PITCH RANGE VARIATION IN ENGLISH TONAL CONTRASTS: CONTINUOUS OR CATEGORICAL?

Laura C. Dilley

Departments of Communication Disorders and Psychology, Bowling Green State University
dilley@bgsu.edu

ABSTRACT

The importance of pitch range variation for intonation theories is well-known, but whether pitch range variation gives rise to distinctive linguistic categories in English is unclear. To test this possibility, three intonation continua were constructed for use in an imitation experiment; all had endpoints with distinct tonal representations under autosegmental-metrical (AM) theory [1]. Responses to all three stimulus sets showed continuous variation in pitch range. The results suggest that pitch range is a phonetic dimension which is gradient in English.

Keywords: intonation, AM theory, pitch range, English, continuous, categorical.

1. INTRODUCTION

Intonation contours are typically assumed to have a linguistically specified shape (i.e., pattern of rises and falls) and independently variable pitch range [2]. For example, differences in fundamental frequency (F0) contour shapes for statements vs. questions are usually assumed to correspond to phonologically distinctive categories. However, differences in pitch range or "vertical scale" have been treated as linguistically distinctive in some cases, and as paralinguistic in others. (See e.g. treatment of "low-rise" vs. "high-rise" nuclear tones in [3] vs. [4].) A growing body of research suggests that pitch range differences may sometimes serve as the phonetic basis of linguistic contrasts [5, 6]. This paper presents an experiment which tests theoretical assumptions regarding whether differences in pitch range serve as the phonetic basis of linguistic distinctions in English.

Within the Autosegmental-Metrical (AM) theory of intonation [1], there is inconsistency in whether differences in vertical scale are contrastive (see discussion in [7]). On the one hand, differences of vertical scale are often assumed to be paralinguistic and gradient [8]. On the other hand, differences in vertical scale are sometimes assumed to give rise to distinct categories, e.g., downstepping vs. non-downstepping accents [1]. Crucially, three pairs of contrasting tonal patterns in AM theory are distinguished by pitch range characteristics when F0 shape is held constant across the syllable sequence. First, pitch range distinguishes H* and L+H* accents. Both accents show a rise across phrase-initial unstressed syllables, but the F0 level(s) of unstressed syllable(s) are low in the pitch range for L+H* [9]. Second, pitch range distinguishes some instances of H* and L*+H. The two accents may have similar shapes: a H* accent may be realized with an F0 peak on a post-stress syllable, just as for L*+H [9]. In this case, the two accent types are distinguished by the fact that the F0 level of a pitch accented syllable is low in the pitch range for L*+H but not for H* [9]. Third, pitch range distinguishes F0 contours with an initial high boundary tone, %H, from contours with no initial high boundary tone or an initial low boundary tone, %L [9]. Unstressed, phrase-initial syllables which have a higher F0 than a following accented syllable are said to have an initial high boundary tone when the unstressed syllables are very high in the speaker's pitch range, but to have no such tone when the unstressed syllables are only moderately high or else low in the speaker's pitch range.

Three intonation continua spanning these AM category pairs were constructed for use in an imitation task. This task is considered to provide the best test of phonological contrastiveness for intonational categories [10]. Previous research has documented a number of cases where intonational continua are reproduced in a discrete manner [e.g., 11, 12]. If intonational continua varying in pitch range give rise to categorical effects in production, it will suggest that pitch range is a phonetic dimension corresponding to categorical distinctions in English, as is generally assumed under AM theory. It will also provide support for AM theory's assumptions about linguistic vs. paralinguistic pitch range effects. Alternatively, if these continua do not give rise to categorical effects in production, then either pitch range is not a dimension which gives rise to categorical distinctions, or else the endpoints of the continua may correspond to a single phonological category.

2. METHOD

2.1. Stimuli

Short phrases were selected containing two-syllable sequences with specific stress patterns in order to test AM theory categories. First, the phrase *some oregano* begins with the weak-weak-strong (WWS) sequence /səmɚɛg/.[1] Next, the phrase *some oranges* begins with the WSW sequence /səmɔrən/. Finally, the utterance *linguistics* begins with the WS sequence /lɪŋgwɪs/. For each phrase, the initial syllables are comprised mostly of voiced, sonorant segments. These phrases were recorded in a sound-attenuated room using a DAT recorder and a high-quality microphone at 22.1 kHz and transferred to a PC.

