

PROSODIC PROPERTIES OF CONTRASTIVE UTTERANCES IN SPONTANEOUS SPEECH

Shari R. Speer & Kiwako Ito

Ohio State University, USA

speer@ling.ohio-state.edu; ito@ling.ohio-state.edu

ABSTRACT

This study examines phonological and phonetic properties of *contrastive* referential expressions. An interactive holiday tree decoration task provided sequences of ornaments to prompt production of target noun phrases in spontaneous conversation. This paper presents an analysis of a subset of referential expressions that were uttered in contrastive discourse contexts, focusing on the production of L+H* vs. H* accents. Blind ToBI analysis revealed that contrastive adjective was produced most frequently with L+H*, whereas the contrastive noun was produced more frequently with H* than with L+H*. The relative F0 height measures showed more consistent relations to the accent choices than the relative durational measures.

Keywords: contrast, discourse analysis, referential expressions, prosody

1. INTRODUCTION

Prosodic variation signal information such as emotional status, illocutionary force, gestures for turn taking and syntactic and pragmatic structuring of the message. Among the range of prosodic options available for encoding a speaker's intent, the placement of *pitch accent* – a distinctive tonal event that leads to perception of relative prominence – is one of the most studied and debated topics in intonational phonology. This is perhaps due to the complex relationship between the physical realization of pitch accents and their phonological, semantic and pragmatic representations, many of which are extremely difficult to manipulate and observe experimentally during spoken language production.

Since Autosegmental-Metrical theory was applied to capture the intonation patterns of American English [3, 12], the convention to describe the continuous F0 contours with discrete tonal units such as H and L (to indicate the relative height of abstract tonal targets) has widely spread, giving rise to the construction of annotation

systems such as ToBI [1, 2]. The ToBI framework was constructed to serve as a tool to analyze both spontaneous speech and speech produced for experimental purposes in the laboratory. However, identifying continuous, gradient tonal events as phonological categories is never easy. The validity of ToBI as an analytic tool has been frequently debated [5, 16], because even trained annotators often cannot reach consensus as to what accent type should be assigned to a given *accented* word [13]. In particular, the distinction between H* and L+H* (*indicates the association between the tone and the stressed syllable of the word) has been intensively discussed, as these two pitch accent types, each with an abstract high tonal target, arguably express differences in the degree of relative prominence or emphasis, leaving open the possibility that they belong to the same accentual category [8, 11, 15]. This view has been partially supported by a perception experiment with naïve listeners, which reports that detection of the location of accentuation is easy but identifying the degree of prominence is not [17].

Researchers across traditions have repeatedly proposed a distinction between accent types, positing one type that denotes the relative importance of a word and another that marks contrast [6, 7, 10]. Supporting evidence for this distinction comes from eye-tracking studies that have shown listeners' immediate interpretation of a prominent accent as marking a contrastive referent [9]. However, the precise nature of the prosodic cues actually used by speakers when they intend to convey contrast remains unclear. The present study examined the prosodic properties of contrastive referential expressions elicited in an interactive conversational task. Participants gave instructions to a partner on how to decorate holiday trees with a given set of ornaments, where the label for each ornament and the decoration sequence were prompted via photographs displayed on a computer monitor. This setup loosely constrained the range of referential expressions and discourse structures produced by

speakers, facilitating the examination of turn taking and the intent behind each utterance.

Referential expressions were ToBI-annotated by a researcher blind to production conditions and discourse context beyond the individual utterance. In a separate analysis, target utterances produced in contrastive discourse contexts were selected according to their positions in the discourse structure. The target expressions were also submitted to duration and F0 measurements to examine the acoustic characteristics of *contrastive* expressions in spontaneous speech and to test whether particular pitch accent types have reliably distinguishable pitch and durational properties.

2. EXPERIMENT

2.1. Participants

Sixteen undergraduate students at the Ohio State University participated in the tree decoration task for partial fulfillment of a course requirement. The present paper reports data from a subset of six speakers.

