

“No tone is an island: Investigating the relational nature of tones in English”

Laura Dilley

MIT/Ohio State University

LSA Annual Meeting

January 6, 2005

I. Introduction and overview

- a. Tonal features are defined in a strong and narrow way under standard autosegmental theory. In particular, there is no clear role in the theory for *relative height relations*, such as whether a tone is higher than or lower than another tone.
- b. This talk details some of the consequences of this assumption for cross-linguistic descriptive and explanatory adequacy in theories of phonology and phonetics. There are three parts to this talk:
 - i. *Background and logical arguments*: This section details for a subset of the world’s languages why it is critical that theories of phonology and phonetics build relative height relations into the representation (III – VIII)
 - ii. *Experimental results*: This section concerns results of an imitation experiment which tests whether relative height relations are part of the phonological or phonetic component of the grammar in English (IX-XII)
 - iii. *Theoretical proposals*: The last section describes a new proposal for incorporating relative height relations into tonal phonology (Dilley 2005). This proposal formalizes music theoretic constructs for tone in language and draws on facts about auditory perception (XIII-XVI).

II. Two competing views of tonal features

- a. There have traditionally been two competing approaches to phonological representations of tonal features (cf. Ladd 1996): *paradigmatic* and *syntagmatic*.
- b. *Paradigmatic*: Tones are defined with respect to a speaker’s pitch range, but not with respect to one another. Under this view, tones are viewed as inherent oppositions like segments (cf. /I/ in *pit* vs. /E/ in *pet*), cf. autosegmental theory (Leben 1973, Goldsmith 1976).
- c. *Syntagmatic*: Tones are defined with respect to one another in sequence. Under this view, tonal features are defined in relative terms, e.g. as higher than or lower than one another, cf. Jakobson, Fant, and Halle (1952).

- d. The predominating view is that of autosegmental theory, which claims that tonal features are exclusively paradigmatic. However, this assumption has been called into question (e.g., Snider 1999, Inkelas *et al.* 1986, Odden 1995).
- e. In the following, I present arguments that a descriptively sufficient theory of the phonetics and phonology of tone which seeks a cross-linguistic account of tonal phenomena *must* define syntagmatic restrictions on relative tone height. I will show that theories which lack restrictions on relative tone height cannot achieve descriptive adequacy for a significant number of languages.

III. Some cross-linguistic differences in the tonal character of languages

- a. *Claim*: While the assumptions of autosegmental theory that tonal features are exclusively paradigmatic have been sufficient to describe a significant number of languages, these assumptions provide an insufficient descriptive device for a subset of the world’s languages. This problem has not received significant attention, and it tends to undermine descriptive and experimental work on these languages.
- b. There are important differences in the phonetic and phonological behavior of tonal systems across languages. In particular:
 - i. In some languages, the lexicon constrains placement of tones in the pitch range to a significant extent (e.g., Mandarin, Yoruba).
 - ii. In other languages, the lexicon does not constrain placement of tones in the pitch range. In these languages, tones are essentially free to be phonetically realized (i.e., scaled) in different parts of the pitch range. (e.g., English, Spanish). I will refer to these languages as *free-scaling languages*.
- c. In the upcoming discussion, I will show that there is a crucial role for relative height relations in theories of phonology and phonetics for free-scaling languages such as English. To demonstrate this role, I will first describe some crucial phonetic facts that theories must account for.

IV. Descriptive adequacy in theories of phonetics and phonology

- a. One of the most significant findings in intonational phonology and phonetics in the last 25 years has been that F0 peaks and/or valleys are consistently aligned temporally with particular segmental positions (Ladd *et al.*, 1999, 2000; Arvaniti *et al.*, 1998, Dilley *et al.*, forthcoming). Moreover, depending on the specific types of F0 points – peaks or valleys – and their alignment, these points have consequences for meaning (e.g., d’Imperio and House 1997, Beckman and Pierrehumbert 1986).
- b. What are the implications of consistent presence and timing of F0 peaks and valleys for theories phonology and phonetics? In particular,

what constitutes a *descriptively adequate* account of these facts in free-scaling languages?

- c. A standard assumption in intonational and tonal phonology is that an F0 peak arises from a High (H) tone, while an F0 valley arises from a Low (L) tone. Then an F0 peak would typically be described as arising from a H tone in a LHL sequence, while an F0 valley would typically be described as arising from a L tone in a HLH sequence.
- d. Deriving a descriptively adequate account of the consistent presence and timing of F0 peaks and valleys is especially challenging in free-scaling languages like English, because of the fact that the lexicon does not constrain tone placement in the pitch range. As a result, individual H and L tones are free to be scaled either quite high or quite low in the pitch range. (See Figure 1).
- e. *Question:* If we assume e.g., that an F0 valley in general arises from the L tone of a HLH sequence, say H_1LH_2 , what *must* be true of the scaling of individual tones in this sequence?
- f. *Answer:* Whatever factors are assumed to scale H and L tones vertically in the pitch range, *they must prohibit L from rising above H_1 , and H_2 from falling below L*. In other words, L must be lower than either H_1 or H_2 .
- g. *Question:* What would be the consequence if these restrictions on the relative heights of adjacent tones were not in place in a theory of phonology and phonetics?
- h. *Answer:* Two types of problems result:
 - i. *Overgeneration* of phonetic contours from phonological sequences. If the relative heights of H_1 , L, and H_2 in H_1LH_2 are not restricted, then any of the schematic contours in Figure 2 might result from this sequence.
 - ii. *Indeterminacy* of phonetic contours with respect to phonological representations. If the relative heights of adjacent H and L tones are not sufficiently constrained, then a contour with an F0 valley could have arisen either from the L tone of a HLH sequence, or from the H tone of a LHL sequence!
 - iii. Implicit in both of these issues is the fact that such a theory would not be testable empirically.
- i. Do theories of phonology and phonetics for free-scaling languages define sufficient restrictions to achieve descriptive adequacy for F0 facts and avoid the potential problems cited above?