To create each stimulus series, F0 contours were stylized as a sequence of straight-line segments and then synthesized using a pitch-synchronous overlap-and-add algorithm in Praat [13]. The *oregano* series was created by shifting the F0 level across the initial WW sequence *some or-* in *some oregano* in 12 equal logarithmic steps; F0 onsets ranged from 125 Hz to 324 Hz. The first and last stimuli in the *oregano* series corresponded to H* and L+H* accents on *reg-*, respectively [10]. Next, the *oranges* series was created by shifting the F0 level across the initial WS sequence *some or-* in *some oranges* in 12 equal logarithmic steps; F0 onsets ranged from 127 Hz to 329 Hz. The first and last stimuli in the *oranges* series corresponded to H* and L*+H on *or-*, respectively [10]. Finally, the *linguistics* series was created by shifting the initial F0 level across the initial WS sequence *linguis-* in *linguistics* in 15 equal logarithmic steps; F0 onsets ranged from 126 Hz to 425 Hz. The first and last stimuli in this series corresponded to contours with %H vs. %L on *ling*, respectively [10].[2] Across the three series, the ratio of F0 values at any given time for successive stimuli was approximately 1.091 (or 1.5 semitones; one semitone equals $2^{1/12}$ = 1.059…), which is well above the difference limen for pitch [14].

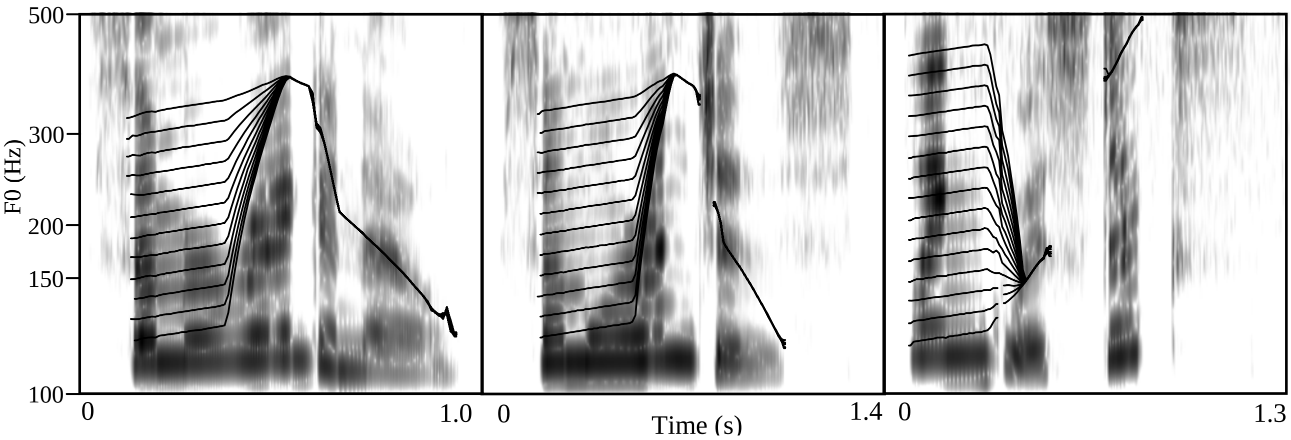
2.2. Participants

Participants were 17 students and staff at colleges in the Boston area (5 males, 12 females), who were self-reported native American English speakers with normal hearing. All were paid a nominal sum. Participants had no known training in phonetics and had a range of musical experience.

2.3. Procedure

Stimuli were recorded onto CD for auditory presentation; stimuli were blocked by series. The three stimulus blocks were repeated three times, for a total of nine blocks, with the order of blocks fixed across participants. Each stimulus block was preceded by a set of practice trials from the upcoming block, and the order of trials within a block was randomized. Stimuli were presented over high-quality headphones at comfortable volume in a sound-treated room, while the text of each phrase was displayed on a computer screen. Participants were told to imitate each phrase as closely as possible in a comfortable pitch range. The imitations were digitized directly to hard disk (16 kHz sampling rate) using custom software

Figure 1: Stimuli used in the imitation experiment. The phrases used in stimuli were *some oregano* (left), *some oranges* (middle) and *linguistics* (right).



(MARSHA v.2.0) written by Mark Tiede. The experiment lasted approximately 35 minutes.

