Original Paper

Phonetica 2010;67:63–81
DOI: 10.1159/000319379

Received: May 7, 2010
Accepted: June 14, 2010

Pitch Range Variation in English Tonal Contrasts: Continuous or Categorical?

Laura C. Dilley

Department of Communicative Sciences and Disorders, Departments of Psychology and Linguistics, Michigan State University, East Lansing, Mich., USA

Abstract

The importance of pitch range variation for intonational meaning and theory is well known; however, whether pitch range is a phonetic dimension which is treated categorically in English remains unclear. To test this possibility, three intonation continua varying in pitch range were constructed which had endpoints with contrastive representations under autosegmental-metrical (AM) theory: H* vs. L+H*, H* with 'peak delay' vs. L*+H, and %H L* vs. L*. The prediction derived from AM theory was that the reproduction of continuous pitch range variation should show a discrete pattern reflecting a change in the phonological representation of tonal sequences and in the number of tonal targets across each continuum. Participants' reproductions of each stimulus set showed continuous variation in pitch range, suggesting that pitch range is a gradient phonetic dimension in English conveying semantic contrast, similar to the formant space for vowels. Moreover, the gradience observed in productions across all parts of the pitch range suggests that contours within each series had the same number of tonal targets. The results support a version of AM theory in which rises and falls are usually comprised of two tonal targets, with strictly monotonic f_0 interpolation between them.

Introduction

Intonation contours are typically assumed to have a linguistically specified shape, i.e., pattern of rises and falls, and independently variable pitch range [Pierrehumbert, 1980; Ladd and Morton, 1997; Ladd, 2008]. For example, differences in fundamental frequency (f_0) contour shapes for (falling) statements vs. (rising) questions are usually assumed to correspond to phonologically distinctive categories, while differences in pitch range or 'vertical scale' have been treated as contrastive in some cases, and as paralinguistic in others [Crystal, 1969; O'Connor and Arnold, 1973; Pierrehumbert, 1980]. This paper investigates pitch range as a basis of linguistic contrasts in English intonation, focusing on proposals put forward within the well-known autosegmental-metrical (AM) theory of intonation [Pierrehumbert, 1980; Liberman

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2010 S. Karger AG, Basel
0031–8388/10/0672–0063
$26.00/0
Accessible online at:
www.karger.com/pho

Dr. Laura C. Dilley
Department of Communicative Sciences and Disorders
Departments of Psychology and Linguistics
Michigan State University, 6 Oyer
East Lansing, MI 48824–1220 (USA)
Tel. +1 517 884 2255, Fax +1 517 353 3176
E-Mail ldilley@msu.edu

and Pierrehumbert, 1984; Beckman and Pierrehumbert, 1986; Pierrehumbert and Beckman, 1988].

In recent years, considerable support has come to light for certain aspects of AM theory. A critical assumption of this theory, which was originally proposed for English, is that the phonological primitives giving rise to intonation contours are discrete tones associated with particular positions in metrical structure which are connected by interpolation functions [Pierrehumbert, 1980]. The assumption that the crucial intonational units are discrete tones or levels is mirrored in previous theories of English intonation [Pike, 1945; Liberman, 1975]. A contrasting theoretical view is that of contours approaches, which assume that intonation contours consist of rising and falling elements, as exemplified by work in the IPO approach [t Hart et al., 1990] and the British school [Halliday, 1967; Crystal, 1969]. Moreover, hybrid theories, such as the proposals of Xu [1998, 1999, 2005] and Xu and Wang [2001], assume that both discrete tones, as well as contours, correspond to intonational primitives.

A significant body of phonetic evidence now shows temporal consistency in the alignment of certain discrete f_0 points, namely maxima (or peaks) and minima (or valleys), with respect to segmental landmarks, even under changes in speech rate [Arvaniti et al., 1998; Xu, 1998; Ladd et al., 1999, 2000; Xu, 2001; Ladd and Schepman, 2003; Dilley et al., 2005]. Moreover, differences in f_0 peak and valley alignment are perceptually salient and often cue semantic distinctions [Purcell, 1976; Bruce, 1977; Kohler, 1987; House, 1990; D'Imperio, 2000; Niebuhr and Kohler, 2004; Niebuhr, 2007a]. These facts are readily accommodated under theories which specify discrete tones or targets as primitives, including AM theory and hybrid theories, but they are not easily reconciled with rise/fall theories [Ladd et al., 1999; Xu and Wang, 2001]. Much of this f_0 alignment evidence has been interpreted as direct support for AM theory [see e.g., Arvaniti et al., 1998; Ladd et al., 1999; Arvaniti et al., 2006].

While consistency in f_0 alignment has provided support for the AM assumption that tones are discrete elements, other aspects of the theory, including its treatment of pitch range, have not been as thoroughly investigated. Indeed, there have been conflicting claims in the AM framework regarding the circumstances under which pitch range may be contrastive [see a review and discussion in Dilley and Brown, 2007]. On the one hand, differences of vertical scale have sometimes been treated as paralinguistic and gradient [Liberman and Pierrehumbert, 1984]. On the other hand, differences in vertical scale have sometimes been treated as arising from distinct categories, e.g., downstepping vs. non-downstepping accents [Pierrehumbert, 1980; but see Ladd, 1983, 1990, 1993]. Controversy about the phonological representation of pitch range characteristics in English is not new; for example, O'Connor and Arnold [1973] proposed that 'low-rise' vs. 'high-rise' nuclear tones are distinctive, while Crystal [1969] treated these differences as arising from noncontrastive variation.

Nevertheless, several tonal patterns have been consistently assumed within the AM framework to be critically distinguished by pitch range. The purpose of the present research is to investigate the phonological representation of pitch range, specifically, to determine whether pitch range gives rise to phonetic evidence of discreteness or 'categorical effects' associated with distinct phonological categories. If so, such categorical effects could present evidence in support of the specific phonological proposals distinctions that have been proposed within AM theory to capture pitch range-based meaning differences.

As background, AM theory assumes two types of tonal categories: pitch accents, which convey prominence, and phrasal boundaries, which separate phrasal constituents. Pitch accents can be single-toned or bitonal; an asterisk or 'star' marks tones which are associated with a stressed syllable. In bitonal accents, an unstarred tone leads or trails a starred tone; for example, in L+H*, the high accentual pitch, H*, is phonologically associated with a stressed syllable, while the low f_0 turning point for unstarred L+ precedes H* on an unstressed syllable or syllables. Tonal patterns of interest in this paper will be described using phonological analyses proposed within the original framework of Pierrehumbert [1980] and Beckman and Pierrehumbert [1986], which will be referred to as 'standard AM theory' or the 'standard AM framework'. This distinguishes the original phonological descriptions of these contours from more recent proposals also based on the AM framework, since the former descriptions are in common use and have been widely accepted [Ladd and Schepman, 2003; Dilley, 2005; Ladd, 2008].

Within standard AM theory, three pairs of tonal patterns are assumed to be critically distinguished in English on the basis of pitch range characteristics, focusing here on phrase-initial position. First, H* describes a high target in the upper part of the speaker's pitch range which typically involves a small rise across a prestress syllable or syllables from the middle or upper part of the pitch range to an f_0 peak [Beckman and Ayers Elam, 1997]. The peak for H* is assumed to occur either on the stressed syllable itself or during a post-stress syllable; the latter represents a 'delayed peak' version of H* [Silverman and Pierrehumbert, 1990; Beckman and Ayers Elam, 1997]. Moreover, the L+H* pitch accent describes a rise across a prestress syllable or syllables to an f_0 peak; f_0 values on any prestress syllables are critically assumed to be low in the speaker's pitch range [Beckman and Ayers Elam, 1997]. Thus, in phrase-initial position H* and L+H* are assumed to differ principally in terms of the positions in the pitch range of any prestress syllables, with prestress syllables for H* being higher than those for L+H*. (See Gussenhoven [2004] for an alternative analysis of distinctive English rising contours which also predicts that representations differing in pitch range will be categorically distinct phonologically.) Note, however, that the two accent types also are assumed to differ phonologically in terms of the *number* of tonal targets, with two for L+H* and one for H* [Beckman and Ayers Elam, 1997]; in other words, prestress syllables are assumed to carry a tone (L+) when they are relatively low in the speaker's range, but to lack a tone when they are higher in the range. Ladd and Schepman [2003] previously found evidence against the assumption of different numbers of tonal targets for H* vs. L+H* in phrase-medial position.

