PROSODIC CUES IN SYNTACTICALLY AMBIGUOUS STRINGS; AN INTERACTIVE SPEECH PLANNING MECHANISM

K. A. Straub

University of Rochester

ABSTRACT

This paper presents data suggesting that the clarity with which the speech signal is produced may be mediated by the presence of extra-syntactic factors such as the presence of strongly biasing context. An elicited speech experiment examining prepositional phrase attachment ambiguities produced within either strongly biasing or non-biasing contexts reveals that although the expected acoustic correlates of syntactic structure are consistently observed, the salience of these cues is diminished when alternate sources of disambiguating information are present for the listener. A second experiment shows that the cues are predictive with respect to listeners' interpretation of the ambiguous segments.

1. INTRODUCTION

Within the domain of sentence processing, the *garden-path* effect is the phenomena whereby a listener or reader pursues a structural analysis of the input which ultimately turns out to be incorrect (1). This effect, which is commonly experienced in sentences like:

i. The horse raced past the barn fell.
ii. John told the boy the girl kissed the story about the horse.

has been studied extensively in the visual input domain, showing that the difficulty experienced by the comprehender is diminished or even circumvented when various alternative sorts of ambiguity resolving information are present. These alternative information sources include lexically driven constraints (2,3), discourse information factors (4,5), local semantic constraints (6) and structural preferences. An additional source of parse-mediating information which has begun to enjoy renewed empirical attention is prosody. It has been long believed that certain parameters of the acoustic signal vary regularly with syntactic structure (7,8,9,10,11). Several recent studies examining the role of prosody in mediating the difficulties associated with structural ambiguities show that prosodic information can facilitate sentence comprehension (12,13,14). However, other studies with the same goal fail to find prosodic facilitation (15). Further, research which focuses on the reliability of prosodic demonstrates that their presence is a far from a perfect predictor of an intended syntactic structure (14,16).

The research presented in this paper accounts for these inconsistencies by postulating an interactive speech planning mechanism. This system estimates the communicative efficacy of an utterance by evaluating the presence of alternative sources of disambiguating information available to the listener. Based on this assessment, the speech planning system adjusts the control it exerts over the signal generation mechanism. When the probability of successful message transmission is low, the

signal production mechanism may marshal its resources to generate a more precisely controlled signal string. Conversely, when the probability of successful message transmission is high, cognitive resources can be shifted to other tasks (such as additional message planning) resulting in a less precisely produced and less acoustically informative speech signal. As such, this work takes as its inspiration previous research examining the interaction of predictive context and acoustic clarity at the lexical level (11,17) and extrapolates from that work to provide an account for the reported irregularities in prosodic cues at the syntactic level.

2. ELICITED SPEECH EXPERIMENT: EVALUATING PROSODIC CUES

In this experiment 4 untrained speakers, naive to the goals of the experiment read a sets of passages which were recorded for later acoustical analysis. Critical passages contained prepositional phrase attachment ambiguities such as the italicized sentence in Table 1. Such sentences are ambiguous across two readings, based on the location of attachment for the final prepositional phrase. On one reading (NP attachment) the man is smoking the cigar. On the other reading (the VP attachment), the chauffeur holds the cigar. Core stimulus sentences were selected based on norms showing no underlying interpretive preferences. Three discourse contexts were constructed for each item. One context strongly biased a VP attachment and the second strongly biased the NP attachment reading. A third context contained several sentences which could logically precede and follow the target sentence but which was non-biasing with respect to the interpretation of the target sentence. Contexts were normed to confirm their biasing effects. Example contexts are also presented in Table 1. Target passages were embedded in a series of unrelated filler passages to create a production script. Given the script, subjects were instructed to familiarize themselves with each passage and to produce it so that they were confident a listener would understand. Although subjects could choose to re-record passages, they were not permitted to review their own recordings.