2.2. Materials and design

Each experimental session included the decoration of four trees. Four sets of slides were created to prompt the instructions for each tree with 24 target and 8 filler ornaments. On each slide, a photo image of an ornament was projected side by side with an image of the entire tree, on which the label for the ornament and its location were superimposed (Fig. 1)

Figure 1: Example slide prompting a tree decoration instruction for the ornament “blue drum.”



The labels for the ornaments combined a color adjective (target: navy, green, orange, gray, beige, blue, brown, clear; filler: white, purple, gold, silver) with an object noun (target: ball, onion, house, drum, doll, bell, candy, egg; filler: snowman, stocking, star, hat). Target adjectives and nouns appeared three times each, while a

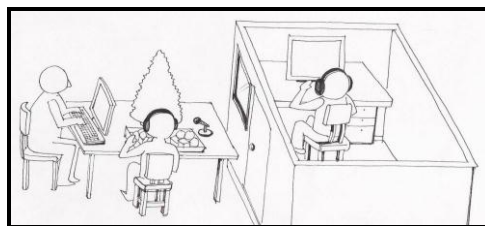
particular combination of an adjective and a noun was never repeated within a tree.

2.3. Procedure

Each participant (the Instructor) was paired with a confederate, the Decorator. The Instructor sat in front of a monitor inside a soundproofed booth while the Decorator sat outside the booth with a set of ornaments and a tree. Both speakers wore a headphone set and used a microphone to communicate with each other. Speakers were not allowed to see each other's faces, but through the window on the booth wall, the Decorator could show the mentioned ornament to the Instructor, who could check the location of the decoration during each trial (Fig. 2). Every time the Decorator finished hanging an ornament and the Instructor confirmed its location on the tree, the experimenter (who sat outside the booth with the Decorator) pressed a key to present the next slide.

Participants' conversations were recorded at 44.1 KHz using Praat [4]. The confederate Decorator used deictic expressions, but never mentioned color adjectives or ornament nouns.

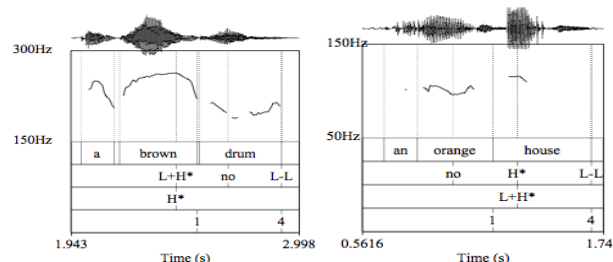
Figure 2: Experimental configuration.



2.4. ToBI annotation

Before the target *contrastive* utterances were selected on the basis of the analysis of discourse structure, utterances mentioning the target ornaments were segmented separately and were annotated according to ToBI conventions. The labeler was blind to the discourse context for each utterance, and the annotations were completed in a pseudo-random order. An example ToBI annotation is given in Fig. 3.

Figure 3: Example ToBI annotation with 'Alt' tier.



Consistent with current annotation conventions [14], the Alt (alternative) tier was used to note instances of pitch accent type that were ambiguous between two categories. Annotations also distinguished ‘deaccented’ (reduced) ‘unaccented’ (words with full vowels but no pitch accent) items.

2.5. Discourse analysis

Each session was carefully transcribed to reflect the patterns of turn taking between the Instructor and the Decorator and the informational status of referential expressions. A referential expression was labeled as *contrastive* when either the adjective or the noun of a target noun phrase was repeated from the Instructor’s most recent adjective-noun mention of an ornament of the same color or object type (see Table 1.)

Table 1: Example text transcription of dialogue.