V. Pierrehumbert (1980) (P80)

- a. This theory proposed an account of English intonation in terms of H and L tones. There were four “types” of these tones: starred pitch

accentual tones, unstarred pitch accentual tones, phrase accents, and boundary tones.

- b. The theory is commonly assumed to have successfully extended autosegmental theory to English.
- c. The number of distinct categorical elements was fairly large, at 11 proposed categories (seven simple or complex pitch accents, two phrase accents, and two boundary tones).¹
- d. The pitch range scaling of individual tones in these 11 tonal categories was assumed to be controlled by approximately *ten* phonetic rules.
- e. These rules controlled the value of a parameter termed “prominence”, which gave rise to an F0 value through a set of transforms. Ladd (1990) has pointed out that P80’s “prominence” parameter bears little resemblance to phonetic or perceptual prominence. The phonetic rules for tone scaling were augmented by three additional (arbitrary) parameters which lack a phonetic interpretation.
- f. The crucial question of interest is whether this theory defines sufficient restrictions to achieve descriptive adequacy for F0 facts, as discussed in IV.
- g. *Question:* Does this theory define sufficient restrictions to achieve descriptive adequacy for F0 facts? **Crucially, does P80 restrict the relative heights of tones in H_1LH_2 so that H_1 does not fall below L, and L does not rise above H_2 ?**
- h. *Answer: No.* The phonetic rules in P80 impose sporadic restrictions on the relative heights of some adjacent tone pairs. These restrictions are insufficient to prevent problems of overgeneration of phonetic contours from the phonology and indeterminacy of phonological representations on the basis of phonetics.
- i. *Proof* (Dilley 2005): The following constitutes a proof that the relative heights of adjacent H and L tones in H^*L+H^* are unconstrained in P80. Thus, this sequence generates not only a falling-rising contour with an F0 valley, but also a rising-falling contour, and other contour shapes. We can first prove that the equations expressing the phonetic rules permitted L to be higher than H_1 in H_1LH_2 . We start with Eqn. 4 (P80, p. 145), which is given in (i). Here, $f(T)$ indicates the F0 value of a tone T, while $p(T)$ indicates the prominence value of T. n and k are arbitrary parameters.

(i)

$$f(L) = n \cdot f(H_1) \frac{p(H_1)}{p(L)}$$

¹ The number of pitch accents was later revised to six in Beckman and Pierrehumbert (1986).

Rearranging terms, we obtain the expression in (ii):

(ii)

$$\frac{f(L)}{f(H_1)} = n \cdot \frac{p(H_1)}{p(L)}$$

In this equation, $0 < n < k$. From Eqn. 3 (P80, p. 145), we know that $0 < k < 1$, so collapsing the two inequalities, we have $0 < n < k < 1$.

When $p(H_1)/p(L) > 1/n$, the F0 level of L will be higher than the F0 level of adjacent H₁. No restrictions are in place to prevent this situation from occurring.

We can also prove that equations permitted the level of the L to be higher than the level of the following adjacent tone, H₂. Rearranging terms in equations (i) and (ii), above, and substituting, we obtain the expression in (iii):

(iii)

$$\frac{f(L)}{f(H_2)} = \frac{n}{k} \cdot \frac{p^2(H_1)}{p(L)p(H_2)}$$

This proves that the F0 level of L is higher than the F0 level of H₂ precisely when $p^2(H_1)/[p(L)p(H_2)] > k/n$. Again, no restrictions are in place to prevent this from happening.

- j. *Question:* Is this an isolated problem with particular tonal sequences, or is the problem more pervasive?
- k. *Answer:* The problems of overgeneration and indeterminacy are widespread in P80 (Dilley 2005).
- l. As a result, the theory is not capable of accounting for robust phonetic facts concerning the presence and timing of F0 peaks and valleys. The theory is also cannot generate testable predictions, since it is impossible to associate a given intonation pattern with a unique phonological representation.
- m. These problems have been alluded to by Ladd (1990), who writes (p. 37): “Unconstrained gradient variability of prominence and other pitch range parameters is, in my view, the most serious empirical weakness of a great many quantitatively explicit models of F0.” However, no explicit discussion or demonstration of these problems has been offered, and no solutions have been proposed. The field has ignored this problem for a significant period of time.
- n. What about other theories of the phonology/phonetics relation?

VI. Pierrehumbert and Beckman (1988) – PB88

- a. This theory was proposed originally for Japanese but has been applied to numerous other languages, including English. One difference with respect to P80 was that PB88 revised the theory of the phonetics-phonology relation assumed under P80.
- b. In this theory, H and L are assumed to be assigned prominence values ranging from 0 to greater than 1.0. Prominence is defined in the sense of P80, rather than in terms of perceived or acoustic prominence.
- c. The prominence value of H tones defined their position with respect to a high reference tone line, *h*. Similarly, the prominence value of H tones defined their position with respect to a low reference tone line, *l*.
- d. *Question: Crucially, did the theory of PB88 restrict the relative heights of adjacent tones in sequence*, so that in e.g., H₁LH₂, L does not rise above H₁, and H₂ does not fall below L?
- e. *Answer: No.* There are *no* restrictions on the relative heights of adjacent tone pairs in PB88, unlike P80, which imposed sporadic restrictions. Thus, PB88 apparently did not consider the problems of overgeneration and indeterminacy associated with the proposed phonetic mechanisms.
- f. These restrictions are insufficient to prevent problems of overgeneration of phonetic contours from the phonology and indeterminacy of phonological representations on the basis of phonetics. No solution to these problems has been proposed.
- g. While restrictions on tone height are not in place in either PB88 or P80, there is a distinct version of the theory in which such restrictions are in place. This is the ToBI transcription system (Beckman and Ayers-Elam 1997).

