ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2004
Nara, Japan
March 23-26, 2004

The Role of Prosody in Parsing Ambiguous Sentences

Nivedita Mani

Phonetics Laboratory
University of Oxford
nivedita.mani@stcatz.ox.ac.uk

Abstract

This paper tests whether listeners are able to use the prosodic characteristics of speech to differentiate between alternative interpretations of syntactically ambiguous stimuli. Most existing research has either employed off-line tasks or provided adequate syntactic information for the listener to recognise ambiguity but inadequate prosodic information to resolve it. In incorporating controls for these limitations, my experiment was able to show that prosodic cues are able to guide initial processing of input irrespective of any putative syntactic parsing preferences.

1. Introduction

This experiment investigates the role of prosody in parsing ambiguous sentences such as:

Packing cases are always newsworthy (1)

Packing cases is always newsworthy (2)

Are listeners intuitively able to differentiate between alternative interpretations of the ambiguous stimuli in (1) and (2) using their knowledge of the prosodic cues associated with the first two words of the structure? In (1), the words *Packing* and *cases* form a phrasal collocation that refers to travelling cases. In (2), however, they form a verb phrase indicating the act of packing. The pitch accent of the phrasal collocation in (1) would typically be placed on the first syllable of the first word, while it would be on first syllable of the second word in (2).

2. Previous Research

The experimental task employed here uses a variation of an on-line response time task described in Marslen-Wilson et al 'Prosodic effects in Minimal Attachment' [9], testing the ability of prosodic cues to disambiguate sentences of the kind:

The workers considered the last offer from the management was a real insult (3)

The workers considered the last offer from the management of the factory (4)

Subjects in their experiment were played the italicised parts of both sentences and the visual probe 'was' followed each auditory input. The probe would be an appropriate continuation of the italicised part of (3), but not of (4). The time taken by listeners to name the probe was measured in both cases. It was found that subjects took longer to name the probe when it followed (4) than when it followed (3) – 397 ms as opposed to 375 ms. If subjects parsed the auditory input according to the default syntactic parsing mechanism suggested by Frazier [5],[6],[7] then they would first have had to construct a minimal attachment or a direct object parse of the stimuli. This would conflict with the continuation offered by the visual probe. Subjects would then have to **re-parse** the sentence as a non-minimal attachment or a complement clause. This would mean that subjects would always take longer to respond to the probe after (3). However, subjects took longer to respond to the stimulus in (4), the direct object reading, than to (3), the complement clause reading. This suggests that subjects were not automatically constructing an initial syntactically motivated direct object parse of all stimuli. Marslen-Wilson et al conclude that subjects were using the prosodic cues contained in the clauses to guide parsing towards the appropriate construction.

Marslen-Wilson et al's work [9] lays the foundation for two areas of research on the influence of prosody on sentence parsing. The first continues with the trend of the experiment detailed above, presenting evidence documenting the influence of prosody on speech processing using on-line and off-line tasks – a distinction that is central to my own research. The second concentrates on the effects of more specific prosodic cues, such as the effect of duration, pitch accent, *f*0 values on speech processing.

I argue that there are two methodological problems with the research conducted so far. These prevent a conclusive result in favour of prosodic first pass analysis independent of syntactic input: Firstly, most studies have used comprehension, paraphrasing or contextual assignment tasks [1],[2],[3],[8],[11],[12],[14],[15]. However, such off-line tasks do not necessarily test the initial processing of the input and are more pertinent to research on the final stages of processing.

Moreover, experiments using on-line tasks [8], [9], [10], [11], [16], [18] have incorporated adequate syntactic information to allow Frazier's default parsing preferences [5], [6], [7] to come into play. In such cases, there may be a period where syntactic ambiguity can be resolved and where there may not be adequate prosodic cues supplied to the parser to resolve suprasegmental ambiguity. Since Frazier's strategies can take over during this period, such an approach allows us to suggest that the first pass structure assignment might have been guided by the syntactic cues available. My experiment is the first to incorporate controls for these two caveats to prosodic first pass analysis.

3. Method

As suggested by Marslen-Wilson et al [9], I tested sentence-initial syntactically ambiguous word pair fragments that contrasted in prosodic information in order to provide prosodic information simultaneously to syntactic information. The contrasting data was taken from Tyler and Marslen-Wilson's experiment [17].