As another example of pitch range dependency in tonal categories, consider that the L*+H accent describes an accented syllable with an f_0 very low in the speaker's pitch range which is followed by a relatively sharp rise to a peak in the upper part of the speaker's pitch range [Beckman and Ayers Elam, 1997]. The pitch peak for this accent usually occurs on a post-stress syllable [Pierrehumbert, 1980; Pierrehumbert and Steele, 1989]. There are thus similarities in the temporal alignment pattern of L*+H and that of the 'delayed peak' variant of H*, with L*+H and 'delayed peak' variant of H* differing phonetically in their assumed pitch range characteristics. Note, however, that these two accent types are also assumed to differ phonologically in terms of the *number* of tonal targets, with two for L*+H and one for the 'delayed peak' variant of H*. Thus, a post-stress syllable is assumed to carry a tone (+H) when the preceding stressed syllable is relatively low in the speaker's range (for L*+H), but that syllable is assumed to lack a tone when the preceding syllable is higher in the range (since the H* tone is phonologically associated with the stressed syllable, not the following unstressed syllable). The relative complexity of the standard AM phonological analysis of pitch range is further reflected in the fact that the phonological analysis of the post-stress syllable depends in this case on the pitch range properties of the preceding stressed syllable (i.e., whether or not it is a L*), rather than the post-stress syllable itself.

Third and finally, the initial high boundary tone, %H, describes the f_0 on initial unstressed syllable(s) in a relatively high to very high part of the pitch range at the beginnings of intonation phrases, while contours with no initial high boundary tone begin with a f_0 in the low to mid- (i.e., 'default') part of the speaker's pitch range [Beckman and Ayers Elam, 1997]. (Note that while an initial low boundary tone, %L, was originally proposed for English [Pierrehumbert, 1980], more recent revisions to the theory associated with the Tones and Break Indices (ToBI) system of intonation transcription [Silverman et al., 1992; Beckman and Ayers Elam, 1997] have eliminated %L as a category.) It has been suggested that in English the use of an initial high pitch on unstressed syllables (%H) indicates 'liveliness', 'vivacity', or 'emphasis' [O'Connor and Arnold, 1973, p. 36; see also Grabe et al., 1998, for Dutch]. Phrases starting with %H are assumed to differ phonetically from those lacking an initial boundary tone on the basis of pitch range characteristics, but once again, it can be seen that the number of tonal targets is assumed to vary across the pitch range. In particular, phrase-initial unstressed syllables are assumed to carry a tone (%H) when they are high in the range, but they are assumed to carry no tone when they are lower in the range. Overall, it can be seen across these cases that a principal means by which pitch range is captured in standard AM theory is in terms of a contrast in the number of tonal targets, depending on the position in the pitch range.

The critical issue in the present paper concerns how pitch range variation operates at a phonetic level in order to give rise to meaningful differences. Of particular interest is whether discreteness or 'categorical effects' might be observed in participants' responses to a pitch range continuum. Such findings could provide support for the phonological analyses proposed within standard AM theory, and particularly for its assumption of different numbers of tonal targets in distinct parts of the pitch range.

That pitch range differences can give rise to meaning contrasts, and therefore that such differences are worthy of capture in phonological theories of intonation, is supported by results from a great many empirical studies. For example, differences in semantic focus are associated with differences in pitch range in a wide variety of languages [e.g., Cooper et al., 1985; Xu and Xu, 2005; Féry and Kügler, 2008; Genzel and Kügler, 2010]. Furthermore, listeners have been shown to interpret pitch range differences in terms of distinct patterns of semantic reference or pragmatic intent [Hadding-Koch and Studdert-Kennedy, 1964; Studdert-Kennedy and Hadding-Koch, 1973; Nash and Mulac, 1980; Bartels and Kingston, 1994; Niebuhr, 2007b; Watson et al., 2008]. Evidence for pitch range as the phonetic basis of meaning contrasts comes from semantic judgment tasks, production studies, and eyetracking studies [Ward and Hirschberg, 1985; Hirschberg and Ward, 1992; Bartels and Kingston, 1994; Gussenhoven and Rietveld, 2000; Calhoun, 2006; Weber et al., 2006; Féry and Kügler, 2008; Watson et al., 2008].

There are a variety of empirical methods that have been used to investigate *phonological* representations of phonetic variation, as opposed to the semantic distinctions they convey. Note that semantic judgment and related psycholinguistic tasks are generally considered insufficient for assessing the nature of phonological contrasts; instead, phonetic evidence, in the form of discreteness (or discontinuity) in production or perception of gradient stimuli, has been considered superior for this purpose [Gussenhoven, 1999, 2004; Ladd, 2008]. Phonetic evidence supporting distinctiveness of phonological representations can come from either production or perception tasks. In production tasks, phonetic evidence of phonological contrast is taken to correspond to discreteness or discontinuity in imitation of continuous phonetic variation, e.g. f_0 maximum timing or F1-F2 values [Pierrehumbert and Steele, 1989; Viechnicki, 2002; Niebuhr, 2003; Redi, 2003; Niebuhr and Kohler, 2004; Dilley, 2005; Braun et al., 2006; Dilley, 2007]. In perception tasks, phonetic evidence of phonological contrast is typically taken to correspond to categorical perception, which corresponds to a maximum in discrimination and an s-shaped labeling function, given continuous phonetic variation [Liberman et al., 1957; Repp, 1984], but other kinds of perceptual tasks have also been taken to provide evidence relevant to evaluating phonologic representations [e.g., Gussenhoven and Rietveld, 2000].

Previous studies which have sought phonetic evidence relating to phonological contrasts in intonation based on pitch range have yielded conflicting results. For example, Ladd and Morton [1997] used a continuum of intonation contours varying in pitch range to investigate the distinction between 'normal' and 'emphatic' high accents. Using categorical perception techniques, they failed to find evidence of a discrimination maximum, leading them to conclude that the contrast may be 'categorically interpreted', but not categorically perceived. On the other hand, Cummins et al. [2006] investigated contrast in English stimuli ranging from a high final rise to a low final fall. Their results showed an s-shaped labeling curve with crossover point aligned at the discrimination peak; however, the pattern did not hold for all subjects, who were each tested extensively. Moreover, Remijsen and van Heuven [1999] found evidence of an s-shaped labeling curve with a crossover point aligned at the discrimination peak in investigating high final rises to low falls in Dutch, but they also found a puzzling second discrimination peak which remained unexplained. Finally, several recent studies have found somewhat conflicting evidence of categorical effects based on pitch range, for meaningful linguistic contrasts in Catalan using a variety of tasks [Vanrell Bosch, 2006; Borras-Comes et al., 2010; Crespo-Sendra et al., 2010; Vanrell et al., 2010].

While these studies therefore appear to demonstrate some amount of phonetic evidence for the phonological contrastiveness of pitch range, Dilley and Brown [2007] have suggested that an extraneous factor other than pitch range might explain some of the findings. Using an imitation task, Dilley and Brown [2007] showed that when the pitch range of two syllables was varied along a continuum, both categorical and continuous effects were observed. On the one hand, categorical effects *in f_0 alignment* were observed precisely when the pitch range manipulation caused the pattern of relative pitch across syllables to differ, i.e., when one syllable switched from being higher than to lower than another in pitch. This suggests a close perception-production link between differences in relative pitch and differences in f_0 alignment; a growing body of evidence supports the latter as a categorical dimension of intonation [e.g., Bruce, 1977; Pierrehumbert and Steele, 1989; House, 1990; Niebuhr and Kohler, 2004]. On the other hand, continuous change in pitch range was observed, whether or not a relative pitch change was introduced. The findings suggest that listeners may perceptually interpret a relative pitch change across two syllables as contrastive, since such a change closely resembles a change in f_0 alignment. Overall, the Dilley and Brown [2007] study suggests that differences of relative pitch and f_0 alignment must be controlled, in order to investigate whether pitch range differences per se can trigger phonetic discontinuities reflecting underlying phonologically distinct categories.

