TEMPORAL CUES SUPPORT SYNTACTIC IDENTIFICATION

Michelle D. Wiley and M. Kathleen Pichora-Fuller
School of Audiology & Speech Sciences, University of British Columbia
5804 Fairview Ave., Vancouver, BC, V6T 1Z3

INTRODUCTION

Previously, we demonstrated that normal-hearing young listeners' comprehension of complex syntactic structures declined more than comprehension of simpler syntactic structures as the signal-to-noise ratio became more adverse, with these declines not being attributable simply to declines in word recognition (Dillon, 1995; Pichora-Fuller, Lloyd, Dillon, & Kirson, 1998). The present study was conducted to examine the roles of temporal and spectral cues in syntactic processing.

METHOD

Participants. Fifteen normal-hearing, female, native English-speakers between the ages of 21 and 29 were paid for their participation in Experiment 1. An additional 5 listeners participated in Experiment 2. Their reading working memory spans (Daneman & Carpenter, 1980) ranged from average (2.7) to excellent (5.7). All were university graduates with sufficient knowledge of linguistics to ensure familiarity with the syntactic classifications used in the experiment.

Materials. Sentence materials were prepared previously (Dillon, 1995) based on the sentences used in a comprehension test developed for use with aphasics (Caplan, Baker, & Dehaut, 1985). Dillon's materials consisted of three equivalent lists of 45 sentences; each list had 5 exemplars of each of 9 different syntactic types (Table 1).

Table 1: Examples of Syntactic Types

Code	Type	Example
A	Active	The duck chased the mouse.
P	Passive	The owl was tapped by the pig.
CS	Cleft Subject	It was the fox that touched the mouse.
CO	Cleft Object	It was the duck that the owl bumped.
D	Dative	The dog smacked the frog to the goose.
DP	Dative Passive	The pig was kicked to the duck by the fox.
C	Coordinated	The owl tripped the dog and grabbed the pig.
SO	Subject-Object	The mouse that the duck grabbed bumped the owl.
OS	Object-Subject	The pig kissed the owl that bumped the mouse.

For the present study, three new versions of the sentence lists were prepared. One list of sentences was spectrally-inverted. For another list, words read in citation form were recorded and then concatenated with a 50 msec silence between each word to form sentences with the same lexical (spectral) content but without normal sentence prosody. For a third list, the concatenated sentences without normal sentence prosody were also spectrally-inverted. All test materials were digitized at a sampling rate of 32 kHz and stored in soundfiles.

Following the technique described by Blesser (1972), spectral inversion was accomplished by band-pass filtering (.2-4 kHz, 48 dB/octave) an audio tape-recording of the intact sentences, applying spectral pre-emphasis of 24 dB/octave between .5 and 3.5 kHz, and then spectrally inverting (around 2.1 kHz) using customized in-house hardware (Benguerel, 1998). The consequence of the spectral inversion manipulation was to severely disrupt the spectral properties of speech while preserving its temporal envelope and some pitch information.

In a pilot experiment with 10 other participants, it was found that words and phrases excised from the spectrally-inverted sentences were almost never recognized when an open-set response format was used, whereas the corresponding segments excised from the intact version of the sentences were recognized nearly perfectly. However, following familiarization with the corresponding intact materials and using a close-choice test format with 34 alternatives, re-testing with the spectrally-inverted materials yielded scores close to 30% correct. The increase in scores seemed to be due mostly to the ability of the listeners to correctly identify the few multi-syllabic alternatives; of the 34 alternatives, 5 had two syllables and 2 had three syllables.

Conditions. In Experiment 1, three blocked conditions were presented in the following fixed order: first, an "intact condition" using the list of sentences with normal spectral and temporal cues; second, a "spectrally-inverted condition" using the list of sentences with reduced spectral cues; third, a "concatenated condition" using the list of sentences with reduced temporal cues. In Experiment 2, we tested a different group of listeners in a "concatenated-inverted condition" using the list of sentences with both spectral and temporal cues reduced.

Procedures. Sentences were presented monaurally at 70 dB SPL in quiet over TDH 39 earphones to listeners in a double-walled IAC booth. CSRE software and TDT hardware were used to control the experiment. An example of each of the nine possible syntactic types was displayed adjacent to the numbers 1 to 9 on a computer screen. Following presentation of each sentence, the listener used the computer mouse to select the number of the sentence structure they thought they heard. All subjects received a practice session prior to the test conditions.

RESULTS

Performance was virtually perfect in the intact and near-perfect (98.9% correct) in the concatenated condition. Even in the spectral inversion condition, most sentence structures remained highly recognizable (81.4% correct). In the spectral inversion condition, fewer errors were made on the simpler syntactic types than on the more complex types (Figure 1). Most errors involved confusions between structures with the same number of syllables (Tables 2 and 3). In contrast, performance was only around chance (12.4% correct) for the sentences with both spectral and temporal cues reduced. Furthermore, there was no obvious pattern to the errors when both the spectral and temporal properties of the sentences were reduced.