2.4. Analysis

Prior to obtaining F0 measurements, segmental landmarks were identified using spectrogram and waveform displays in Praat. First, the boundaries between /m/ and /ɚ/ in *Some oregano* and between /m/ and /ɔr/ in *Some oranges* were labeled as the locations of amplitude increase across frequencies. Next, the boundaries between /ɚ/ and /ɛ/ in *oregano* and between /ɔr/ and /ən/ in *oranges* were taken as the location of F3 frequency increase, if present, or else the point of amplitude increase in F2 and higher formants. The onset and offset of /ŋ/ in *linguistics* were taken as the locations of amplitude decrease and increase, respectively, across frequencies. Finally, the onset and offset of /wɪ/ in *linguistics* were taken as the locations of amplitude increase across frequencies and of the start of high-frequency energy for /s/, respectively.

Estimates of pitch range were then obtained separately for each stimulus imitation. In particular, two F0 values, T_1 and T_2, were measured for each imitation of a stimulus, where T_1 and T_2 were estimates of the F0 values associated with expected or possible tonal targets under AM theory.[3] For the *oregano* series, T_1 was the average F0 across /ɚ/ in <u>or</u>-, while T_2 was the peak F0 on /ɛ/ in (*or*)<u>*eg*</u>-. For the *oranges* series, T_1 was the average F0 across /ɔr/ in <u>or</u>*(an)*-, while T_2 was the peak F0 on /ən/ in *(or)*<u>*an*</u>-. Finally, for the *linguistics* series, T_1 was the average F0 on /ŋ/ in *li*<u>*ng*</u>-, while T_2 was the average F0 on /wɪ/ in <u>*guis*</u>-. Finally, pitch range estimates were calculated as an *interval metric* using the equation in (1), consistent with the logarithmic (or ratio) scale that was used in constructing stimuli.[4]

(1) $$\log(\text{interval}) = \log\left(\frac{T_2}{T_1}\right)$$

To check consistency across participants' responses, a two-tailed, pairwise bivariate correlation analysis (Pearson's product-moment) was carried out on pairs of subjects based on estimates of average interval size to each stimulus. Subjects who were not significantly correlated at $p < .05$ with half or more of the other subjects were judged to be poor imitators and were discarded from the analysis for that series. This resulted in discarding one participant from the *oranges* series and three from the *oregano* series.

3. RESULTS

Figure 2 shows log produced intervals in participants' imitations (open circles) plotted against mean log stimulus intervals for each series. Participants' imitations are well-described by straight lines; linear regressions to mean produced intervals are shown as solid lines. R^2 values for best-fit lines are high, ranging from 0.972 to 0.991 ($p < .001$ for all). Crucially, no evidence of categorical effects in production is observed for any of the three stimulus series. Finally, produced intervals in participants' imitations show displacement from the line $y = x$ (dashed line), suggesting an overall linear transformation and compression of the pitch range relative to that of the stimuli.

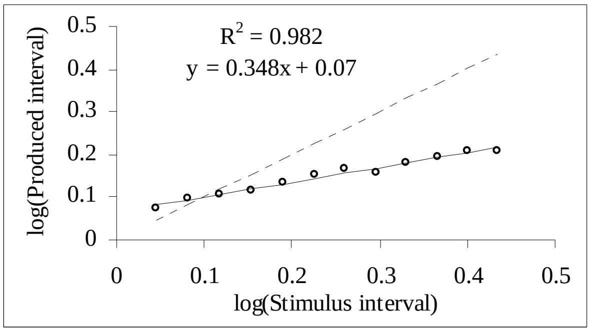
4. DISCUSSION

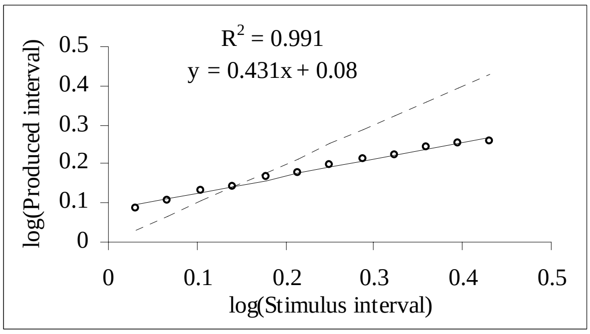
In this experiment participants imitated stimulus continua varying in vertical scale (i.e., pitch range), where the endpoints of continua corresponded to distinctive AM phonological representations. If pitch range were a phonetic dimension conveying the distinction between contrastive tonal pairs, speakers should have produced categorical values for pitch range in their imitations of stimuli. However, no evidence of categorical production of pitch range was observed. Rather, mean responses within each of the stimulus continua were well fit by straight lines (with R^2 greater than 0.97 in all cases). The fact that the data are offset from $y = x$ suggests pitch range was compressed in participants' imitations.