Several studies investigating pitch range have also involved changes in relative pitch and/or f_0 alignment, including Remijsen and van Heuven [1999], Cummins et al. [2006], and Gussenhoven and Rietveld [2000]. (Note, however, that Cummins et al. [2006] explicitly assumed that relative height, rather than pitch range per se, was responsible for evidence of phonetic contrast.) Thus, evidence of phonological contrast obtained in these cases might actually be explained by a change in relative pitch, rather than pitch range variation per se. The study by Ladd and Morton [1997] controlled for both relative pitch and alignment, yet this study failed to find phonetic evidence of contrast based on pitch range differences alone.

One possible explanation for Ladd and Morton's [1997] failure to find categorical effects in response to their pitch range continuum is that categorical perception is simply a weak and/or unstable task for assessing phonological representations in intonation. This is the view put forward by Gussenhoven [2004], based on a review of studies using categorical perception tasks; instead, Gussenhoven [2004] proposed that imitation tasks provide a better means of uncovering phonetic evidence of distinct underlying phonological categories. If pitch range were to give rise to evidence of discreteness in an imitation task, it will suggest that the manner by which pitch range conveys distinctions is similar to that of stop consonants, for which phonetic continua, e.g., in F2 or voice onset time, readily give rise to perception of discrete categories [e.g., Liberman et al., 1957; Lisker and Abramson, 1970]. On the other hand, a different explanation for the lack of categorical effects observed in Ladd and Morton's [1997] study is that pitch range per se is not a phonetic dimension that produces evidence of discreteness or discontinuity in perception or production, even though pitch range differences are clearly associated with semantic distinctions. This interpretation is supported by initial evidence from the study by Dilley and Brown [2007] that pitch range was reproduced continuously, even when relative pitch changed. Indeed, it could be that the manner by which pitch range conveys meaningful distinctions is similar to that of vowels, for which phonetic continua, e.g., in F1-F3, give rise to perceptual distinctions that are perceived relatively continuously [e.g., Fry et al., 1962; Stevens et al., 1969] with some steps being heard as more prototypical [Kuhl, 1991]. If the phonological categories associated with pitch range variation are conveyed via continuous variation in a manner similar to how vowel categories are conveyed, then finding evidence supporting or refuting particular phonological proposals for underlying contrasts will not be a simple matter of choosing the right methodological paradigm and controlling for confounding factors. In sum, these considerations suggest that if phonetic evidence of contrasts based on pitch range is to be found, it is best discovered by (1) using stimuli in which the relative height and f_0 alignment of stimuli is carefully controlled, and (2) using an imitation task.

The present experiment therefore investigated several proposed standard AM categories using an optimal task (i.e., imitation) and enhanced experimental control over pitch range and f_0 alignment. The first goal of the present research was to determine whether there is support for the particular phonological representations proposed in standard AM theory to underlie pitch range variation, in particular, to determine whether there is support for the theory's treatment of pitch range as involving different numbers of tonal targets at the endpoints of pitch range continua. Given the widespread assumption that phrase-initial syllables with no tonal specification are produced in a default region of the speaker's pitch range [Beckman and Ayers Elam, 1997], there is a clear prediction within the standard AM framework that imitation of a pitch range continuum across such syllables should give rise to consistency in placement of those syllables in a set mid or default part of the range and the concomitant appearance of discontinuity in reproducing the entire continuum. The second, related goal was to determine whether pitch range is a phonetic dimension which gives rise to categories associated with discreteness, akin to consonants in segmental phonology, or else more subtly graded and continuous categories, akin to vowels. We focused on three cases where pitch range is assumed to contrast two different tonal categories assumed to be distinguished in terms of the number of tonal targets, controlling for relative height and f_0 alignment. For the first continuum, the endpoints ranged from single-toned H* with a small rise across prestress syllables to a peak on a stressed syllable, to bitonal L+H* with a steeper rise across prestress syllables to a peak on a stressed syllable. For the second continuum, the endpoints ranged from single-toned H* with 'peak delay' involving a small rise from the high part of the speaker's pitch range peaking on a high-pitched post-stress syllable at one end, to bitonal L*+H involving a steep rise from the lowest part of the speaker's pitch range peaking on a high-pitched post-stress syllable at the other end. Finally, for the third continuum, the endpoints ranged from the two-tone sequence %H L* with initial high pitch falling steeply to low at one end, to single-toned L* with an initial low gradually rising to a slightly higher pitch at the other end. For this last series, the relative pitch of the initial two syllables was additionally manipulated specifically in order to investigate whether an explicit change of relative height yielded a categorical or continuous effect on pitch range variation, following up on the initial results of Dilley and Brown [2007].

If phonetic evidence of discreteness is found using an imitation task, it will support the distinct phonological representations proposed under AM theory. In particular, it will strengthen the AM assumption that the number of tonal targets differs depending on pitch range, i.e., that contours with no tone are produced in a default region of the speaker's pitch range, while comparable contours with L or H tones are produced in low vs. high parts of the speaker's pitch range, respectively [Beckman and Ayers Elam, 1997]. Such different numbers of tonal targets should be evidenced as discreteness reflecting consistent use of a default or 'mid' pitch range on tonally unmarked syllables [Beckman and Ayers Elam, 1997, p. 21] and/or better vocal control over parts of the pitch range where there are more tonal targets than over those parts of the range where there are fewer tonal targets. Alternatively if pitch range is treated as continuous in the phonetic system, it will be consistent with the notion that the same number of tonal targets are present across the entire pitch range when controlling for phonetic differences of relative height and f_0 alignment, and may point to the need for revisions in the phonological categories assumed to underlie meaning contrasts. Phonetic evidence of continuous change in an imitation task would suggest that meaningful pitch range variation derives from a phonetic dimension not unlike that of vowel contrasts in segmental phonology, which are perceived and produced continuously [Fry et al., 1962; Stevens et al., 1969; Vinegrad, 1972; Viechnicki, 2002].

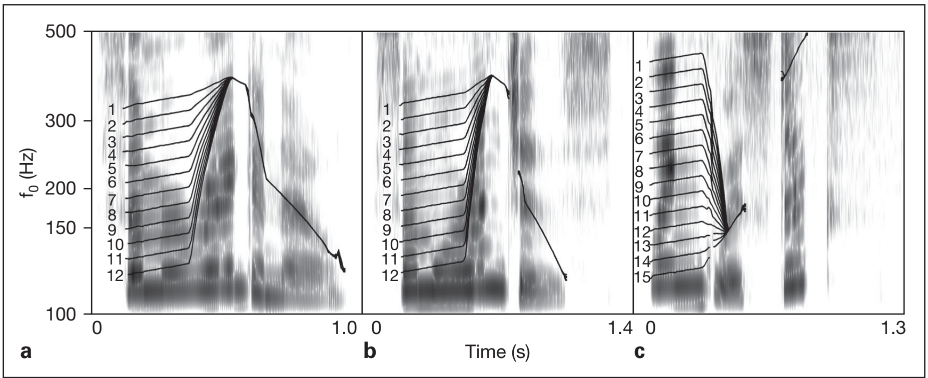

Fig. 1. Stimuli used in the imitation experiment. The first phrase was *some oregano* (**a**); the high and low endpoints of this continuum corresponded to H* and L+H*, respectively. The next phrase was *some oranges* (**b**); the high and low endpoints of this continuum corresponded to H* with delayed peak and L*+H, respectively. The final phrase was *linguistics* (**c**); the high and low endpoints of this continuum corresponded to %H L* and L*, respectively.

Method

Stimuli

Three 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 unstressed-unstressed-stressed (UUS) sequence /səmɚrɛg/. (Note that in General American English, the word *oregano* has main stress on the second syllable.) Next, the phrase *some oranges* begins with the USU sequence /səmɔrən/. Finally, the utterance *linguistics* begins with the US sequence /lɪŋgwɪs/. For each phrase, the initial syllables are comprised almost entirely of voiced, sonorant segments. These phrases provided critical phonetic environments for creating pitch range continua with particular AM tonal categories as endpoints. Phrases were produced by the author, who is a native speaker of General American English from the Central Midwest US. Recordings were made in a sound-attenuated room using a DAT recorder and a high-quality microphone at 22.1 kHz and transferred to a PC. A single version of each phrase was selected for subsequent resynthesis. The pitch range characteristics of critical portions of each selected recording were in the mid range between maximum and minimum f_0 values associated with endpoints of continua as described below, making base stimuli well suited to interpretations in terms of tonal categories associated with either endpoint in the respective stimulus series. Pitch range characteristics of original recordings were as follows. First, the sequence /səmɚ/ in the recording of *some oregano* had a mean f_0 of 210 Hz; there was also a peak f_0 on /rɛg/ of 289 Hz. Moreover, the sequence /səmɔ/ in the recording of *some oranges* had a mean f_0 of 217 Hz; the original production had a 'late peak' on /rən/ of 339 Hz. Finally, the syllable /lɪŋ/ in the recording of *linguistics* had a mean f_0 of 298 Hz; the f_0 valley on /gwɪs/ was 180 Hz.