VP-Attachment Context: The business man was pleased with the limousine service that he had found. The only thing that troubled him was that he never seemed to get the same driver. Some drivers were quite friendly, and the man didn't mind talking to them. Other drivers just drove him nuts. They either smoked or swore loudly at the traffic the whole ride. Today the man seemed to have gotten one of those irritating drivers. *The chauffeur annoyed the man with the cigar.* The driver listened to the man's reprimand, but he refused to stop smoking his cigar.

NP-Attachment Context: The man was pleased with the limousine service that he had found. The only thing that bothered him was that he sometimes had to share the car with other passengers. Yesterday he was riding with two other businessmen. One was reading the paper and the other was puffing on a big cigar and gazing out the window. When there were this many riders, the driver always seemed to irritate someone. This had been no exception: *The chauffeur annoyed the man with the cigar.* The driver listened to the man's reprimand calmly, but in the end, he refused to quit singing along with the radio.

Non-biasing Context: The driver had promised his boss that he would try to be more polite to the clients. Even though he was trying to be more polite, *the chauffeur annoyed the man with the cigar.* The man reported the chauffeur to the company and he was reprimanded again.

Table 1: An Example Stimulus Item

2.1. Results and Discussion

The durations of the collected utterances were relativized over minimal pairs durations (calculated within item, speaker, and context condition) to account for rate differences across utterance tokens. Phonetic analyzers blind to the goals of the study identified word boundaries from which pause durations were calculated. A control pause, between *chauffeur* and *annoyed*, was measured. Since there are no putative structural differences between the two reading of the sentence at this juncture, no duration differences were expected to obtain. In fact, no significant duration differences were identified at this location between the PP-attachment types.
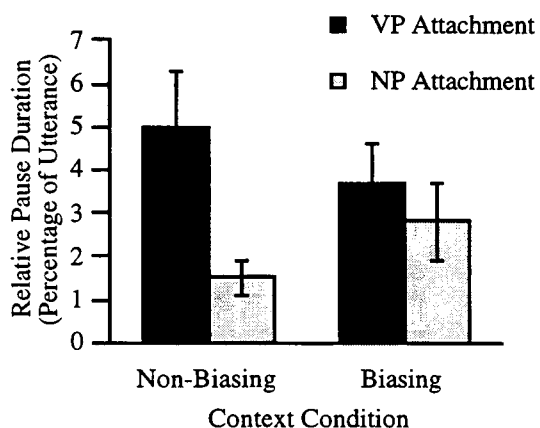

Figure 1: Critical Pause durations for Elicited Speech Experiment by Naming Word and Intended Attachment

In contrast, previous acoustic research (9, among others) predicts that because of identifiable phrase structure differences, the critical pause, located between *man* and *with* in the example, will be longer for the VP-attachment than the NP-attachment items. While this predicted result obtains within the items produced in the *absence* of strongly biasing context ($F_1(1,3)=11.85$, $p< .05$; $F_2(1,8)=3.29$, $p < .11$), it does not hold for the items produced within a strongly biasing discourse context (Fs<1). These results can be seen in Figure 1

In the utterances collected here, the precision of the acoustic signal was reduced in ambiguous strings whose intended interpretations were strongly supported by contextual information. In contrast, the precision of the acoustic signal for sentences whose communicative efficacy was not enhanced by contextual information was increased. This finding supports a description of a speech planning mechanism which can identify and evaluate the communicative efficacy of non-syntactic sources of disambiguating information. Based on its assessment of the probability of successful message transmission, the speech planning system adjusts the level of precision it exercises over the signal generator. The results obtained in this experiment suggest that the inconsistencies reported by researchers examining the regularity of the production prosodic cues to syntactic structure might be explainable if the circumstances under which the items produced are carefully examined: A seemingly reasonable strategy for eliciting orthogonal sets of readings of syntactically underspecified strings might be to embed those sentences into strongly biasing contexts. However, these findings suggest that employing such a strategy would reduce the likelihood of obtaining reliable prosodic cues.

