H%
(B) [*Peter verspricht #*][*ANna zu entlasten #*]

To summarize the predictions of syntax-prosody mapping, three predictable prosodic parameters have to be distinguished, namely (1) accent position in terms of relative metrical prominence, (2) accent type in terms of the association of prominence with tonal sequences and (3) boundary marking in terms of the tonal or durational realization of the edges of IPhs.

1.3. Results of the Acoustic Speech Signal Analyses

48 sentences of both conditions (A) and (B) were produced by a female speaker of standard German and recorded in a soundproof chamber. Each digitized speech signal (44.1 kHz/16 bit sampling rate) was analyzed with respect to word and pause durations, pitch contour (fundamental frequency, F0), and loudness function (amplitude squares), and then statistically analyzed in paired *t*-tests or ANOVAs. The data clearly confirm the predictions derived from the models of syntax-prosody mapping. Prosodic differences between conditions A and B occurred long before the structures were disambiguated lexically by the verb_2.

1. The different accent positions, i.e. *’arbeiten’* in (A) versus *’Anna’* in (B), were realized by both local pitch maxima and local loudness maxima (cf. **Figure 2**). As expected, both parameters revealed significant interactions between condition and accent position ($p<0.0001$ and $p<0.01$, respectively).

2. The additional IPh boundary in condition (B) was marked by a significant pause insertion of some 150 ms between verb_1 and *’Anna’* ($p<0.0001$). Moreover, the sentence initial constituent *’Peter verspricht’* preceding the boundary was considerably lengthened in (B) as compared to (A) ($p<0.0001$).

(A)
(B)

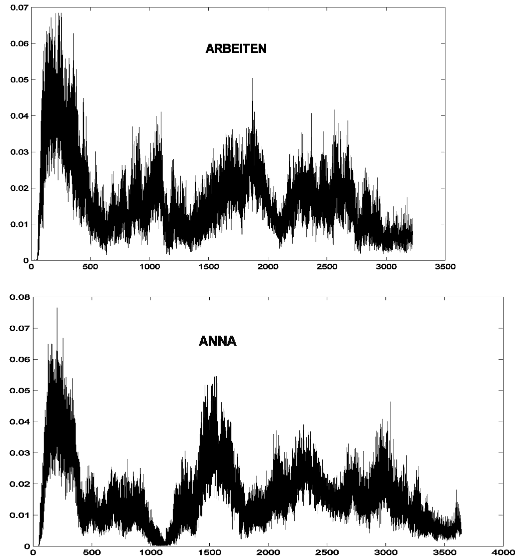

Figure 2: Loudness functions across the whole sentence in condition (A) (upper panel) and condition (B) (lower panel). Accent positions correspond to local amplitude maxima (on *’arbeiten’* in A, and on *’Anna’* in B). Due to the constituent lengthening in (B), however, both maxima occur at approximately the same time.

2. PERCEPTION

2.1. Misanalyses during Reading: The Garden-Path Effect

When locally ambiguous sentences such as (A) and (B) are presented visually, readers tend to interpret the NP2 *’Anna’* initially as the object of the preceding rather than the subsequent verb, as required in (B). This initial preference requires a revision and results in increased reading times in (B), a phenomenon called the garden-path effect (Frazier & Rayner, 1982; Steinhauer & Alter, in preparation). However, whereas a reader can read the sentence again, listeners rely much more on the first analysis of the input. Our hypothesis was, that prosodic cues should prevent listeners from being led up the garden-path. In particular, the additional IPh boundary in (B) should secure an immediate correct attachment of *’Anna’* to the second verb. To test this hypothesis, we cross-spliced the speech signals resulting in a prosody-syntax mismatch condition (C):

(C) * [**IPh1** *Peter verspricht #*] [**IPh2** *ANNA zu arbeiten*]
Peter promises # to work Anna
[**IPh3** *und das Büro zu putzen.*]
and to clean the office

(C) consists of the same lexical elements as (A) but has the same initial prosody as (B). If prosody guides the reader’s initial analysis, this should result in a severe reversed garden-path for this normally easy-to-process sentence structure. When encountering the verb the signal should initially be perceived as *'*Peter promises to work Anna ...'* which is clearly not grammatical and requires a revision.