To create the stimulus series, f_0 contours for each phrase were stylized as a sequence of straight-line segments and resynthesized using a pitch-synchronous overlap-and-add algorithm in Praat [Boersma and Weenink, 2002] (fig. 1). The *oregano* series was created by shifting the f_0 level across the initial UU sequence *some or-* in *some oregano* in 12 equal logarithmic steps; f_0 onsets ranged from 125 Hz, which was low in the speaker's pitch range and thus representative of phrase-initial L+H*, and 324 Hz, which was high in the speaker's pitch range and thus representative of phrase-initial H*. The first and last stimuli in the *oregano* series thus corresponded to H* and L+H* accents on *reg-*, respectively [Beckman and Ayers Elam, 1997]. Next, the *oranges* series was created by shifting the f_0 level across the initial US sequence *some or-* in *some oranges* in 12 equal logarithmic steps; f_0

onsets ranged from 127 Hz, which was low in the speaker's pitch range and thus representative of phrase-initial L*+H, to 329 Hz, which was high in the speaker's pitch range and thus representative of phrase-initial H*. The first and last stimuli in the *oranges* series thus corresponded to H* and L*+H on *or-*, respectively [Beckman and Ayers Elam, 1997]. Finally, the *linguistics* series was created by shifting the initial f_0 level across the initial US sequence *linguis-* in *linguistics* in 15 equal logarithmic steps; f_0 onsets ranged from 126 Hz, which was low in the speaker's pitch range and thus representative of the case of no initial boundary tone, to 425 Hz, which was high in the speaker's pitch range and thus representative of initial %H. The first and last stimuli in this series thus corresponded to contours with %H vs. no tone on *ling-*, respectively [Beckman and Ayers Elam, 1997]. Across the three series, the ratio of initial f_0 values for successive stimuli within a continuum was approximately 1.091 (or 1.5 semitones; one semitone equals $2^{1/12} = 1.059$ …); this is well above the difference limen for pitch [Flanagan and Saslow, 1958].

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.

Procedure

Stimuli were recorded onto CD for auditory presentation and were blocked by stimulus series. The three stimulus blocks were presented in the order *oranges-oregano-linguistics*; this block ordering was repeated 3 times, for a total of 9 blocks. Each stimulus block was preceded by a set of practice trials from the upcoming block, and the order of trials within each block was randomized. Participants were told to imitate each phrase as closely as possible in a comfortable pitch range. Stimuli were presented over high-quality headphones at comfortable volume in a sound-attenuated room; the text of each phrase was simultaneously displayed on a computer screen. The imitations were digitized directly to PC hard disk using a 16-kHz sampling rate via custom software (MARSHA v.2.0 by Mark Tiede). The experiment lasted approximately 35 min.

Analysis

Prior to obtaining f_0 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 /ɚr/ 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 f_0 values, V_1 and V_2, were measured for each imitation of a stimulus, where V_1 and V_2 were estimates of the f_0 values associated with expected or possible tonal targets under AM theory. For the *oregano* series, V_1 was the average f_0 across /ɚr/ in *or-*, while V_2 was the peak f_0 on /ɛ/ in *(or)eg-*. For the *oranges* series, V_1 was the average f_0 across /ɔr/ in *or(an)-*, while V_2 was the peak f_0 on /ən/ in *(or)an-*. Finally, for the *linguistics* series, V_1 was the average f_0 on /ŋ/ in *ling-*, while V_2 was the average f_0 on /wɪ/ in *guis-*. If glottalization interrupted regions of interest, the longest modal or diplophonic portion was used for the f_0 estimate. In the case of diplophonia (defined here as an alternating pattern of *ABAB*… in the amplitude or shape of pitch periods in the waveform) f_0 estimates were multiplied by a factor of 2 to compensate for pitch halving. Finally, pitch range estimates were calculated as an interval (ratio) metric using the equation in (1), consistent with the logarithmic scale that was used in constructing stimuli.

$$\log(\text{interval}) = \log\left(\frac{V_1}{V_2}\right) \qquad (1)$$

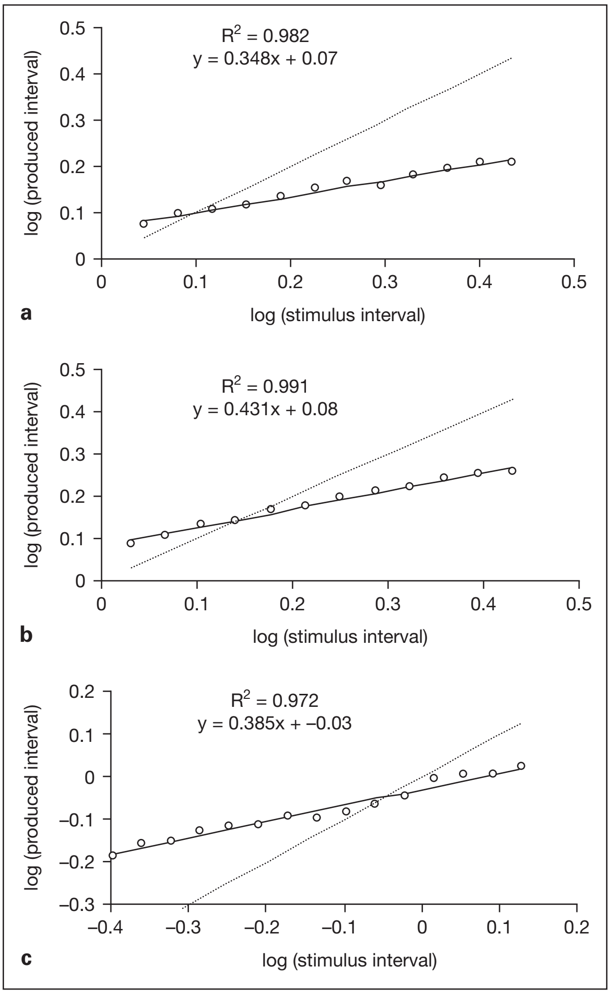

Fig. 2. Comparison of log interval sizes in stimuli (abscissa) with log interval sizes in imitated versions of stimuli (ordinate). Results for *oregano* (**a**), *oranges* (**b**), and *linguistics* (**c**) series are shown. The dashed line corresponds to the equation y = x (see text).

To check consistency across participants' responses, a two-tailed, pairwise bivariate Pearson's product-moment correlation analysis was carried out on pairs of subjects based on estimates of average interval size to each stimulus. Subjects who were not significantly correlated at $p < 0.05$ 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 1 participant from the *oranges* series and 3 from the *oregano* series.

Results

Figure 2 shows log-produced intervals in participants' imitations plotted as open circles against mean log stimulus intervals for each series. It can be seen that participants' imitations are well described by straight lines for each of the three stimulus series. Linear regressions to mean produced intervals are shown as solid lines.

Pearson's R^2 values for best-fit lines were high, ranging from 0.972 to 0.991 ($p < 0.001$ for all). Crucially, no evidence of categorical effects in production is observed for any of the three stimulus series. It can also be observed that produced intervals in participants' imitations show displacement from the line y = x (indicated as a dashed line). This suggests an overall linear transformation and compression of the pitch range relative to that of the stimuli, which was likely due to the wide range of pitch variation in the stimuli.

Finally, the *linguistics* series was examined to determine whether the change in relative height, which occurred at around stimulus 12, induced categorical effects in imitation of pitch range. To test this, the average Euclidean distance for each participant between log values of produced intervals for stimuli *near* the change in relative height (stimuli 9–14) and the best-fit line was compared with the average Euclidean distance between log values of produced intervals for stimuli *far* from the change in relative height (stimuli 1–8 and 15) and the best-fit line. Any categorical effects in pitch range resulting from perseveration of relative pitch should correspond to less accurate pitch range reproduction (i.e., greater Euclidean distance) near the change in relative height than far from this change. A two-way, paired-samples t test revealed no significant difference in average Euclidean distance between log values of produced intervals for stimuli near the change in relative height (M = 0.056) versus far from the change in relative height (M = 0.062), $t(16) = 0.757$, $p = 0.46$.