Each fragment consisted of word pairs such as:

Packing cases (5)

16 repeats of both prosodic alternations of the fragments were included. As soon as the fragment ended, a visual probe appeared on a screen in front of the subject. This probe would either be *is/are*. The probes contributed to the parse intended by the prosodic contour fifty percent of the time and conflicted with it otherwise. The probe *is* following the first two words of (1) and the probe *are* following the first two words of (2) both conflict with the prosodic contour of the intended stimulus. Conversely, the probe *is* following the first two words of (2) and the probe *are* following the first two words of (1) would be appropriate continuations of the interpretations intended by the speaker. Subjects were provided with handsets that had two buttons, one labelled *is*, and the other *are*. Subjects were told to press the button that corresponded to the probe on the screen as soon as it appeared. The time taken to respond to the probes was measured in each case. Half the stimuli – the test condition – were chosen so that the alternative interpretations were prosodically distinct, as in (1), (2) or (5) above. The rest – the control condition – were such that the alternative interpretations were prosodically similar, as in (6) and (7). Subjects were again played only the first two words of the sentences:

Frying eggs are always newsworthy (6)

Frying eggs is always newsworthy (7)

Subjects were given five seconds after they had responded to the probe to decide if they thought it was a 'good' or 'bad' continuation of the fragment.

4. Hypothesis

If the subjects are using prosodic information to construct an appropriate parse of the fragment, then I predict that naming latencies for the experimental condition followed by inappropriate probes (Phrasal Collocations followed by *is* and Verb phrases followed by *are*) will be greater than the naming latencies for the experimental condition followed by appropriate probes (Phrasal Collocations followed by *are* and Verb phrases followed by *is*). I suggest that this is because subjects are led into expecting the probe consistent with the prosodic contour of the clause and are therefore forced to recheck the inappropriate probe presented when it clashes with the prosodic information. This leads to an increase in response time. In the control condition, I predict that there will be no significant difference between the response times to appropriate and inappropriate probes. This is because neither of the probes would clash with the prosodic cues of the control sentences, as they are prosodically similar.

5. Results

18 out of 5760 responses were discounted because of incorrect button pressing (pressing the *are* button for an *is* cue). 64 responses later than 1500 msec were excluded in order to disallow any chance of reflective button pressing over immediate and unconscious responses. The difference in response times of the remaining measures is given in Table 1.

Table 1: Reaction times of match and mismatch experimental and control conditions tested (ms)

	Match	Mismatch
Expt Condition	551	573
Control Condition	560	564

The results of a one-way anova would have to be treated with caution, as the data were not normally distributed. Conversely, Independent sample t-tests do not rely quite so heavily on the normal distribution of the curve. A t-test analysis of the data revealed that response time does vary significantly with match and mismatch condition in the experimental category ($t(2822) = 2.97$; $p < .01$). This suggests that there is a 3 in 1000th chance that variance in response times is not affected by the match or mismatch of the probes to the experimental stimuli.

Conversely, the mean response times to the control sentences did not vary with the match or mismatch probes. An independent samples t-test of the response times to the control condition revealed that the response times of subjects do not vary significantly according to the appropriateness or inappropriateness of the probes as continuations of the audio input ($t(2852) = .542$; $p > .05$).

Since I could not assume that the data came from an underlying normal distribution, thereby undermining the assurance of the t-test results. I also performed nonparametric tests on the significance of the relation between response time and the match and mismatch of the continuation offered by the probes and the interpretation intended by the prosody of the stimulus. The results of a Mann Whitney test indicated that the response times of subjects to the experimental stimuli was significantly affected by the match or mismatch of the probe to the interpretation intended by the speaker ($z = 3.73$; $p < .01$). Conversely there was no significant interaction between response times and the match or mismatch of the probes in the control condition ($z = .75$; $p > .05$).

Further, using the medians of the subjects' responses to each individual stimulus, I also performed Wilcoxson Signed ranks tests on the significance of the interaction between response times of subjects and the match or mismatch of the probes. Again, the results established a significant interaction between the response times of subjects to the experimental stimuli and the match and mismatch of the visual probe to the interpretation intended by the speaker ($z = 2.44$; $p < .01$). The interaction between response times to the control stimuli and the match or mismatch of the probes to the intended interpretation was, however, not significant ($z = .69$; $p > .01$).