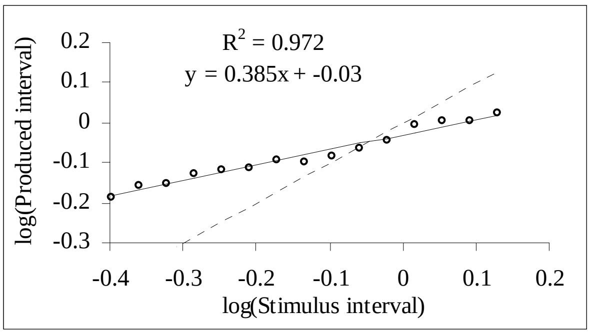
There are two possible theoretical interpretations of these findings. One is that some or all of the pitch range continua were perceived as two categories for participants, who simply did not produce categorical responses to these continua. If so, participants' continuous responses to pitch range continua contrast with typical responses to F0 peak alignment continua, which have previously been shown to give rise to discrete responses in imitation tasks [11, 12]. The possibility that pitch range is categorically perceived, but continuously reproduced, seems unlikely, however. A second possibility is that the endpoints of pitch range continua in this experiment corresponded to graded variations within a single linguistic category. This is a reasonable possibility, since little empirical research has investigated the AM categories proposed in [1] for English. Moreover, it has been argued based on previous production data that one of the tonal pairs studied here, L+H* vs. H*, represent extremes on a continuous dimension consisting of a single category [16]. The present data support such an interpretation.

In summary, the present experiment demonstrates that when pitch range is varied along a continuum in English with distinctive AM tonal representations as endpoints, speakers produce continuous, rather than categorical, responses.

Figure 2: Comparison of log interval sizes in stimuli (abscissa) with log interval sizes in imitated versions of stimuli (ordinate). Results for *oregano*, *oranges*, and *linguistics* series are shown in the top, middle, and bottom graphs, respectively. See text.







5. REFERENCES

[1] Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. Ph.D. dissertation, MIT, Cambridge, MA.

[2] Ladd, D. R., Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *J. Phon.*, 25, 313-342.

[3] O'Connor, R. J., Arnold, G. F. (1973). *Intonation of Colloquial English*. Bristol, U.K.: Longman Group, Ltd.

[4] Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

[5] Hirschberg, J., Ward, G. (1992). The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *J. Phon.*, 20, 241-251.

[6] Vanrell, M. (2005). A scaling contrast in Majorcan Catalan interrogatives. *Proc. Speech Prosody*, Dresden.

[7] Dilley, L., Brown, M. Effects of pitch range variation on F0 extrema in an imitation task. *J. Phon.*, in press.

[8] Liberman, M., Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In: Aronoff, M., Oerhle, R. (eds.), *Language Sound Structure*. Cambridge, MA: MIT Press, 157-233.

[9] Beckman, M., Ayers Elam, G. (1997). Guidelines for ToBI labeling, v. 3.0: Ohio State University. http:// ling.ohio-state.edu/~tobi/ame_tobi/ visited 25-Feb-07

[10] Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

[11] Pierrehumbert, J., Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, 46, 181-196.

[12] Redi, L. (2003). Categorical effects in production of pitch contours in English. *Proc. 15th ICPhS*, Barcelona, 2921-2924.

[13] Boersma, P., Weenink, D. (2002). Praat, a system for doing phonetics by computer. http://www.praat.org, visited 25-Feb-07

[14] Flanagan, J., Saslow, M. (1958). Pitch discrimination for synthetic vowels. J. Acoust. Soc. Am., 30, 435-442.

[15] Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. *Proc. 15th ICPhS*, Barcelona, 771-774.

[16] Ladd, D. R., Schepman, A. (2003). “Sagging transitions” between high accent peaks in English: experimental evidence. *J. Phon.*, 31, 81-112.

---

1 Note that in General American English, the word *oregano* has main stress on the second syllable.

2 Whether %L is a category in English has been debated (cf. [1, 9]). Category endpoints of the *linguistics* series might also be interpreted as H+!H* and L+H* under AM theory. Regardless, the endpoints correspond to distinct phonological representations in AM theory.

3 If glottalization interrupted regions of interest, the longest modal or diplophonic portion was used for the F0 estimate. In the case of diplophonia, F0 estimates were multiplied by a factor of 2.

4 The logarithmic scale used for pitch range estimates is consistent with experimental data [15]. However, other choices of metric are possible; space constraints preclude discussion of these alternatives.