General Discussion

The purpose of the present research was to investigate the phonological representation of pitch range, specifically, to determine whether a pitch range continuum gives rise to phonetic evidence of discreteness or discontinuity consistent with distinct phonological categories proposed within standard AM theory: H* vs. L+H*, H* with 'delayed peak' vs. L*+H, and %H L* vs. L*. Because an imitation task is considered a preferred method of investigating phonological categories in intonation [Gussenhoven, 2004; Ladd, 2008], the current study presented an advantage over previous studies which have used only perceptual methods in attempting to detect categorical effects in pitch range variation. Moreover, the present study controlled for relative pitch change and f_0 alignment within each stimulus series, thereby eliminating two potentially confounding variables compared with previous studies.

If pitch range variation were to have given rise to evidence of discreteness in an imitation task, then it would have provided support for the phonological distinctions proposed as part of standard AM theory, and in particular, for its assumption of different numbers of tonal targets, depending on the pitch range of the speaker. Such discreteness is predicted under the common assumption that phrase-initial syllables lacking a tonal specification are consistently produced in a 'default' or middle part of the pitch range [Beckman and Ayers Elam, 1997]. In contrast to this prediction, participants produced continuous pitch range variation in response to each stimulus series. No evidence of categorical production of pitch range was observed; instead, straight lines provided excellent fits to produced responses for each stimulus series. These findings provide evidence for a phonological analysis in which phrase-initial pitch accents contrasting in pitch range have the same number of tonal targets, and are also interpreted as support for the notion that pitch range differences convey semantic contrasts in English in a manner similar to vowels (i.e., continuously) [Fry et al., 1962; Stevens et al., 1969]. These points are considered in more detail below.

First, note that the finding that speakers produced continuous variation in pitch range in response to stimuli cannot simply be explained in terms of listeners repeatedly being exposed to very similar stimuli and thus operating in a 'phonetic' rather than a 'phonological' mode in which they automatically accurately encoded the acoustic continuum. This is because a number of previous experiments have used imitation tasks in which participants listened to very similar or identical stimuli many times but still produced categorical effects in imitation, at least when the manipulations have involved a continuum of f_0 alignment [Pierrehumbert and Steele, 1989; Redi, 2003; Dilley, 2005; Gili Fivela, 2008]. Such categorical responses would not have been obtained if the imitation paradigm itself were not capable of eliciting evidence of discreteness in reproduction of an intonation continuum, or if repetition of very similar stimuli automatically led to perception in a 'phonetic' mode in which accuracy in encoding acoustic details is obligatory and unavoidable.

It was determined that the English pitch range contrasts investigated here appear to operate like vowels, for which phonetic continua give rise to continuously differentiable contrasts, rather than like consonants, for which phonetic continua tend to give rise to discreteness and discontinuity [Liberman et al., 1957; Lisker and Abramson, 1970]. It is reasonable to assume that each pitch range continuum traverses up to two, but not more than three, semantic distinctions or categories (e.g., broad vs. narrow focus), each of which presumably has a psychologically distinct reality for listeners that is coded in terms of a unique phonological representation. In this vein, it is furthermore reasonable to assume that subjects recoded stimuli into semantic categories during the task, where these are indexed via gradient, within-category phonetic variation. Imitation of semantic or phonological categories conveyed by gradient pitch range differences may therefore be seen as analogous to successfully imitating within-category variation for vowels stemming from dialectic or speech style differences [Hagiwara, 1997; Kuhl et al., 1997; Burnham and Kitamura, 2002; Clopper and Pisoni, 2004].

Note that the stimulus continua were constructed by closely following both descriptions of canonical categories in the ToBI training materials of Beckman and Ayers Elam [1997], as well as standard assumptions about phonetic characteristics which are likely to occur for each accent category. Thus the endpoints of continua can be considered representative of the AM categories in question, so that the continua were a fair test of whether gradual acoustic changes separating these accent categories give rise to categorical effects distinguished by pitch range. Of particular interest in this regard is the designation of the high endpoint of the *oranges* series as the category H* with 'delayed peak'. While the 'delayed peak' position does not represent the canonical peak alignment position for this accent, it is nevertheless a well-recognized variant of the H* accent type. The selection of the 'delayed peak' shape for this endpoint was justified by the facts that H* frequently surfaces American English with an f_0 peak on a post-stress syllable, and that H* is the label typically assigned during ToBI annotation to contours with a high post-stress peak [Shattuck-Hufnagel et al., 2004]. Moreover, treating H* with 'delayed peak' as the high endpoint of the *oranges* series is justified by the fact that Dilley [2005] demonstrated evidence across several experiments that the f_0 contour corresponding to H* with 'delayed peak' is perceived and produced as a distinctive intonation category (and therefore one which is legitimate to use as a continuum endpoint in the present experiment); this was shown in experiments involving both discrimination and imitation tasks using a continuum of f_0 peak alignments spanning canonical H* (with no peak delay) to H* with 'delayed peak'. In addition, recall that it was necessary to hold constant f_0 alignment and relative height across syllables within a pitch range continuum in order to avoid the confounding variable of f_0 alignment on obtaining categorical effects. Finally, note that both peak alignment positions traditionally associated with the standard AM category of H* (i.e., on the stress vs. in post-stress or 'delayed peak' position) were tested in distinct stimulus series *(oranges, oregano)*, so both variants were tested for their potential to give rise to evidence of discreteness. The findings of continuous reproduction of pitch range characteristics therefore suggest that when the pitch range is made lower for the standard AM H* accent category, the result does not give rise to categorical effects in an imitation task, regardless of the peak alignment position for the H*.

One possible qualification to these conclusions is that there could be other distinctions that are based on pitch range that were not tested here which might give rise to discreteness. The present experiment sampled a subset of the contrasts that are assumed within the AM framework and was limited to contrasts associated with phrase-initial position. However, there are several other AM contrasts that are assumed to be distinguished based on pitch range (e.g., the phrase-final distinctions between H-H% vs. L-H% and H-L% vs. L-L%). The existence of these other presumed contrasts that are assumed to depend on pitch range means that these data do not conclusively indicate that pitch range does not anywhere give rise to discreteness or discontinuity associated with distinct categories in English. However, we see no reason to believe that pitch range will behave any differently in these other cases, at least when alignment and relative f_0 level differences are controlled for.

Other details of the findings are additionally worthy of note. First, reproduced pitch range values were offset from the line y = x. This implies the presence of pitch range compression in participants' imitations. This compression likely resulted from the fact that the overall pitch range in each stimulus series was very wide, so that participants chose to adapt the pitch range in the stimuli to that of their own voices. It is clear, however, that in spite of this transformation and compression, no evidence of categorical effects in production was observed. Moreover, note that a change in relative height for the *linguistics* series did not induce categorical effects in pitch range variation. This is consistent with earlier findings from Dilley and Brown [2007], who presented initial results that when pitch range was manipulated, a change in relative pitch resulted in categorical effects on f_0 alignment but continuous effects on pitch range.

These findings have important implications for the assumption in AM theory that the number of tonal targets in intonation contours depends on the position within the speaker's pitch range. For example, the initial unstressed syllables in the *oregano* series carry no tonal target when they rise from the mid- to high part of the pitch range (i.e., for H*), but they carry an L tone when they rise from the low part of the pitch range (i.e., for L+H*). Similarly, contours at one end of each continuum for the *oranges* and *linguistics* series are assumed to lack a tonal target on the unstressed initial syllable(s), but contours at the other end of each continuum are claimed to carry a tonal target on the same syllable(s).