VII. The ToBI system

- a. The ToBI transcription system associates particular F0 contour shapes with particular pitch accentual categories. As a result, it *implicitly* places restrictions on syntagmatic relative height relationships. However, these restrictions are nowhere codified theoretically (e.g., not in P80 or PB88) and they are assumed to exist between tone pairs sporadically rather than systematically.
- b. By placing implicit restrictions on relative tone height, ToBI makes many of the original claims of P80 testable.
- c. Dilley (2005) conducted several perception and production experiments testing categories claimed in ToBI. Support was found for some distinctions, e.g., H+L* vs. H*. However, some ToBI categories, e.g. H*, apparently corresponded to more than one category, while other claimed category distinctions (e.g., H* vs. L+H*) should apparently be collapsed.

- d. Would it be a productive approach to codify the restrictions on relative height implicit in ToBI?
- e. Several lines of evidence in addition to the results of Dilley (2005) suggest that codifying the implicit restrictions in ToBI would yield only limited returns.
 - i. Recent phonetic findings converge with Dilley (2005) by presenting evidence against the formulation of P80 (Arvaniti *et al.* 1998; Ladd *et al.*, 1999, 2000; Ladd and Schepman 2003)
 - ii. The theory proposed in P80 lacks phonetic transparency (Ladd 2000).
 - iii. Even when restrictions on the scaling of adjacent tones is assumed, as in ToBI, the theory overgenerates possible pitch accentual sequences (Pierrehumbert 2000).
 - iv. Some of the original Africanist analyses and assumptions on which P80's analysis of English is based have since been cast into doubt (Snider and van der Hulst 1993, Ladd 1990, 1993).
- f. What other existing options are available for addressing these problems?

VIII. Other approaches to phonetics and phonology

- a. Some less comprehensive theories of the relationship between phonetics and phonology have been proposed (e.g., Liberman and Pierrehumbert 1984, van den Berg *et al.* 1992).
- b. How productive would it be to modify these theories to include sufficient syntagmatic restrictions on relative tone height, and how would this be accomplished?
- c. These theories include the common assumption that tones are scaled with respect to abstract reference lines. It would be necessary to specify restrictions on the relative heights of abstract reference lines, as well as on the scaling of tones with respect to these lines, so that adjacent tones have restricted relative height relations.
- d. Some *prima facie* disadvantages of the reference line analysis:
 - i. In most of these theories, the level of a tone can vary freely with respect to the level of the reference line; thus, the pitch of the reference line cannot be recovered in any systematic way from the level of the tone itself. In other words, a crucial part of the representation under such theories (i.e., the level of reference line or lines) is indeterminate.
 - ii. Under versions of these theories which have not yet been proposed, but which might be, a tone might be assumed to be situated on a reference line or a fixed distance from the reference line. Then in theory the indeterminacy in (i) could

be eliminated, since the level of the reference line would be phonetically recoverable.

- iii. Then restricting the relative heights of tones would be accomplished by restricting the relative heights of reference lines. The problem is that in such a theory the reference lines "do no work" and simply constitute extra parameters in the theory.
- iv. Thus, theories which might be proposed in which relative height was restricted by means of phrasal reference lines inevitably are more complicated than theories which involve syntagmatic restrictions on the levels of tones directly in the phonology, all else being equal.
- e. *Conclusion:* In the limit, theories which might seek to account for syntagmatic relations of relative height between tones by means of paradigmatic features plus phonetic rules for scaling tones with respect to phrasal reference lines are more complex and less phonetically transparent than theories which build syntagmatic restrictions on tone level directly into the phonology.

IX. Investigating the status of relative pitch level as phonetic or phonological

- a. The arguments presented in III-VI suggest that in free-scaling languages there must be syntagmatic restrictions on the relative heights of adjacent H and L tones at some level of the grammar.
- b. *Questions:* Are syntagmatic relations of relative height part of the phonetics, or the phonology? How can we determine whether syntagmatic restrictions on relative height are part of the phonetic or phonological component of the grammar?
- c. The approach in P80 and PB88 was to assign syntagmatic relations of relative height in whole or in part to the phonetic component of the grammar.
- d. Work in laboratory phonology suggests that the hallmark of phonological categories is *demonstrating that a gradient or continuous phenomenon is interpreted in terms of discrete categories*. Thus, to test phonetic status we should ideally design an experiment in which some phonetic parameter is manipulated along a continuum to determine whether speakers and listeners treat that parameter as continuous or discrete.
- e. *Question:* What sort of experimental paradigm might be used to investigate the status of relative pitch level as phonetic vs. phonological?
- f. Pierrehumbert and Steele (1989), or PS89, successfully used an imitation task to investigate phonological categories in intonation. In their experiment, the timing of an F0 peak was shifted along a

continuum through the SW syllable sequence *million-* in *millionaire*. (See Figure 3.) Subjects appeared to impose categories on the timing of F0 peaks. They were not able to faithfully reproduce the full range of continuously-spaced F0 peak positions, but rather imitated stimuli in a manner that gave rise to a discrete distribution of F0 peak positions.

- g. This suggests that some version of the PS89 task could be used to investigate whether syntagmatic relative height relations are part of the phonetic or phonological component of the grammar, by determining whether such relations give rise to a discrete or continuous distribution of phonetic points.
- h. It turns out that in laboratory phonology studies of intonation, discreteness in the timing of F0 peaks and valleys in production tasks is held to be the “gold standard” as an indicator of distinct phonological categories (Gussenhoven 2004).
- i. **Question: Can we elicit a discrete distribution of F0 peaks and valleys in an imitation experiment in which relative pitch level has been manipulated along a continuum in the stimuli?**
- j. The intuition behind this approach involves a reinterpretation of the PS89 results. PS89 interpreted the observed discreteness in F0 peak timing in terms of two paradigmatically-defined pitch accentual categories: L*+H and L+H*. Several points can be made.
- k. First, the discussion earlier in IV-VI showed that, strictly speaking, the phonetic rules and mechanisms in P80 and PB88 lead to problems of overgeneration and indeterminacy. Thus it is not possible to uniquely associate the contours used in PS89 with L*+H and L+H* but with many other phonological categories as well. An interpretation in terms of L*+H and L+H* is possible only under ToBI, but in this system the phonetic mechanisms/rules defining restrictions on relative height relations of adjacent tones have not been made explicit.
- l. Second, there is another interpretation of the results from PS89 which is compatible with an account in terms of *syntagmatic phonological relations*. Consider what happens to the relative pitch levels of the S and W syllables as an F0 peak is shifted from early in the SW sequence to successively later time positions (cf. Figure 3). Initially, the S syllable will sound higher in pitch than the W syllable, until a point in time when the W syllable begins to sound higher in pitch than the S syllable. Thus, *shifting an F0 peak in time across a SW syllable sequence leads to a (categorical) change in the relative pitch level of the S and W syllables*.
- m. **Question: How do peaks and valleys relate to relative pitch level?**
- n. A rather widespread assumption is that F0 peaks and valleys are the most important phonetic correlates of H and L tones, i.e., these points are assumed to be direct phonetic realizations of (paradigmatically-