311.4	INST:	um blue drum
316.16	DEC:	yeah [shows ornament through booth window]
320.9	INST:	‘k
325.0	INST:	uh brown drum
328.6	DEC:	yeah
337.8	DEC:	ok, next?
340.0	INST:	uh silver snowman, and this is a five, has five in this row, yeah
355.0	DEC:	‘k
357.9	INST:	then a clear bell , [Dec shows through booth window]
358.4	DEC:	‘k. [places on tree] so like there?
370.7	INST:	an clear doll , yep
380.6	DEC:	ok

2.6. Duration & F0 analysis

For each *contrastive* noun phrase, the beginning and the ending of each word and the beginning and the ending of its stressed vowel were marked for both the adjective and the noun. The absolute duration of the stressed vowel was divided by the absolute duration of the word containing the vowel. The mean F0 value for the stressed vowel was also divided by the mean F0 value for the noun phrase.

3. RESULTS

ToBI annotations and phonetic measures were made for 190 *adjective-contrast* utterances and 162 *noun-contrast* utterances from six speakers.

3.1. Distribution of pitch accents

The ToBI annotations revealed clear differences in pitch accent distribution between the utterance types (*adjective-* vs. *noun-contrast*). As shown in Table 2, the adjective was *always* accented in the *adjective-contrast* utterances, and it was most frequently produced with L+H*.

Table 2: Pitch accent counts for *adjective-contrast* utterances (n=190).

Pitch Accent	Adjective		Noun	
	Tone-tier	Alt-tier	Tone-tier	Alt-tier
H*	56	10	33	3
!H*	2	0	25	6
H*>	2	0	0	0
H*+!H*	8	2	4	0
L+H*	110	8	1	0
L+!H*	1	0	2	0
L*	7	4	15	6
L*+H	4	0	0	0
Deacc	0	0	83	3
Unacc	0	0	27	0

Unsurprisingly, the most frequent alternative annotations for these adjective accents were also H* and L+H*, accounting for 18 of 24 ambiguous cases. The noun in *adjective-contrast* utterances was most frequently deaccented, and often appeared unaccented. While pitch accents did appear on the noun, L+H* and L+!H* were rare in this location.

Table 3 shows that adjectives in *noun-contrast* utterances most frequently carried H*, but never L+H*. Nouns were most likely to carry H* or !H*, but a substantial number of L+H* accents were also found. Adjectives were sometimes deaccented or unaccented, most often when preceding L+H* on the noun. Overall there were more accent pattern types, and more types of ambiguity shown in the Tone vs. Alt tier annotations for *noun-* than for *adjective-contrast* utterances.

Table 3: Pitch accent counts for *noun-contrast* utterances (n=162).

Pitch Accent	Adjective		Noun	
	Tone-tier	Alt-tier	Tone-tier	Alt-tier
H*	108	10	41	4
!H*	5	0	53	12
H*>	4	1	0	0
H*+!H*	1	3	9	0
L+H*	24	3	30	2
L+!H*	0	0	6	2
L*	7	0	13	7
L*+H	0	0	1	0
Deacc	9	1	6	0
Unacc	4	0	3	0

3.2. Duration & F0 properties

For each utterance type, the relative duration and F0 measures were compared across three subgroups sorted by the type of accent for the contrastive word: (a) unambiguous H* (including

H* > /!H*), (b) ambiguous H* and L+H* with alternative L+H* and H*, and (c) unambiguous L+H*. In *adjective-contrast* utterances (Table 4), the relative F0 range of the adjective's stressed vowel was significantly higher for unambiguous L+H* than for unambiguous H* ($t=2.57, p<.05$), though it did not differ between unambiguous L+H* and ambiguous H*/L+H* ($t=1.31, p>.05$).

Table 4: Relative duration and F0 for *adjective-contrast* NPs grouped by accent type on adjective.

Pitch Accent	Adjective		Noun	
	Duration	F0	Duration	F0
(a) H*	.44	1.02	.54	.92
(b) amb	.49	1.05	.46	.88
(c) L+H*	.43	1.13	.50	.91

In the *noun-contrast* utterances (Table 5), the relative F0 range of the noun's stressed vowel was significantly higher for unambiguous L+H* than for ambiguous H*/L+H* ($t=1.77, p<.05$), and it was also higher for absolute H* than for ambiguous H*/L+H* ($t=1.64, p<.05$). In addition, the relative F0 range of the adjective's stressed vowel was significantly lower before the nouns with unambiguous L+H* than before the nouns with ambiguous H*/L+H* ($t=2.67, p<.05$). The relative duration of the stressed vowels did not show consistent patterns in either utterance type.