A standard assumption in intonational phonology and phonetics is that a change in phonological composition of tones should result in discrete behavior in response to a continuum for the relevant phonetic dimension [Pierrehumbert and Steele, 1989; Ladd, 2008]. In particular, phrase-initial syllables lacking a tone are assumed to be produced in a consistent 'default' or middle part of the pitch range, while phrase-initial syllables with an L or H tone are assumed to be produced in a characteristically low or high part of the range, respectively [Beckman and Ayers Elam, 1997]. Given these assumptions, the phonological change in number of tones should have resulted in a discrete appearance of the pitch range data across each series. However, this discrete pattern was not observed; instead, the results showed continuous variation across all stimuli, reflecting a similar degree of accuracy and control of pitch range across the entire series. This suggests that the number of tonal targets serving as loci of pitch range control were the same for all parts of the pitch range. In other words, regardless of position in the pitch range, the number of tonal targets was constant across all stimuli within a continuum.

The simplest explanation for this production consistency is that each phrase-initial rise or fall in the stimuli arose from *two tonal targets*, one each at the start and end of the rise or fall. A similar conclusion was reached by Ladd and Schepman [2003] for English rising contours, i.e., H* vs. L+H* in standard AM theory, based on examination of a different type of phonetic variation, namely, the alignment of phrase-medial low f_0 valleys, which in addition to pitch range has been claimed to distinguish these two accent types. For H*, a phrase-medial preceding low f_0 valley is claimed to correspond to a nonphonological 'sagging transition' [Pierrehumbert, 1980], predicting no segmental anchoring under changes in phrase length. In contrast, for L+H* a phrase-medial low f_0 valley is claimed to correspond to the unstarred L tone [Pierrehumbert, 1980], predicting consistent segmental anchoring under changes in phrase length. Using read sentences, Ladd and Schepman [2003] observed consistent alignment of phrase-medial low f_0 valleys with respect to segments under changes in phrase length, suggesting that all phrase-medial rises ending in a high, stressed accent involved the same phonological representation. They interpreted these results as suggesting that all rising accent contours peaking on a high, stressed syllable in English have the same number of tonal targets, i.e., two. Moreover, they proposed treating both standard AM H* and L+H* as instances of one rising accent category, which they called (L+H)* to capture the idea that all such rising contours had two tonal targets.

The present results support and extend Ladd and Schepman's [2003] work in several ways. First, rather than f_0 valley alignment (and associated segmental anchoring), the current study examined a different type of phonetic variation – pitch range – which is also assumed to distinguish the English rising accents H* and L+H* in standard AM theory. Using a different type of evidence and a controlled experimental task, the present study reached the same conclusions as Ladd and Schepman [2003] regarding consistency of the number of tones for these rising accents. Second, Ladd and Schepman's [2003] study examined rising patterns only for phrase-medial position; the present study extends this earlier result by showing consistency in the number of tonal targets, even for phrase-initial position. Finally, a wider range of tonal contrasts was investigated in the present study than was explored in the Ladd and Schepman [2003] study, including rising accents peaking on a stressed syllable, rising accents peaking on a post-stress syllable, and falling accents, again suggesting that even for these additional patterns, the number of tonal targets does not depend on pitch range, contrary to standard AM assumptions.

The present results can be accounted for in phonological terms by extending the proposal put forward by Ladd and Schepman [2003] with slight modifications. Ladd and Schepman [2003] proposed that all rising accents peaking on a stressed syllable should be described as bitonal (L+H)*, while the notation L+H* should be reserved for a rising accent with an expanded pitch range, following Pierrehumbert [1980]. This proposal allows all rising accents peaking on a stressed syllable in English to be treated in a similar way, one in which the endpoints of such rises are marked by tones. However, Ladd and Schepman's [2003] original proposal does not afford a means of uniquely referring rising accents with a *compressed* pitch range. The distinction between narrow vs. expanded pitch range is important, since it has been tied to semantic contrasts involving e.g., focus and/or speaker attitude [Cooper et al., 1985; Hirschberg and Ward, 1992; Calhoun, 2004; Liu and Xu, 2005] and was critical to the original distinction proposed between H* and L+H* by Pierrehumbert [1980]. It is therefore proposed that parentheses be used around a sequence of two tonal elements to denote *compressed* pitch range, while a lack of parentheses be used to denote *expanded* pitch range. Moreover, by enclosing the entire tonal sequence in brackets for rises with compressed pitch range, e.g., (L+H*), it is clear that only the H element, not both elements, has a 'star', and the notation can more readily be extended to other sequences of tones.

Given these minor changes to Ladd and Schepman's [2003] proposal, phonological descriptions can be generated within the general AM framework according to which all contours within each stimulus series have the same number of tones. In particular, for stimuli in the *linguistics* series, all phrase-initial falling contours can be treated as a sequence of a high boundary tone followed by a low tone. In particular, it is proposed that %H L* be used to represent a large initial fall with expanded pitch range, while (%H L*) be used to represent a shallow initial fall with compressed pitch range. Moreover, stimuli from this series with a small rise from a low pitch to a somewhat higher but (still) low pitch can be described as (%L L*). Finally, for the *oranges* series, all phrase-initial rising contours peaking on a post-stress syllable can be described as L*+H when the pitch range is expanded, following Pierrehumbert's [1980] original proposal for this accent, but as (L*+H) when the pitch range is compressed. To be clear, all three pairs of similar distinctions – L+H* vs. (L+H*), %H L* vs. (%H L*), and L*+H vs. (L*+H) – are assumed here to signal meaningful contrasts which are separated by a phonetic boundary which is gradient and continuous, much like vowels.

In addition to providing an account of rising and falling tonal patterns in which the number of tonal targets is consistent, the present proposal also overcomes a shortcoming in the original theory regarding the mapping from phonological tone sequences to phonetic f_0 values in phrase-initial position. Strictly speaking, the general AM theoretical framework allows only two mechanisms by which f_0 values can be assigned. First, it is assumed that when a given position in an utterance corresponds to an underlying phonological tone, an f_0 value for that tone is generated from that tone's prominence or pitch value [Pierrehumbert, 1980; Pierrehumbert and Beckman, 1988; Dilley and Brown, 2007]. Alternatively, when a given position in an utterance does not correspond to an underlying phonological tone, an f_0 value may be assigned via a phonetic interpolation function. Given these mechanisms, a longstanding theoretical gap concerns how f_0 values of utterance-initial syllables lacking a phonological tone are determined, since phonetic interpolation does not occur across intonation phrases or gaps of silence delimiting utterances [Pierrehumbert, 1980]. While it has been claimed that f_0 values of phrase-initial unstressed syllables are filled in with a 'default' value [Beckman and Ayers Elam, 1997], no mechanism for generating such values has been proposed in formal phonetic models. The present proposal simplifies the phonology-phonetics interface such that phrase-initial unstressed syllables in English rising and falling patterns are assumed to always have a tone (see Dilley [2005] and Dilley and Brown [2005] for similar proposals). As a result, the two basic phonetic mechanisms already proposed for specifying f_0 values are sufficient to deal with the phrase-initial case, and there is no need to posit a special phonetic mechanism responsible for 'filling in' default f_0 values on phrase-initial unstressed syllables.

Finally, some comment is required on the fact that one endpoint for two of three stimulus continua in the present experiment are treated in standard AM theory as the same accent, H*. In particular, in the *oregano* series, the peak occurred on the stressed syllable for H*, while in the *oranges* series, the peak occurred on the post-stress syllable, corresponding to a 'peak-delayed' variant of H*. While standard AM theory treats these two contours as belonging to the same phonological category, evidence against this interpretation was recently presented by Dilley [2005], as discussed earlier. Dilley [2005] proposed that the findings could be accounted for by assuming that standard AM H* encompasses two distinct phonological categories, depending on f_0 alignment characteristics. These categories could be described as arising from different patterns of phonological association of a H tone with a stressed syllable vs. a post-stress syllable. In Dilley's [2005] proposal, a small rise to a peak on the stressed syllable was treated as H*, while a small rise to a peak on a post-stress syllable was treated as H*+H. Based on the present proposals, contours with a small rise and a peak on the stressed syllable are described as (L+H*), while contours with a small rise and a peak on the post-stress syllable are described as (L*+H). Given the present data suggesting that the rise is initiated starting with a tonal target, it seems more appropriate to describe the peak-on-stress case as (L+H*), in contrast to Dilley's [2005] proposal that this contour be H*. However, the choice of whether to describe a small rise to a peak on a post-stress syllable as (H*+H) or as (L*+H) hinges on resolving conflicting treatment of cues to tone identity in standard AM theory. Tonal identity in the AM system reflects different kinds of phonetic properties – sometimes the height of a tone relative to another, and sometimes the tone's position in the pitch range (see Dilley and Brown [2007] for discussion). Deciding whether (H*+H) or (L*+H) provides a better treatment of rises peaking on a post-stress syllable would require defining a clearer and more consistent relationship between phonological 'H' and 'L' and phonetic f_0 properties, an issue which is beyond the scope of the present paper.