defined) phonological entities. Consistent with this assumption, the timing of F0 peaks and/or valleys is considered diagnostic of particular categories in intonation in ToBI.

- o. In contrast, Dilley (2005) claims that the perceived pitch levels of TBU’s relative to one another are the most significant phonetic correlates of tonal elements, rather than F0 peaks and valleys. Moreover, it is claimed that *consistent F0 peak or valley timing occurs because speakers attempt to make a particular syllable/TBU have the locally highest or lowest pitch relative to other syllables/TBU’s*.
- p. **Experimental questions:**
 - i. Are syntagmatic relations of relative height part of the phonological or the phonetic component of the grammar?
 - ii. Can categorical effects in F0 peak and valley timing be elicited in an imitation task by varying relative pitch level, rather than F0 peak and valley timing *per se*?
 - iii. What is the most important phonetic correlate of phonological categories in F0: relative pitch level, or F0 peaks and valleys?

X. Imitation experiment - Methods

- a. The full experiment is described in Dilley (2005).
- b. Stimuli were based on the English phrase *Some lemonade*. The phrase had an overall rising-falling contour (i.e., statement intonation) or a falling-rising contour (i.e., question intonation).
- c. The F0 contours for the stimuli were altered using Praat (Boersma and Weenink 2002). In particular, the F0 across each syllable of *lemon-* was replaced with a flat F0 (Figure 4). This gave rise to four stimulus series:
 - i. In Series A & C, there were 10 steps of ½ semitone. For Series A, the highest step on *le-* was paired with the lowest step on *-mon* to create the first stimulus, the next highest step was paired with the next lowest step to create the second stimulus, etc.
 - ii. In Series B & D, there were 6 steps of ¾ semitone. The level of *le-* was fixed while the level of *-mon* was varied.
- d. Noise was spliced in to eliminate information about the direction of an F0 transition to and from a syllable, i.e., whether the contours were rising or falling, since this info is redundant with the temporal location of an F0 peak or valley.
- e. Thus the following things were true of these stimuli:
 - i. Pitch level was manipulated along a continuum.
 - ii. There are no explicit phonetic cues to *either the presence or the timing* of an F0 peak or a valley.
- f. **Task.** There were 13 subjects. The subjects were told that they would hear the phrase “Some lemonade” together with some noise, and that

they should repeat the phrase as closely as possible, ignoring the noise. Stimuli in Series A & B, C & D were randomized within two blocks and presented over headphones, after which subjects produced an imitation of each stimulus into a high-quality microphone.

- g. If the most important phonetic characteristic of tones in free-scaling languages is relative pitch level, then speakers should produce F0 peaks and valleys in response to these stimuli, even though there are no such points in the stimuli. In particular, they should produce F0 peaks in response to Series A & B and F0 valleys in response to Series C & D.
- h. If syntagmatic relations of relative height are part of the phonology rather than the phonetics, then speakers should interpret this continuous variation in relative pitch level in a categorical way. In particular, speakers should produce e.g., an early-timed peak on *le-* in imitated versions of stimuli 1-5 in Series A but a late-timed peak on *-mon* in stimuli 6-10 in this series. Moreover, we expect to see evidence of a break point between Stimuli 5 and 6.
- i. In contrast, P80, PB88 and ToBI do not generate any clear predictions about the outcome of this experiment. In particular, since it is assumed that tones are scaled individually with respect to pitch range parameters, there is no reason that we would expect a difference in subject behavior in response to the initial syllable in Series B & D. The level of this syllable is constant across all stimuli, so there is no reason to expect under these theories that the syllable would be treated as a “H” tone over one range of stimuli but not as a “H” tone over another range of stimuli.
- j. *Analysis.* The timing of any F0 peaks and valleys in imitated versions of stimuli was determined with respect to the target two-syllable sequence. In particular, normalized peak or valley time was calculated by determining the time of the peak or valley relative to the beginning of the /l/ in *lemon-*, divided by the duration of *lemon-*.

XI. Imitation experiment – Results

- a. Subjects produced F0 peaks in response to stimuli in Series A & B and F0 valleys in response to stimuli in Series C & D (Figure 5).
- b. Speakers interpreted continuous variation in relative pitch level in a categorical way.
 - i. For Series A & B, subjects produced F0 peaks during the first syllable, *le-*, in response to stimuli with a higher relative pitch on the first syllable, but peaks during the second syllable, *mon-*, in response to stimuli with a higher relative pitch on the second syllable (Figures 6 and 7).
 - ii. For Series C & D, subjects produced F0 valleys during the first syllable, *le-*, in response to stimuli with a lower relative

pitch on the first syllable, but F0 peaks during the second syllable, *mon-* in response to stimuli with a lower relative pitch on the second syllable (Figure 8).

- c. Speakers interpreted continuous variation in relative pitch level in a categorical way, as evidenced by a clear discontinuity or break point at the locations of a change in relative pitch level, in three of the stimulus series. Discontinuities were observed between stimuli 5 and 6 in Series A, between stimuli 5 and 6 in Series C, and between stimuli 3 and 4 in series D.