Figure 1: Mean number of sentences of each syntactic type identified correctly in the four conditions (intact, inverted, concatenated, concatenated and inverted).

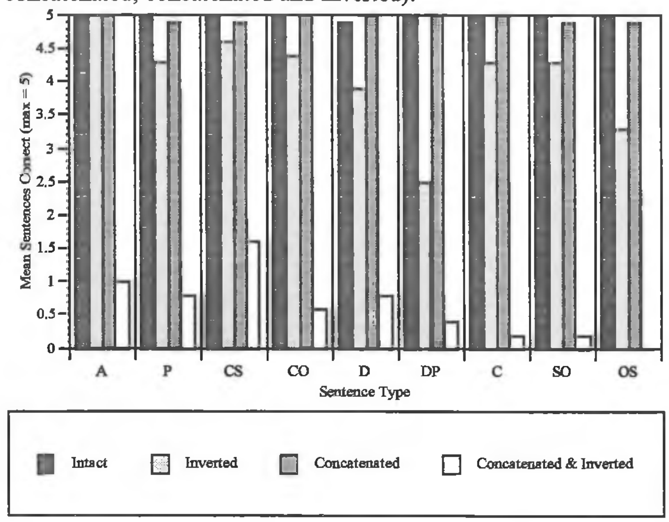

Table 2: Confusion matrix for target-response patterns in the spectral inversion condition by sentence type. Number of responses are shown (max. = 75).

	Target Sentence Type								
Response	A	CS	CO	C	P	SO	D	OS	DP
A	75		1		8				
CS		69	6	1		2		2	
CO		3	66			1		1	2
C				65			3	16	14
P		3	1		64	2	8		5
SO						64	1	1	4
D				3	2	4	59	3	12
OS				6		1		50	1
DP			1		1	1	4	2	37

Table 3: Confusion matrix for target-response patterns according to number of syllables. Sentence type A has 5 syllables; sentence type P has 7 syllables; sentence types CS, CO, and D have 8 syllables; sentence types C, SO, and OS have 9 syllables; sentence type DP has 10 syllables.

	Number of Syllables in Target Sentence				
Number of Syllables in Response	5	7	8	9	10
5	100%	11%	1%		
7		85%	5%	1%	7%
8		3%	90%	8%	19%
9			2%	90%	25%
10		1%	2%	1%	49%

Also of interest is the finding that subjects with low working memory spans were less able to identify spectrally inverted sentences than were those with high working memory spans. Correlations between working memory span and sentence identification scores were highly significant (p<.05) for all syntactic types except OS and A, with the highest correlations being observed for the sentence types C (r = .83) and SO (r = .80) both of which are two-verb structures. Note that it was not possible to observe meaningful correlations with scores for sentence type A since no errors were made on this sentence type and that the majority (64%) of errors on type OS, the third two-verb structure, involved the choice of one other sentence type with the same syllabic pattern (C).

DISCUSSION

The results of this experiment demonstrate that listeners are able to correctly identify most common syntactic structures when spectral cues are severely reduced but prosodic cues are largely preserved or when normal sentence-level prosodic cues are altered but spectral cues are preserved, whereas they are unable to perform better than chance when both types of cues are reduced.

An examination of the pattern of errors in the spectral inversion condition suggests that listeners are guided by syllabicity cues insofar as they usually choose sentence types with the same (or slightly fewer) syllables than the target sentence type. For the two sentence types (OS and DP) that were most poorly recognized, the most frequent response was a more common sentence type with a shallower syntactic tree structure (C). These findings are compatible with the recent arguments presented by Greenberg (1998) regarding he usefulness of syllabic information for speech intelligibility in everyday listening conditions.

The high correlation between working memory span and accuracy in identifying the sentence types, especially two of the three more complex two-verb types with a high number of syllables, is consistent with the idea that subjects are storing information about the number and relative timing of syllables and using this information to select one of the alternative sentence types.

REFERENCES

Benguerel, A.-P. (1998). Stress-timing vs syllable-timing vs. mora-timing: The perception of speech rhythm by native speakers of different languages. Etudes et Travaux, Institute des Langues Vivantes et de Phonétique, Université Libre de Bruxelles.

Blesser, B. (1972). Speech perception under conditions of spectral transformations, I: Phonetic characteristics. Journal of Speech & Hearing Research, 15, 5-41.

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. Cognition, 21, 117-175.

CSRE (4.2) (1995). Computer Speech Research Environment. London, Ontario: AVAAZ Innovations Inc.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behaviour, 19, 450-466.

Dillon, L.M. (1995). The effect of noise and syntactic complexity on listening comprehension. Master of Science Thesis, University of British Columbia.

Greenberg, S. (1998). Auditory processing of speech. Invited tutorial, NATO Advanced Study Institute on Computational Hearing, Il Ciocco, Italy.

Pichora-Fuller, M.K., Lloyd, V., Dillon, L., & Kirson, R. Effects of sentence type and background noise on comprehension by listeners with low- or high-span working memory. Contributed poster, Seventh Cognitive Aging Conference, Atlanta, Georgia, April 1998.