The recently proposed RaP (Rhythm and Pitch) intonation transcription system [Dilley and Brown, 2005; Breen et al., 2006; Dilley et al., 2006], which builds on AM theory, provides a means of resolving this impasse. In the RaP system, every tone is named according to its relative pitch level in relation to other tones. Thus L or H indicates a tone which is lower or higher, respectively, than the tone to its left, with the exception of utterance-initial tones, for which L or H indicates a tone that is lower or higher, respectively, than a tone to its right. Unstarred tones may both lead and trail the same starred tone, building on the work of Grice [1995] as well as Dilley [2005]. Contours with a rise to a post-stress syllable from the *oranges* series would thus be described as: L+ H* +H L% in the RaP system. In other words, the RaP system resolves the impasse about the description of contours with a rise from a stressed syllable to a post-stress peak by suggesting a basic two-tone label *sequence* of H* +H (indicating two independent tones with temporally and metrically coordinated alignment properties; see Dilley [2005]). This is the equivalent of bitonal H*+H in the present paper. (Note that the RaP system captures pitch range contrasts with the use of the '!' diacritic, rather than parentheses.)

In summary, the present experiment demonstrated that when pitch range is varied along a continuum in English with distinctive AM tonal representations as endpoints, speakers produced continuous rather than categorical responses. Such continuous responses occurred even though an optimal phonetic task was used and confounding factors were eliminated. This experiment therefore provides the strongest evidence yet that pitch range variation, like vowel contrasts, conveys meaningful information in a gradient fashion, at least for English.

References

Arvaniti, A.; Ladd, D.R.; Mennen, I.: Stability of tonal alignment: the case of Greek prenuclear accents. J. Phonet. *26:* 3–25 (1998).

Arvaniti, A.; Ladd, D.R.; Mennen, I.: Tonal association and tonal alignment: Evidence from Greek polar questions and contrastive statements. Lang. Speech *49:* 421–450 (2006).

Bartels, C.; Kingston, J.: Salient pitch cues in the perception of contrastive focus; in Bosch, van der Sandt, Focus and natural language processing. Proc. J. Semantics Conf. on Focus. IBM Working Papers, TR-80.94-006 (1994).

Beckman, M.; Ayers Elam, G.: Guidelines for ToBI labeling, version 3 (Ohio State University, 1997).

Beckman, M.; Pierrehumbert, J.: Intonational structure in Japanese and English. Phonol. Yb. *3:* 255–309 (1986).

Boersma, P.; Weenink, D.: Praat, a system for doing phonetics by computer. Software and manual available online at http://www.praat.org 2002.

Borras-Comes, J.; Vanrell, M.d.M.; Prieto, P.: The role of pitch range in establishing intonational contrasts in Catalan. Proc. Speech Prosody 2010, Chicago 2010.

Braun, B.; Kochanski, G.; Grabe, E.; Rosner, B.: Evidence for attractors in English intonation. J. acoust. Soc. Am. *119:* 4006–4015 (2006).

Breen, M.; Dilley, L.C.; Gibson, E.; Bolivar, M.: Advances in prosodic annotation: a test of inter-coder reliability for the RaP (Rhythm and Pitch) and ToBI (Tones and Break Indices) transcription systems. 19th Annu. CUNY Conf. on Human Sentence Processing, New York 2006.

Bruce, G.: Swedish word accents in sentence perspective (Gleerups, Lund 1977).

Burnham, D.; Kitamura, C.: What's new, pussycat? On talking to babies and animals. Science *296:* 1435 (2002).

Calhoun, S.: Phonetic dimensions of intonational categories: the case of L+H* and H*. Proc. Speech Prosody 2004, Nara 2004.

Calhoun, S.: Information structure and the prosodic structure of English: a probabilistic relationship (Department of Linguistics, University of Edinburgh, 2006).

Clopper, C.; Pisoni, D.B.: Some acoustic cues for the perceptual categorization of American English regional dialects. J. Phonet. *32:* 111–140 (2004).

Cooper, W.E.; Eady, S.J.; Mueller, P.R.: Acoustical aspects of contrastive stress in question-answer contexts. J. acoust. Soc. Am. *77:* 2142–2156 (1985).

Crespo-Sendra, V.; Vanrell, M.d.M.; Prieto, P.: Information-seeking questions and incredulity questions: gradient or categorical contrast? Proc. Speech Prosody 2010, Chicago 2010.

Crystal, D.: Prosodic systems and intonation in English (Cambridge University Press, Cambridge 1969).

Cummins, F.; Doherty, C.P.; Dilley, L.C.: Phrase-final pitch discrimination in English. Proc. Speech Prosody, Dresden 2006.

D'Imperio, M.: The role of perception in defining tonal targets and their alignment (The Ohio State University, 2000).

Dilley, L.C.: The phonetics and phonology of tonal systems (MIT, Cambridge 2005).

Dilley, L.C.: Pitch range variation in English tonal contrasts: continuous or categorical? Proc. Int. Congr. Phonet. Sci., Saarbrücken 2007.

Dilley, L.C.; Breen, M.; Bolivar, M.; Kraemer, J.; Gibson, E.: A comparison of inter-transcriber reliability for two systems of prosodic annotation: RaP (Rhythm and Pitch) and ToBI (Tones and Break Indices). Proc. Interspeech 2006, Pittsburgh 2006.

Dilley, L.C.; Brown, M.: The RaP (Rhythm and Pitch) Labeling System, version 1.0. Available at http://tedlab.mit.edu/tedlab_website/Links.html 2005.

Dilley, L.C.; Brown, M.: Effects of pitch range variation on F0 extrema in an imitation task. J. Phonet. *35:* 523–551 (2007).

Dilley, L.C.; Ladd, D.R.; Schepman, A.: Alignment of L and H in bitonal pitch accents: testing two hypotheses. J. Phonet. *33:* 115–119 (2005).

Féry C.; Kügler F.: Pitch accent scaling on given, new and focused constituents in German. J. Phonet. *36:* 680–703 (2008).

Flanagan, J.L.; Saslow, M.: Pitch discrimination for synthetic vowels. J. acoust. Soc. Am. *30:* 435–442 (1958).

Fry, D.B.; Abramson, A.S.; Eimas, P.; Liberman, A.M.: The identification and discrimination of synthetic vowels. Lang. Speech *5:* 171–189 (1962).

Genzel, S.; Kügler, F.: The prosodic expression of contrast in Hindi. Proc. Speech Prosody 2010, Chicago 2010.