XII. Imitation experiment – Summary and Conclusions

- a. Discreteness in the timing of F0 peaks and valleys was elicited in an imitation task. This is the “gold standard” for categorical distinctiveness in studying intonation in laboratory phonology.
- b. These categorical effects in timing were elicited by continuous variation in relative pitch level. Locations of category boundaries were at positions in the stimulus series where relative pitch level changed.
- c. These results therefore suggest that syntagmatic restrictions on relative tone height are *part of the phonological, rather than the phonetic, component of the grammar in English.*
- d. The results also support the proposal of Dilley (2005) that the most significant phonetic attribute of F0 contours of the phonology is the relative pitch level of TBU’s, rather than F0 peak and valley timing *per se.*
- e. Finally, this is the first experiment to demonstrate categorical effects in F0 valley placement in a production task in response to a stimulus continuum.

XIII. What about modifying previous approaches to tonal description?

- a. I have so far shown that in order to avoid problems of phonological overgeneration and phonetic indeterminacy, it is necessary to include syntagmatic restrictions on relative tone height. Results of an imitation experiment showed that these restrictions should be incorporated into the phonological component of the grammar, rather than the phonetic component.
- b. *Question:* How can we build syntagmatic restrictions into a phonology of tone?
- c. *Answer:* We need relational, syntagmatic features for tones.
- d. *Question:* Is there evidence from other languages that syntagmatic restrictions are warranted in the phonology?
- e. *Answer:* Yes, there is.
 - i. An example of contrastive use of downstep comes from Igbo: *ámá* ‘street’ vs. *ám’á* ‘distinguishing mark’ (Williamson

- 1972).² Distributional information for tone does not support a treatment of downstepped H as a distinct tonal category (V. Manfredi, pers. comm.).
- ii. An example of contrastive use of upstep comes from Acatlán Mixtec (Pike and Wistrand 1974): *?ikumida* ‘we (incl.) have’ is MHHH, while *?ikumida* ‘you (pl. fam.) have’ is MHHU, where U indicates upstep.³
 - iii. Strings of identical H tones show three different relative height relations in several Bantu languages. For example, in Kenyang, sequences of H’s step down; in Kimatuumbi, they step up; and in Kipare, they stay at the same level (Odden 1995).
 - iv. To account for cross-linguistic facts, we need a phonological means of making a three-way relative height contrast: going higher, going lower, and staying at the same level.
- f. Thus, we have evidence from both lexical tone languages and free-scaling languages that syntagmatic restrictions on relative tone level are part of the phonology/UG.
 - g. *Question*: Can standard autosegmental theory account for these facts?
 - h. *Answer*: No. Standard autosegmental theory encodes only paradigmatic tonal features. The only mechanism which deals with relations between tones is the Obligatory Contour Principle (OCP). The OCP claimed that adjacent tones “spread” giving rise phonetically to a level contour. Absence of a phonetically level contour is then described as an OCP violation. There is no clear way to map a binary distinction (OCP violation vs. no violation) onto a three-way distinction. Moreover, the predictions of the OCP have not held up to empirical data (Odden 1995).
 - i. Snider (1999) has proposed a modification of autosegmental theory which includes syntagmatic features. He proposes two tone features (H, L) and two register features (h, l), building on the work of Yip (1980, 1993), Inkelas *et al.* (1986), Pulleyblank (1986), and others. Two arguments suggested that this approach would not be appropriate for an account of a free-scaling language like English:
 - i. The proposal was developed based on data from languages in which pitch range is lexically determined, determining the pitch range placement of e.g., H and L. It is not clear how this proposal could be extended without running into the same problems as P80.

² I would like to thank Victor Manfredi for suggesting this example.

³ I would like to express my appreciation to Keith Snider for suggesting this example in response to my query.

- ii. The proposal assumes a powerful component, the tone-register ratio, which can take a continuous range of values and which underdetermines the syntagmatic relation of some tones. This seems unsatisfactory for a language in which syntagmatic relations are of paramount importance in the phonology.
- j. Another approach to syntagmatic features was the work of Clark (1978), who proposed a theory of dynamic syntagmatic tone features. However, Clark’s proposal treated variation in pitch range placement as strictly phonetic, rather than phonological. Thus it is at a disadvantage compared with theories in which the phonology determines tone placement.
- k. It appears that we need another approach which blends syntagmatic and paradigmatic tone features.

XIV. Tone interval theory (Dilley 2005)

- a. This theory proposes that the representational structures of tonal patterns in language and musical melody are similar. The inspiration for this proposal comes from music theory, where this domain presents a straightforward means of integrating syntagmatic and paradigmatic tonal features into a unified linguistic framework. The framework is therefore not arbitrary, but attested in another domain of auditory perception and cognition.
- b. ♪ Musical melodies are represented in terms of *relationships among notes*. Burns (1999: 218) states: “...melodic information in music is mediated by the *frequency ratio relationships* among tones (i.e., the musical intervals), not by their absolute pitches.”⁴
- c. ♪ Properties of musical melodies include the following (Handel 1989):
 - i. *Contour*: The pattern of ups and downs of successive notes.
 - ii. *Interval*: The tonal (or pitch) distance between successive notes.
 - iii. *Scale Note (or Key)*: The intervallic relationship between each note and a referent note (i.e., the musical key).
- d. Tone interval theory proposes that tonal patterns are represented in terms of *tone intervals*, which are abstractions of frequency ratios, e.g., musical intervals.
- e. Under this proposal, tones (*T*) are joined into tone intervals (*I*). A tone interval $I_{r,T}$ relates a tone *T* to a referent *r* according to the following general formula:

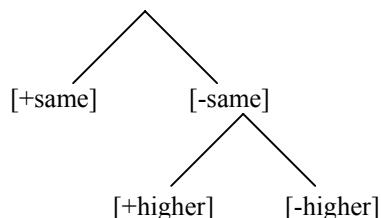
$$I_{r,T} = T/r$$
- f. Two types of referents may be defined, giving rise to two distinct types of tone intervals:

⁴ Emphasis supplied.