Figures 1 and 2 below illustrate clearly the difference in response times between the experimental and control conditions of the stimuli used in my experiment given in Table 1. The graphs are plotted at a 95% confidence interval to illustrate the difference between the two conditions. The mean response times of subjects to the match and mismatch control conditions are remarkably similar, while the mean response times to the contrasting experimental stimuli are significantly different. These error plots are centred on sample means +/- 1.96 times the standard error. It is clear that the slowest possible response time to the visual probe in the experimental match condition is still faster than the quickest possible response to the probe in the experimental mismatch condition. Conversely, the variation in response times around the mean in the control condition overlap considerably, as had been expected, and the slowest possible response to the probe in the match control condition is in fact slower than the slowest response to the mismatch condition, and much slower than the quickest possible response.

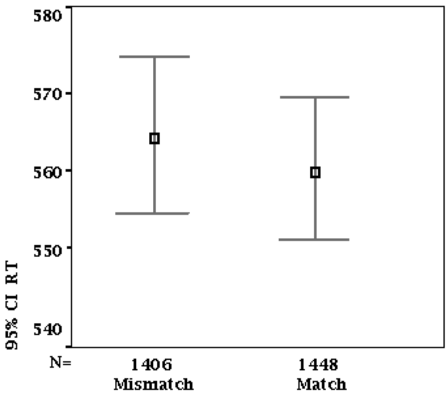

Figure 1. Control Condition: Error plot of match and mismatch against response time

6. Discussion and analysis of the results

The results show that response times were quicker when the visual probes were appropriate continuations of the parses indicated by the prosodic input than when they opposed the prosodic contour in the prosodically alternating conditions. There was almost no difference between the reaction times of the alternative interpretations of the control condition, indicating that subjects were not using syntactic or semantic information to distinguish between them. Furthermore, subjects did not tend towards constructing collocation parses in the control condition, as might have been expected if the tendency to associate any individual stress pattern had been learned and applied. The results of this experiment suggest that listeners do, in fact, have a repertoire of the effects that certain prosodic cues have on structure assignment. Given sufficient prosodic information, subjects **can** bring this knowledge to bear on constructing parses of input.

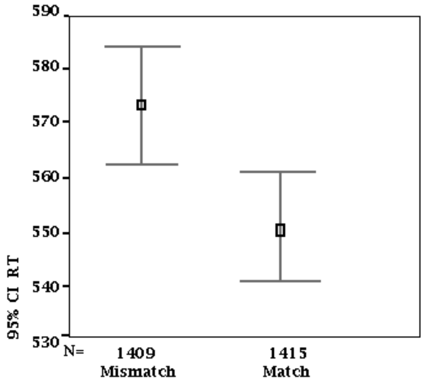

Figure 2. Experimental Condition: Error plot of match and mismatch against response time

As the stimuli did not contain any cues to syntactic preference [5],[6],[7] I propose that there is no reason to suggest that an initial syntactic construction of all plausible parses occurred. There would have been no way that either default syntactically motivated parse could have been chosen over the other for prosodic structure assignment. Structure was assigned in the test condition using almost solely prosodic cues. Moreover, in as much as parsing was conducted on-line, structure was being assigned to the stimulus before completion of the clause.

My research is able to conclude that first pass input analysis can be carried out using almost exclusively prosodic information. Subjects are able to use their knowledge of the correlations between suprasegmental and grammatical structure to parse the input that they receive. I suggest that all the information that is available to the parser is *accessible* simultaneously. Additionally I propose that the syntactic, semantic and prosodic information available is separately parsed in parallel. This information might even be parsed in different syntactic, semantic and prosodic modules, as suggested by Fodor [4]. Fodor had suggested the existence of a syntactic module. I advance his argument with my own proposal of prosodic and semantic modules. I suggest that upon receiving speech input the different modules immediately and involuntarily compete to create the first parse of the input. All the modules attempt to use the information specific to them to construct an initial parse. The module with enough information to construct a first parse presents it to be checked for compatibility with the other modules. At this stage, the other modules stop constructing their parses of the input, and assess the parse for compatibility with the information specific to them. They are now reduced to solely inhibitory action. If the syntactic module has adequate information to construct a parse before any of the other modules, then the other modules are reduced to merely inhibitory action (i.e. if the parse that is constructed is inconsistent with any information that they possess that is specific to them alone). Conversely, if the prosodic module has enough information to construct an initial parse, then the syntactic and pragmatic modules, among others, are reduced to merely inhibitory action. Interaction between the modules is restricted to inhibitory action post first pass analysis of the input. This avoids the criticism of first pass analysis in strong interaction models that it incorporates too much information [4], [5], [6], [7]. It is important to emphasize that the syntactic module is used to provide a description of the input, irrespective of whether the other modules are the first to construct a parse.