Table 5: Relative duration and F0 for *Noun-contrast* NPs grouped by accent type on noun.

Pitch Accent	Adjective		Noun	
	Duration	F0	Duration	F0
(a) H*	.49	1.02	.56	.97
(b) amb	.43	1.06	.60	.88
(c) L+H*	.68	.96	.58	1.02

4. DISCUSSION AND CONCLUSION

Using a carefully designed interactive dyad task, the present study examined the tonal patterns and phonetic properties of contrastive referential expressions in spontaneous speech. When the discourse context prompted contrast for the adjective, the adjective was never deaccented or unaccented and was produced most frequently with L+H*. When contrast was on the noun, speakers produced H* more frequently than L+H*. The relatively infrequent use of L+H* for the noun-contrast utterances may reflect the fact that the target noun phrase was often produced at the end of a prosodic phrase, where the compressed pitch range may have constrained tonal excursion. The phonetic analysis of the noun phrases indicated

that F0 measures distinguished accent types better than the durational measures. Data from additional speakers would confirm whether the ambiguity between L+H* and H* is solely related the word's F0 range, or the F0 range of adjacent word also affects the judgment of tonal prominence.

5. REFERENCES

- [1] Beckman, M., Ayers, G. 1997. *Guidelines for ToBI Labelling, vers. 3.0*. OSU.
- [2] Beckman, M., Hirschberg, J., Shattuck-Hufnagel, S. 2005. The original ToBI system and the evolution of the ToBI framework. In Jun, S.-A. (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford, 9-54.
- [3] Beckman, M., Pierrehumbert, J. 1986. Intonational structure in Japanese and English. *Phonology Yearbook III*. 15-70.
- [4] Boersma, P., Weenink, D. 1992-2010. Praat: doing phonetics by computer. <http://www.praat.org>
- [5] Breen, M., Dilley, L., Gibson, E., Bolivar, M., Kraemer, J. 2006. Advances in prosodic annotation: A test of inter-coder reliability for the RaP and ToBI transcription systems. *19th CUNY Conf NY*.
- [6] Calhoun, S. 2010. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language* 86, 1-42.
- [7] Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In Li, C.N. (ed.), *Subject and Topic*. Academic Press, 25-56.
- [8] Dilley, L., Ladd, D., Schepman, A. 2005. Alignment of L and H in bitonal pitch accents: testing two hypotheses. *J. Phonetics*, 33, 115-119.
- [9] Ito, K., Speer, S.R. 2008. Anticipatory effect of intonation: Eye movements during instructed visual search. *J. Memory and Language* 58, 541-573.
- [10] Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. MIT Press.
- [11] Ladd, R., Morton, R. 1997. The perception of intonational emphasis: continuous or categorical? *J. Phonetics* 25, 313-342.
- [12] Pierrehumbert, J.B. 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT.
- [13] Pitrelli, J., Beckman, M., Hirschberg, J. 1994. Evaluation of Prosodic transcription labeling reliability in the ToBI framework. *Proc. Intern'l Conf Spoken Language Processing Yokohama, Japan*, 2, 123-126.
- [14] Veilleux, N., Shattuck-Hufnagel, S., Brugos, A. 2006. Transcribing prosodic structure of spoken utterances with ToBI. MIT. <http://anita.simmons.edu/~tobi/index.html>
- [15] Watson, D., Tanenhaus, M., Gunlogson, C., 2008. Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science* 32, 1232-1244.
- [16] Wightman, C.W. 2002. ToBI or not ToBI, *Proc. 1st International Conf. Speech Prosody Aix en Provence* 25-29.
- [17] Yoonsock, M., Cole, J. 2008. Naive listeners' prominence and boundary perception, *Proc. Speech Prosody Campinas, Brazil*.