The duration differences observed for utterances produced within a strongly biasing context are not statistically reliable, however, the differences consistently obtain in the predicted direction. Ferreira (19) suggests that "although prosodic information is not a perfect predictor of syntactic structure...the patterns are regular enough to be potentially useful to the comprehension system. (p.330)" It is feasible that the sentence comprehension system, ignorant of the strictures of statistical reliability is sensitive to these small differences. This possibility must be evaluated, since such a finding would diminish the explanatory power of an informationally integrated, speech planning system. To evaluate that possibility, the collected target sentences were presented to naive subjects in a comprehension study to evaluate the predictive efficacy of duration cues for sentence comprehension.

3. SPEECH PERCEPTION EXPERIMENT

In this experiment the target sentences produced in the previous experiment were excised from their discourse contexts and the final word of each sentence was removed. Items were then paired with two naming words matched on length, frequency and initial phoneme. The semantics of one naming word forced a reading in which the ambiguous PP was attached to the VP. The other naming word forced an NP attachment. Example (*iv*) provides an example one target string with its possible completion words.

high attachment: snickering
iv. The chauffeur annoyed the man with the
     low attachment: suspenders

These stimulus items were integrated into a set of items for use in a naming experiment. In this paradigm, subjects listen to a sentence fragment. At some critical point, the auditory presentation ends and a word is flashed on a computer screen. The subject's task is to read the word aloud as quickly as possible and then indicate, via a button press, whether or not the read word constituted a coherent continuation for the fragment. Half of the auditory fragments were paired with semantically anomalous naming words. Thus, subjects' "makes sense" judgments were be used to evaluate whether the subject was attending to the auditory stimulus. It is generally believed that naming latency reflects the ease with which a word is integrated into the current sentence (2, 18). Thus, if prosodic cues have a parse mediating influence, it is expected that a prosodic contour consistent with the semantics of the naming word (henceforth, *prosody-match*) should facilitate that integration process and reduce naming time. Conversely, a prosodic contour inconsistent with the semantic constraints imposed by the naming word (henceforth, *prosody-mismatch*) should cause an increased naming latency.

Since duration is not the only acoustic parameter underlying prosody it is possible that other cues, such as amplitude change, or F_0 contour, which were not evaluated in the previous experiment, or some *combined* (cue-trading) description is necessary to effectively predict the prosodic influence (14). If this is the case, the duration differences based on the availability of alternative disambiguating information sources observed in the previous experiment might not accurately predict the influence of prosody on the comprehension of those utterance strings. On the other hand, if pause duration is a reasonable predictor for the utility of any combination of prosodic cues in the determination of syntactically underspecified strings, significantly different naming latencies for the prosody-match versus prosody-mismatch conditions for items originally produced *without* biasing context should be observed. Further, items produced with the benefit of a biasing context should show no prosody-match/mismatch effects.

Twenty right-handed, native English, undergraduates participated in this cross-modal naming experiment in return for payment. They were tested individually. Voice trip latencies and makes sense judgments were collected for later analysis.

3.1. Results and Discussion

Response latencies more that 2.5 standard deviations away from the average response time for any subject were replaced by that outlier boundary value. Preliminary analysis showed no effect of speaker voice on response times. Although an ANOVA examining the effects of naming word type and prosodic contour, using both subjects and items as random factors showed no significant main effects, a interaction between the two variables was uncovered ($F_1(1,19)= 5.21$, $p<$ .05; $F_2(1,8)=5.61$, $p < .05$). As can be observed in Figure 2, further analysis reveals that the same interaction can be observed for items produced *without* the benefit of strongly biasing context ($F_1(1,19)=14.30$ p<.01; $F_2(1,8)=4.95$, p<.056), but not for items produced within strong contexts (Fs <1). No main effects obtained for either stimulus set.