2.2. Behavioral Data

In the first two studies, 40 students listened to 48 sentences in each of the three conditions A, B, and C. Prosodic acceptability rates and comprehension data showed that both (A) and (B) were equally acceptable (>80%), whereas (C) was accepted in only 6% of the trials. Interestingly, this pattern did not change in the third experiment (N=16) in which we carefully removed the pause preceding *'Anna'* in conditions (B) and (C). Thus, even without the pause, the remaining prosodic cues were sufficient to mark the prosodic boundary and to reverse the preferred syntactic analysis. (The present data leave it open, however, what role a pause may play in the absence of other parameters.)

3. BRAIN ACTIVITY

3.1. The Brain at Work: Event-related Potentials (ERPs) and Language Processing

Most psycholinguistic research is based on behavioral studies measuring error rates and reaction times. However, due to its implicit on-line characteristics and its high time resolution the employment of ERP measures has joined the list of on-line methods as an additional approach to study language processing. ERPs are a transient change of voltage, reflecting a systematic brain activity which is triggered by a physical event.

If this event is a word presented either in a semantically appropriate context or in an inappropriate context (e.g. *'He spread the warm bread with* ***socks****'*), the ERP differences between the two conditions reflect the brain's activity while processing a semantic violation. This experiment was conducted by Kutas and Hillyard (1980) who found a negative brain potential around 400 ms after onset of the anomalous word, the N400 component. The N400 has been replicated in hundreds of studies and reflects difficulties in lexical/semantic integration. The costs of processing syntactic difficulties, on the other hand, generally elicit a late positivity around 600 ms (P600) rather than an N400 (e.g. Osterhout & Holcomb, 1992). Thus, in contrast to response times, ERP patterns can more easily distinguish between different levels of linguistic processing. Here we used ERP measures to monitor the processing of prosodic features.

3.2. A Brain Response to Prosodic Boundaries

Brain potentials of all 56 participants in the three experiments were recorded while they listened to the sentences. As expected, the disambiguating verb_2 *'arbeiten'* of the mismatch condition (C) elicited a biphasic N400-P600 pattern of components, which is typical for verb-argument structure violations. These components reflect the reverse garden-path effect (Steinhauer, Alter, & Friederici, 1999).

The most exciting finding was that the ERPs displayed a large positive potential shift at IPh boundaries. Whereas only one such shift was found in (A), both conditions (B) and (C) displayed an additional positive shift at the additional boundary preceding *'Anna'*. We termed this component *Closure Positive Shift* (CPS) as we assume that it reflects the closure of Intonational Phrases. An alternative explanation was that it reflects simply the lack of speech input during the pause. In this case, one would expect that the CPS should not be present after pause removal in the third experiment even though the boundary was perceived by the listeners and did influence their syntactic analysis. The ERP data of the third experiment clearly reject this alternative hypothesis. As in the previous studies, the *Closure Positive Shift* occurred exactly at the respective prosodic boundaries: once in condition (A) and twice in (B) and (C) (cf. **Figure 3**). The CPS can be viewed as a valuable tool to monitor prosodic processing. It is an on-line measure with high time resolution which, in contrast to behavioral measures, does not require to interrupt speech presentation in order to collect data.