- i. *Syntagmatic tone interval*: Defines a relationship between a tone and another tone in sequence, e.g. for sequentially-ordered tones $T_1 T_2$:

$$I_{1,2} = T_2/T_1$$
 - ii. *Paradigmatic tone interval*: Defines a relationship between a tone and a referent tone level or “tonic”, μ . Then for a tone T_0 :

$$I_{\mu,0} = T_0/\mu$$
- g. Syntagmatic tone intervals are based on a direct analogy with the notions of *Contour* and *Interval* in music; paradigmatic tone intervals are based on a direct analogy with the notion of *Scale Note* in music
- h. In this theory, *tones do not have features; rather, tone intervals have features*.
- i. Tones are featureless, discrete-time events which are associated with metrical grids.
 - ii. In contrast, tone intervals are defined by two relational features: $[\pm\text{same}]$ and $[\pm\text{higher}]$, arranged in the following hierarchical structure. These features are proposed to be universal. Because of the feature geometry, three relations are expressed in tone intervals: *higher* ($[-\text{same}, +\text{higher}]$), *lower* ($[-\text{same}, -\text{higher}]$), and *same* ($[+\text{same}]$).



- i. Each relational feature specification (*higher*, *lower*, or *same*) defines a particular range of tone interval values (cf. frequency ratio values):
 - i. *higher* $\Rightarrow I_{r,T} > 1$ ($T > r$)
 - ii. *lower* $\Rightarrow I_{r,T} < 1$ ($T < r$)
 - iii. *same* $\Rightarrow I_{r,T} = 1$ ($T = r$)
- j. Restrictions on tone interval values specify a relative position of a tone with respect to a referent in the abstract tonal space.
- k. Syntagmatic tone intervals involving the relations *higher*, *lower*, and *same* ($I > 1$, $I < 1$ and $I = 1$) are assumed to be *universally attested in all linguistic tonal systems*.
- l. Language-specific restrictions on tone interval values give rise to language-specific syntagmatic and/or paradigmatic constructs.
- m. Paradigmatic tone intervals are assumed to be represented in *only a subset of languages*.

- n. Finally, restrictions on syntagmatic tone interval values are assumed to be specified in language-specific ways.
- o. Tone interval theory retains the benefits of autosegmental theory by assuming that tones are autonomous from segments, but temporally coordinated with them.
- p. So far, the theory has been worked out in detail only to English intonation (Dilley 2005), but preliminary work suggests that the system is capable of accounting for a wide range of phenomena in African and Asian tonal systems. If this is correct, it would obviate the need for two “tiers” of tonal information for representing e.g., tonal register (cf. Yip 1980, 1993; Hyman 1993, others).

XV. Some parallels between music and tone interval theory

- a. ♩ Both universal commonalities and striking differences can be seen across musical cultures and genres; these commonalities and differences are captured by the musical notions of *Contour*, *Interval*, and *Scale Note*.
- b. Similarly, across linguistic tonal systems, both universal commonalities and striking differences can be seen; these commonalities and differences are captured in the tone interval theoretic equivalents of the musical notions *Contour*, *Interval*, and *Scale Note*.
- c. ♩ In music, *Contour* is a universal and basic aspect of all musical melodies (Dowling and Fujitani 1971, Dowling and Harwood 1986).
- d. Similarly, syntagmatic tone intervals defining the up-and-down patterning of tones are assumed to be common to all linguistic tonal systems.
- e. ♩ Out of all musical melodies, a subset of melodies define specific cutoff values for frequency ratios, thereby defining *Interval* and *Scale Note* constructs (Burns 1999).
 - i. Some melodic systems apparently define no interval or key constructs, e.g. Australian Aboriginal music (U. Will, pers. comm.)
 - ii. For those systems which do define *Interval* and *Scale Note* constructs, the specific frequency ratios that are defined in the representation vary widely across melodic systems of the world’s cultures (Burns 1999).
- f. Similarly, it is assumed in the theory that individual linguistic tonal systems may define specific cutoff values for (syntagmatic or paradigmatic) tone interval ratios, and that these cutoffs are highly specific to individual languages.
- g. Moreover, drawing an analogy with the notion of *Interval*, language-specific contrastive differences in the tonal distances between two

tones in sequence may be represented by defining language-specific cutoff bounds on syntagmatic tone interval ratios.

- h. ♪ Out of all melodies across musical cultures and forms, a *Key* is defined only for a subset of melodies, e.g. tonal melodies in Western music (e.g., Dowling and Fujitani 1971, Dowling and Harwood 1986).
- i. The fact that *Scale Notes* are defined only for a subset of melodies is reflected in the assumption that paradigmatic tone intervals are represented only in a subset of linguistic systems.
- j. ♪ Significantly, there is an implicative relationship among the properties of *Contour*, *Interval*, and *Scale Note*. In particular, if the *Scale Notes* is defined, then the distances of all notes to a referent “tonic” note (i.e., the musical key) are known, and sequential information about the *Interval* and *Contour* between notes can be readily derived.
- k. Note that autosegmental theory did not codify the fact that a paradigmatically-defined sequence of H and L tones by default gives rise to a representation in terms of ups and downs between adjacent tones.
- l. ♪ Musical melodies are defined with respect to a hierarchical metrical structure or “grid” (Lerdahl and Jackendoff 1983, Jackendoff 1989).
- m. Similarly, tone interval theory assumes that tones associate universally with respect to metrical grid structures.