Compared to items produced within strongly biasing contexts, items produced without the benefit of biasing context showed significant naming time facilitation for both prosody-match items ($t(613)=2.06$, $p < .05$) and prosody mismatch items ($t(613)=2.66$, However, within the Non-Biasing Context condition, prosody-match items were named significantly faster than Prosody-Mismatch items ($F_1(1,19)= 4.89$, $p< .05$; $F_2(1,8)=5.53$, $p < .05$). No parallel differences were observed between the match and mismatch conditions for items produced in biasing contexts.
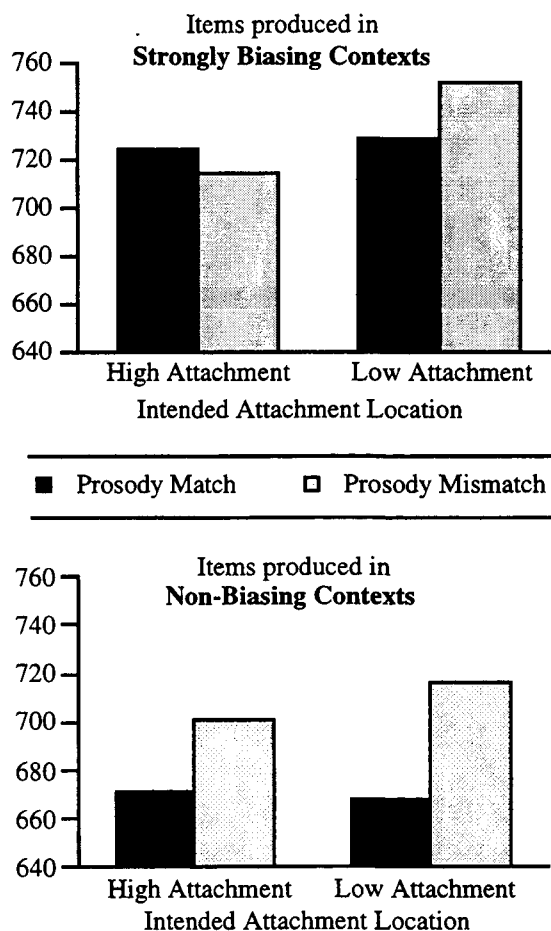

Figure 2: Speech Perception Experiment Results: Naming Latencies by Context and Prosody-Match Condition

These results support the use of pause duration as a metric for predicting the influence of prosodic cues on syntactic ambiguity resolution. Naming showed no prosody match/mismatch differences for the utterances produced in strongly biasing context, just as acoustic evaluation of those utterances revealed no reliable pause differences at the critical phrasal juncture. Conversely, utterances originally produced without the benefit of strongly biasing context showed both overall facilitation for naming responses as well as significant facilitation for prosody-match items over for prosody-mismatch items. Although this by no means, suggests that pause duration is the *only* acoustic cue that listeners exploit to derive

syntactic structure, it strongly suggests that pause duration is reliable predictor of the utility of those other cues.

4. General Discussion

The production experiment shows that the availability of context as an source of ambiguity resolving information influences the level of control which the speech planning system exerts over the signal generator. When the probability of successful message transmission is very high, such as when strongly biasing context supports the intended parse of a structurally ambiguous string, the planning system can exploit that redundancy by reducing the precision exercised over signal generation. However, when the likelihood of communicative efficacy is low, such as when a structurally anomalous string is to be produced without the benefit of strongly biasing context, the speech planning system notes that potential difficulty and assigns additional cognitive resources to the task of generating more acoustically informative speech signals.

The validity of the measure of pause duration as a predictor of prosodic facilitation, and as such the validity of positing an interactive speech planning mechanism, was evaluated through an on-line measure of the influence of the prosodic cues on sentence comprehension. In that study, word naming latencies were facilitated when the to-be-named word was consistent with the parse cued by the prosodic contour and were delayed when the word was inconsistent with the prosodically cued attachment. However, this naming latency difference was observable only for items originally produced without the support of a biasing context. Items originally produced in a biasing context showed no naming latency effects for either prosody match or mismatch conditions.