Gili Fivela, B.: Intonation in production and perception: the case of Pisa Italian (Edizioni dell'Orso, Alessandria 2008).
Grabe, E.; Gussenhoven, C.; Haan, J.; Marsi, E.; Post, B.: Preaccentual pitch and speaker attitude in Dutch. Lang. Speech *41:* 63–85 (1998).
Grice, M.: Leading tones and downstep in English. Phonology *12:* 183–233 (1995).
Gussenhoven, C.: Discreteness and gradience in intonational contrasts. Lang. Speech *42:* 283–305 (1999).
Gussenhoven, C.: The phonology of tone and intonation (Cambridge University Press, Cambridge 2004).
Gussenhoven, C.; Rietveld, T.: The behavior of H* and L* under variations in pitch range in Dutch rising contours. Lang. Speech *43:* 183–203 (2000).
Hadding-Koch, K.; Studdert-Kennedy, M.: An experimental study of some intonation contours. Phonetica *11:* 175–185 (1964).
Hagiwara, R.: Dialect variation and formant frequency: the American English vowels revisited. J. acoust. Soc. Am. *102:* 655–658 (1997).
Halliday, M.A.K.: Intonation and grammar in British English (Mouton, Paris 1967).
Hirschberg, J.; Ward, G.: The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. J. Phonet. *20:* 241–251 (1992).
House, D.: Tonal perception in speech (Lund University Press, Lund 1990).
Kohler, K.J.: Categorical pitch perception; in Viks, Proc. 11th Int. Congr. Phonet. Sci., Tallinn 1987, pp. 331–333 (1987).
Kuhl, P.: Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. Perception Psychophysics *50:* 93–107 (1991).
Kuhl, P.K.; Andruski, J.E.; Chistovich, I.A.; Chistovich, L.A.; Kozhevnikova, E.V.; Ryskina, V.L.; Stolyarova, E.I.; Sundberg, U.; Lacerda, F.: Cross-language analysis of phonetic units in language addressed to infants. Science *277:* 684–686 (1997).
Ladd, D.R.: Peak features and overall slope; in Cutler, Ladd, Prosody: models and measurements (New York 1983).
Ladd, D.R.: Metrical representation of pitch register; in Kingston, Beckman, Papers in Laboratory Phonology. I: Between the grammar and the physics of speech, pp. 35–57 (Cambridge University Press, Cambridge 1990).
Ladd D.R.: In defense of a metrical theory of intonational downstep; in van der Hulst, Snider, The phonology of tone: the representation of tonal register, pp. 109–132 (Mouton de Gruyter, Berlin 1993).
Ladd, D.R.: Intonational phonology (Cambridge University Press, Cambridge 2008).
Ladd, D.R.; Faulkner, D.; Faulkner, H.; Schepman, A.: Constant 'segmental anchoring' of F0 movements under changes in speech rate. J. acoust. Soc. Am. *106:* 1543–1554 (1999).
Ladd, D.R.; Mennen, I.; Schepman, A.: Phonological conditioning of peak alignment in rising pitch accents in Dutch. J. acoust. Soc. Am. *107:* 2685–2696 (2000).
Ladd, D.R.; Morton, R.: The perception of intonational emphasis: continuous or categorical? J. Phonet. *25:* 313–342 (1997).
Ladd, D.R.; Schepman, A.: 'Sagging transitions' between high accent peaks in English: experimental evidence. J. Phonet. *31:* 81–112 (2003).
Liberman, A.M.; Harris, K.S.; Hoffman, H.S.; Griffith, B.C.: The discrimination of speech sounds within and across phoneme boundaries. J. exp. Psychol. *54:* 358–368 (1957).
Liberman, M.: The intonation system of English (MIT, Cambridge 1975).
Liberman, M.; Pierrehumbert J.: Intonational invariance under changes in pitch range and length; in Aronoff, Oerhle, Language sound structure, pp. 157–233 (MIT Press, Cambridge 1984).
Lisker, L.; Abramson, A.S.: The voicing dimension: some experiments in comparative phonetics. Proc. 6th Int. Congr. Phonet. Sci., 1967 (Academia Publishing House of Czechoslovak Academy of Sciences, Prague 1970).
Liu, F.; Xu, Y.: Parallel encoding of focus and interrogative meaning in Mandarin intonation. Phonetica *62:* 70–87 (2005).
Nash R.; Mulac A.: The intonation of verifiability; in Waugh, van Schooneveld, The melody of language, pp. 219–242 (University Park Press, Baltimore 1980).
Niebuhr, O.: Perceptual study of timing variables in F0 peaks. Proc. 15th Int. Congr. Phonet. Sci., Barcelona 2003, pp. 1225–1228.
Niebuhr, O.: Perzeption und kognitive Verarbeitung der Sprechmelodie: theoretische Grundlagen und empirische Untersuchungen (de Gruyter, Berlin 2007a).
Niebuhr, O.: The signalling of German rising-falling intonation categories: the interplay of synchronization, shape, and height. Phonetica *64:* 174–193 (2007b).
Niebuhr, O.; Kohler, K.J.: Perception and cognitive processing of tonal alignment in German. Proc. Int. Symp. Tonal Aspects of Lang.: Emphasis on Tone Lang., Beijing 2004, pp. 155–158.
O'Connor, R.J.; Arnold, G.F.: Intonation of colloquial English (Longman, Bristol 1973).
Pierrehumbert, J.: The phonology and phonetics of English intonation (MIT, Cambridge 1980).
Pierrehumbert, J.; Beckman, M.: Japanese tone structure (MIT Press, Cambridge 1988).
Pierrehumbert, J.; Steele, S.A.: Categories of tonal alignment in English. Phonetica *46:* 181–196 (1989).
Pike, K.L.: The intonation of American English (University of Michigan Publications, Ann Arbor 1945).
Purcell, E.: Pitch peak location and the perception of Serbo-Croatian word tone. J. Phonet. *4:* 265–270 (1976).

Redi, L.C.: Categorical effects in production of pitch contours in English. Proc. 15th Int. Congr. Phonet. Sci., Barcelona 2003, pp. 1647–1650.
Remijsen, B.; van Heuven, V.J.: Gradient and categorical pitch dimensions in Dutch: diagnostic tests. Proc. 14th Int. Congr. Phonet. Sci., San Francisco 1999, pp. 1865–1868.
Repp, B.H.: Categorical perception: issues, methods, findings; in Lass, Speech and language, advances in basic research and practice, pp. 243–335 (Academic Press, Orlando 1984).
Shattuck-Hufnagel, S.; Dilley, L.C.; Veilleux, N.; Brugos, A.; Speer, R.: F0 peaks and valleys aligned with non-prominent syllables can influence perceived prominence in adjacent syllables. Proc. Speech Prosody, Nara, 2004.
Silverman, K.; Beckman, M.; Pierrehumbert, J.; Ostendorf, M.; Wightman, C.W.S.; Price, P.; Hirschberg, J.: ToBI: a standard scheme for labeling prosody. Proc. 2nd Int. Conf. on Spoken Lang. Processing, Banff 1992, pp. 867–879.
Silverman, K.; Pierrehumbert, J.: The timing of prenuclear high accents in English; in Kingston, Beckman, Papers in Laboratory Phonology. I: Between the grammar and the physics of speech, pp. 71–106 (Cambridge University Press, Cambridge 1990).
Stevens, K.N.; Liberman, A.M.; Studdert-Kennedy, M.; Ohman, S.: Cross-language study of vowel perception. Lang. Speech *12:* 1–23 (1969).
Studdert-Kennedy, M.; Hadding-Koch, K.: Auditory and linguistic processes in the perception of intonation contours. Lang. Speech *16:* 293–313 (1973).
t Hart, J.; Collier, R.; Cohen, A.: A perceptual study of intonation (Cambridge University Press, Cambridge 1990).
Vanrell Bosch, M.d.M.: A scaling contrast in Majorcan Catalan interrogatives. Proc. Speech Prosody, Dresden 2006.
Vanrell, M.d.M.; Mascaro, I.; Torres-Tamarit, F.; Prieto, P.: When intonation plays the main character: information-vs. confirmation-seeking questions in Majorcan Catalan. Proc. Speech Prosody 2010, Chicago 2010.
Viechnicki, P.: Composition and granularity of vowel production targets (University of Chicago, 2002).
Vinegrad, M.D.: A direct magnitude scaling method to investigate categorical versus continuous modes of speech perception. Lang. Speech *15:* 114–121 (1972).
Ward, G.; Hirschberg, J.: Implicating uncertainty: the pragmatics of fall-rise intonation. Language *61:* 747–776 (1985).
Watson, D.G.; Tanenhaus, M.K.; Gunlogson, C.A.: Interpreting pitch accents in on-line comprehension: H* vs. L+H*. Cognit. Sci. *32:* 1232–1244 (2008).
Weber, A.; Braun B.; Crocker, M.W.: Finding referents in time: eye-tracking evidence for the role of contrastive accents. Lang. Speech *49:* 367–392 (2006).
Xu, Y.: Consistency of tone-syllable alignment across different syllable structures and speaking rates. Phonetica *55:* 179–203 (1998).
Xu, Y.: Effects of tone and focus on the formation and alignment of F0 contours. J. Phonet. *27:* 55–105 (1999).
Xu, Y.: Fundamental frequency peak delay in Mandarin. Phonetica *58:* 26–52 (2001).
Xu, Y.: Speech melody as articulatorily implemented communicative functions. Speech Commun. *46:* 220–251 (2005).
Xu, Y.; Wang, Q.E.: Pitch targets and their realization: evidence from Mandarin Chinese. Speech Commun. *33:* 319–337 (2001).
Xu, Y.; Xu, C.: Phonetic realization of focus in English declarative intonation. J. Phonet. *33:* 159–197 (2005).