XVI. A tone interval theoretic approach to English intonation

- a. The primary advantage of this framework so far is to present a straightforward means of formalizing the relationship between phonology and phonetics for free-scaling languages like English in which placement of tones in the pitch range is not lexically determined.
- b. In particular, tone interval theory permits a formal account of tone phenomena in free-scaling languages while avoiding problems of overgeneration of phonetic contours from the phonology and indeterminacy of phonetic contours with respect to the phonology.
- c. This was not accomplished in P80 nor in subsequent work, and the problems have been outstanding for approximately 30 years since the proposals of Goldsmith (1976).
- d. Tone interval theory provides a means of integrating syntagmatic and paradigmatic features into a unified framework. It does this by developing a proposal which is based on an attested tonal representation system, that of music.
- e. Rather than charting a course which diverges from recent work, *the tone interval theoretic account of English intonation builds on previous work by placing descriptive linguistic studies of different languages carried out in the framework of P80 and subsequent work on firmer theoretical footing.*

- f. Dilley (2005) shows that tone interval theory permits an account of English intonational phenomena using just *six tonal elements and no phonetic implementation rules*.
 - i. Two types of tones (T^* and T)
 - ii. Three types of syntagmatic tone intervals ($I > 1$, $I < 1$ and $I = 1$ specifying *higher*, *lower*, and *same*).⁵
 - iii. Compare P80’s 11 tonal elements and ~10 phonetic implementation rules.
- g. Universal principles of phonological association between tones and metrical grids are proposed which require e.g. unstarred tones, T , to align with respect to leftward or rightward starred tones, giving $T+$ or $+T$
- h. A simple change in notation permits a representation in terms of tone intervals to give rise to a ToBI-like description of tonal patterns.
 - i. E.g., starred tones which are *higher*, *lower*, or the *same* level as a previous tone are H^* , L^* or E^* ⁶
 - ii. E.g., the distinction claimed in P80 between bitonal $L+H^*$ and L^*+H is captured in tone interval theory as a sequence of tones, $L+H^*$ and L^*+H .
 - iii. Representing “bitonal” evens as sequences of independent but coordinated tones brings the phonology into line with findings from recent phonetic studies (Arvaniti et al. 1998; Ladd et al. 1999, 2000; Dilley et al., in press).
- i. The theory provides a solution to several problems which are outstanding in the intonation literature, including those cited by Ladd and Schepman (2003) and Arvaniti, Ladd and Mennen (2000).
- j. *Summary*: By defining tonal categories as in tone interval theory, we avoid problems of overgeneration and indeterminacy while putting descriptive linguistic work that has taken place over the past 25 years on firmer theoretical footing. A simple change of notation permits previous linguistic analyses based on the work of P80 and colleagues to be re-expressed in tone interval theoretic terms in a virtually transparent manner.

XVII. Overall summary and conclusions

- a. Previous work has failed to successfully extend autosegmental theory to free-scaling languages, despite widespread assumptions to the contrary.

⁵ One language-specific cutoff on a syntagmatic tone interval value is additionally needed to account for the calling contour (e.g., Liberman 1975).

⁶ Here, italics are used to distinguish tones as defined in tone interval theory from paradigmatically-defined tones assumed in P80 and later work.

- b. In particular, a lack of sufficient restrictions on the relative heights of tones in free-scaling languages leads to chronic problems:
 - i. *Overgeneration* of phonetic contours from phonological sequences
 - ii. *Indeterminacy* of phonetic contours with respect to the phonological representation
- c. The theories of phonology and phonetics described in both P80 and PB88 gave rise to these problems.
- d. An imitation experiment showed evidence of categorical distinctions in the form of discreteness in production data in the timing of F0 peaks and valleys, which is the “gold standard” for phonological distinctiveness in laboratory phonology.
- e. Converging evidence across languages suggests that syntagmatic relations are needed in a descriptively adequate theory of the phonetics and phonology of tone. It is not clear how autosegmental theory could be modified to eliminate the problems with overgeneration and indeterminacy in free-scaling languages.
- f. Tone interval theory integrates syntagmatic and paradigmatic features by building on the attested representational system for tone found in theories of musical melody.
- g. By building syntagmatic features into the phonology, tone interval theory proposes a solution to the problems of indeterminacy and overgeneration which have persisted since early proposals of autosegmental theory.
- h. Finally, a simple change of notation permits earlier descriptive work on free-scaling languages to be re-expressed in tone interval theoretic terms.
- i. If syntagmatic features are universally attested, as claimed in tone interval theory, then it would truly be the case that “no tone is an island”. Rather, tones are interrelated and relative to one another, whether the domain is language, or music, or auditory perception in general.