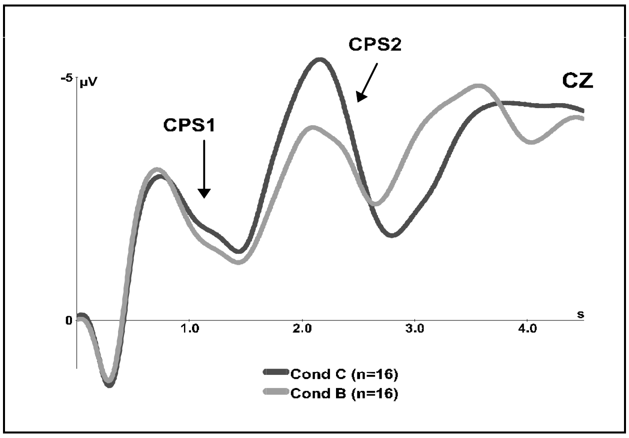

Figure 3: The two Closure Positive Shifts (CPS) at the CZ electrode in conditions (B) and (C) of Experiment 3, i.e. after pause removal. Negative amplitudes are plotted upwards.

4. CONCLUSION

We are just beginning to understand the relationship between prosodic and syntactic parsing and the underlying brain mechanisms. The present study demonstrates that a combination of acoustic analyses of the speech signals and psycho-physiological techniques may shed new light on this exciting aspect of language processing.

Acoustic analyses allow to test and adjust models of syntax-prosody mapping. Furthermore, they enable to systematically manipulate the speech signal in order to separate and investigate the contribution of single prosodic parameters such as pause insertions.

Behavioral and ERP data of 56 listeners show that prosodic differences are immediately decoded and used to prevent misunderstandings often observed during reading. This seems to be necessary as listeners, unlike readers, "lack the luxury of instant replay and may be severely misled by an incorrect initial parse" (Van Petten & Bloom, 1999). Given the importance of prosody in speech perception, it is not very surprising to find a specific brain response to prosodic boundaries, i.e. the CPS.

In contrast to purely behavioral measures from which intonational phrasing can only be indirectly inferred, the CPS in the ERPs provides an on-line indication of this phrasing. An additional advantage of ERPs is that the sentences can be presented as a whole. Behavioral on-line studies employing cross-modal naming tasks (e.g. Warren et al., 1995), present sentence fragments only and require to perform in a quite unnatural task. In contrast, ERP studies enable to examine prosodic processing as close to normal speech as possible.

ACKNOWLEDGMENTS

We thank Katja Kühn, Annett Schirmer, and Franziska Kopp for their assistance in generating and analyzing the speech signals. The work was supported by the Deutsche Forschungsgemeinschaft (FR 519/17-1).

REFERENCES

1. Alter, K., Steinhauer, K. & Friederici, A.D. De-accentuation: Linguistic environments and prosodic realizations. *Proceedings of the 5th International Conference on Spoken Language Processing.* Rundle Mall, South Australia: Causal Productions. Sydney, Australia (1998).
2. Cutler, A., Dahan, D. & van Donselaar, W. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* **40**, 141-201 (1997)
3. Frazier, L. & Rayner, K. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* **14**, 178-210 (1982).
4. Jacobs, J., „Integration," *Linguistische Arbeiten 306, Reis, M., (ed.), Wortstellung und Informationsstruktur*, Tübingen, 63-116 (1993).
5. Kutas, M. & Hillyard, St. A. Reading senseless sentences: Brain potentials reflect semantic incongruency. *Science* **207**, 203-205 (1980).
6. Osterhout, L. & Holcomb, P. J. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* **31**, 785-804 (1992).
7. Reyelt, M., Grice, M., Benzmüller, R., Mayer, J., and Batliner, A. Prosodische Etikettierung des Deutschen mit ToBI, In: *Gibbon, D. (ed.): Natural Language Processing and Speech Technology,* Mouton de Gruyter, Berlin, 144-155 (1996).
8. Steinhauer, K. & Alter, K. Are commas equivalent to intonational boundaries? (in preparation)
9. Steinhauer, K., Alter, K. & Friederici, A.D. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience,* **2**, 191-196 (1999).
10. Warren, P., Grabe, E. & Nolan, F. Prosody, phonology, and parsing in closure ambiguities. *Language and Cognitive Processes* **10**, 457-486 (1995).
11. Van Petten, C. & Bloom, P. Speech boundaries, syntax and the brain. *Nature Neuroscience,* **2**, 103-104 (1999).