Taken together, the results of these experiments strongly support the position that the mediating effect of strongly biasing context previously observed at the lexical level (11,17) also applies at the level of structural ambiguity.

References

1. Bever, T.G. "The Cognitive Basis for Linstuistic Structures," In Hayes, J.D. (Ed.) **Cognition and the Development of Language.** Wiley: 1970.
1. Boland, J.E. "The role of verb argument structure in sentence processing: Distinguishing between syntactic and semantic effects", *J. Psycholing. Res., Vol. 22, 1993, 133-152.*
2. Trueswell, J. & Tanenhaus, M. "Tense, temporal context and syntactic ambiguity resolution", *L.C.P., Vol. 6, 1991,, 303-338.*
3. Spivey-Knowlton, M., Trueswell, J., and Tanenhaus, M. "Context effects in parsing reduced relative clauses", *Can. J. of Exp. Psy. Vol. 47, 1993, 276-309.*
4. Altman, G. & Steedman, M. "Interaction with context during human sentence processing", -*Cognition, Vol. 30, 1988, p. 191-238.,*
5. Spivey-Knowlton, M., and Tanenhaus, M. "Referential Context and Syntactic Ambiguity" in Clifton, C. (Ed.) **Perspectives on Sentence Processing.** LEA Press: 1994

6. Cooper, W. and Sorrenson, J. **Fundamental Frequency in Sentence Production.** Springer Verlag: 1981.
7. Lehiste, I, Olive, J. and Streeter, L. "Role of duration in disambiguating syntactically ambiguous sentences", *J. Acoustic. Soc. Amer, Vol. . 60, 1199-1202.*
8. Scholes, R., **Acoustic Cues for constituent structure: A series of Experiments on the Nature of Spoken Sentence Structures.** Mouton:1971.
9. Goldman-Eisler, F. **Psycholinguistics: Experiments in Spontaneous Speech.** Academic Press: 1968.
10. Leiberman, P. "Some effects of semantic and grammatical context on the production and perception of speech", *Language and Speech, Vol. 6,1963, 172-17.*
11. Schlepmann, A., Altmann, G. and Garnham, A., "The Status of Prosodic Cues in Early Sentence Processing: A Time Course Analysis". Poster presented at the 9th Annual CUNY Conference on Human Sentence Processing, 1996
12. Dobroth, K., and Speer, S. "A time course analysis of prosodic effects on the resolution of closure ambiguity", Poster presented at the 6th Annual CUNY Conference on Human Sentence Processing, 1995.
13. Beech, C.M. "The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cued trading relations", *J. Mem. & Lang, Vol. 30, 1991, p.644-663*
14. Watt, S. and Murray, W., "Complement sentence parsing in the auditory modality", Poster presented at the 6th Annual CUNY Conference on Human Sentence Processing, 1993.
15. Fierrera, F. "The creation of prosody during sentence production", *Psy. Rev., Vol. Vol. 100, 1993, 233-253.*
16. Fowler, C. "Differential Shortening of Repeated Content Words Produced in Various Communicative Contexts", *Lang. and Speech, Vol. 31,* 1988, 307-319.
17. Marlen-Wilson, W. and Tyler, L. "Processing Structure of Sentence Perception" *Nature, Vol. 257, 1975, 784-786.*
18. Ferreira, F, Henderson, J., Anes, M., Weeks, P., and McFarlane, D. "Effects of Lexical Frequency and Syntactic Complexity in Spoken Language Comprehension: Evidence from the Auditory Moving Window Technique", *J.E.P.: L.M.C., Vol. 22,1996, 324-335*

Please direct questions or comments to:
K. Straub
Cognitive Science
Johns Hopkins University
Baltimore, MD 21218
kath@mail.cog.jhu.edu