In the case of most utterances, the parse that is constructed by the module is compatible with the information that the other modules possess. However, garden-path situations may occur when one of the modules presents a parse with insufficient information, perhaps influenced by the frequency of production of one alternative over the other. The parse produced is incompatible with the inhibitory action of the other modules, which then triggers all the modules into competing to construct an appropriate parse from scratch.

No one of the modules has de facto priority over any of the others. Rather, priority is suggested by one of the components having information that helps it to parse the signal more appropriately and faster than the others. In my experiment, the prosodic parser was provided with more information than any of the other modules and was able to construct a first parse before the others. In other situations, for instance when contextual information exceeds other information, the semantic or pragmatic parser may produce an appropriate parse before the others. In the same vein, I suggest that experiments conducted by proponents of initial syntactic first pass analysis [3],[5],[18] have been able to produce supporting results because they provided the listener with information compatible with a syntactically motivated initial parse.

Bibliography

[1] 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.

[2] Cohen, H.; Douaire, J.; & Elsabagh, M., 2001. The role of prosody in Discourse Processing. *Brain and Cognition* 45, 73-82.

[3] Ferreira, F.; Anes, M.S.; & Horine, M.D., 1996. Exploring the use of Prosody during Language Comprehension using the Auditory Moving Window Technique. *Journal of Psycholinguistic Research* 25(2), 273-289.

[4] Fodor, J.A., 1983. *Modularity of the Mind*. Massachussets: MIT Press.

[5] Frazier, L., 1979. *On Comprehending sentences: Syntactic Parsing Strategies*. Bloomington: Indiana University Linguistics Club.

[6] Frazier, L., 1987. 'Sentence Processing: A Tutorial review'. In *Attention and Performance XII: The Psychology of Reading*. M Cotlheart (ed). Erlbaum.

[7] Frazier, L., 1990. 'Exploring the architecture of the language-processing system'. In *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. G.T.M Altmann (ed). MIT Press.

[8] Kjelgaard, M.M.; & Speer, S.R., 1999. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40, 153-194.

[9] 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.

[10] Nagel, H.; Shapiro, L.; Tuller, B.; & Nawy, R., 1995. Prosodic influences on the processing of attachment ambiguities. *Journal of Psycholinguistic Research* 25 (2), 319-343.

[11] Price, P.J.; Ostendorf, M.; Shattuck-Hafnagel, S.; & Fong, C., 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 90, 2956- 2970.

[12] Schafer, A.; Carlson, K.; Clifton, C,Jr.; & Frazier, L, 2000. Focus and the Interpretation of Pitch accent: Disambiguating embedded questions. *Language and Speech* 43 (1), 75-105.

[13] Shapiro,L.; & Nagel, H., 1995. Lexical properties, prosody, and syntax: Implications for normal and disordered language. *Brain and Language* 50, 240-257.

[14] Speer, S.R.; Shih, C.L.; & Slowiaczek, M., 1989. Prosodic Structure in Language Understanding: Evidence from Tone Sandhi in Mandarin. *Language and Speech* 32(4), 337-354.

[15] Speer, S.R.; Crowder, R.G.; & Thomas, L.M., 1993. Prosodic structure and Sentence recognition. *Journal of Memory and Language* 32, 336-358.

[16] Speer, S.R.; Kjelgaard, M.M.; & Dobroth, K.M., 1996. The influence of prosodic structure on the resolution of syntactic closure ambiguities. *Journal of Psycholinguistic Research* 25, 249-272.

[17] Tyler, L.K. & Marslen-Wilson, W.D, 1977. The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behaviour* 16, 683-692.

[18] Watt, S.M. & Murray, W.S., 1996. Prosodic form and parsing commitments. *Journal of Psycholinguistic Research* 25(2), 291- 317.