References

- Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *J. of Phonetics*, 26, 3-25.
- Arvaniti, A., R., L. D., & Mennen, I. (2000). What is a starred tone? Evidence from Greek. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V* (pp. 119-130). Cambridge, UK: Cambridge University Press.
- Beckman, M., & Ayers-Elam, G. (1997). *Guidelines for ToBI labeling, version 3.0*: The Ohio State University.
- Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255-309.
- Boersma, P., & Weenink, D. (2002). Praat, a system for doing phonetics by computer (Version 4.0.26): Software and manual available online at <http://www.praat.org>.
- Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed., pp. 215-264). San Diego: Academic Press.
- Clark, M. M. (1978). *A dynamic treatment of tone, with special attention to the tonal system of Igbo*. Ph.D. dissertation, University of Massachusetts at Amherst.
- Dilley, L. (2005) The phonetics and phonology of tonal systems. Ph. D. Dissertation, MIT.
- Dilley, L., Ladd, D. R., & Schepman, A. (in press). Alignment of L and H in bitonal pitch accents: Testing two hypotheses. *J. of Phonetics*.
- d'Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment*. Unpublished Ph.D. dissertation, The Ohio State University.
- d'Imperio, M., & House, D. (1997). Perception of questions and statements in Neapolitan Italian, *Proceedings of Eurospeech 97* (Vol. 1, pp. 251-254). Rhodes, Greece.
- Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *J. Acoust. Soc. Am.*, 49(2), 524-531.
- Dowling, W. J., & Harwood, D. L. (1986). *Music Cognition*. Orlando, FL: Academic Press, Harcourt Brace Jovanovich Publishers.
- Goldsmith, J. (1976). *Autosegmental phonology*. Ph.D. dissertation, MIT, Cambridge, MA.
- Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology*, 17, 143-185.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Handel, S. (1989). *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press.
- House, D. (1990). *Tonal Perception in Speech*. Lund: Lund University Press.
- Hyman, L. M. (1993). Register tones and tonal geometry. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 75-108). Berlin, New York: Mouton de Gruyter.
- Inkelas, S., Leben, W. R., & Cobler, M. (1986). The phonology of intonation in Hausa. In S. Berman & J. Choe & J. McDonough (Eds.), *Proceedings of the 16th Annual Meeting of NELS*. Amherst: GLSA, University of Massachusetts.
- Jackendoff, R. (1989) A comparison of rhythmic structures in music and language. In P. Kiparsky and G. Youmans (eds.), *Phonetics and Phonology*, Vol. 1, Academic Press, New York, 15-44.
- Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.
- Kohler, K. J. (1987). Categorical pitch perception. In U. Viks (Ed.), *Proceedings of the 11th International Congress of Phonetic Sciences* (Vol. 5, pp. 331-333). Tallinn.
- Ladd, D. R. (1990). Metrical representation of pitch register. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 35-57). Cambridge: Cambridge University Press.
- Ladd, D. R. (1993). In defense of a metrical theory of intonational downstep. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 109-132). Berlin, New York: Mouton de Gruyter.
- Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *J. Acoust. Soc. Am.*, 106(3), 1543-1554.
- Ladd, D. R., Mennen, I., & Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *J. Acoust. Soc. Am.*, 107(5), 2685-2696.
- Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high accent peaks in English: experimental evidence. *J. of Phonetics*, 31, 81-112.
- Leben, W. R. (1973). *Suprasegmental phonology*. Ph.D. dissertation, MIT, Cambridge, MA.
- Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Lieberman, M. (1975). *The intonational system of English*. Ph.D. dissertation, MIT.
- Lieberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oerhle (Eds.), *Language Sound Structure* (pp. 157-233). Cambridge, MA: MIT Press.
- Odden, D. (1995). Tone: African languages. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 444-475): Blackwell.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Ph.D. dissertation, MIT, Cambridge, MA.
- Pierrehumbert, J. (2000). Tonal elements and their alignment. In M. Horne (Ed.), *Prosody: Theory and Experiment* (pp. 11-36). Dordrecht: Kluwer Academic Publishers.
- Pierrehumbert, J., & Beckman, M. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.
- Pierrehumbert, J., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, 46, 181-196.
- Pike, E. V., & Wistrand, K. (1974). Step-up terrace tone in Acatlan Mixtec (Mexico). In R. M. Brend (Ed.), *Advances in Tagmemics, North Holland Linguistic Series* (Vol. 9, pp. 81-104). Amsterdam: North-Holland.
- Pulleyblank, D. (1986). *Tone in Lexical Phonology*. Dordrecht: D. Reidel Publishing Co.
- Snider, K. (1999). *The Geometry and Features of Tone*. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington Publications in Linguistics 133.
- Snider, K., & van der Hulst, H. (1993). Issues in the representation of tonal register. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 1-27). Berlin, New York: Mouton de Gruyter.
- van den Berg, R., Gussenhoven, C., & Rietveld, A. C. M. (1992). Downstep in Dutch: Implications for a model. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (pp. 335-359). Cambridge: Cambridge University Press.
- Yip, M. (1980). *The tonal phonology of Chinese*. Ph.D. dissertation, MIT, Cambridge, MA.
- Yip, M. (1993). Tonal Register in East Asian Languages. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 245-268). Berlin, New York: Mouton de Gruyter.

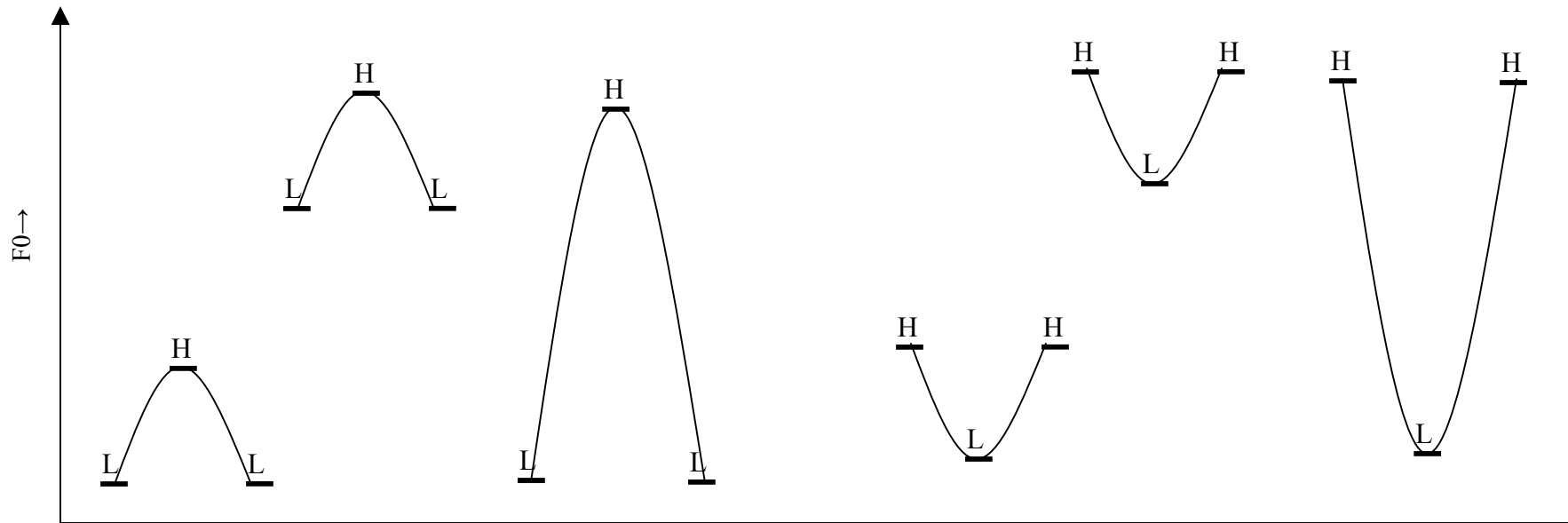


Figure 1. Schematic representation of how the positions of H and L tones can vary in their placement in the pitch range in a free-scaling language.

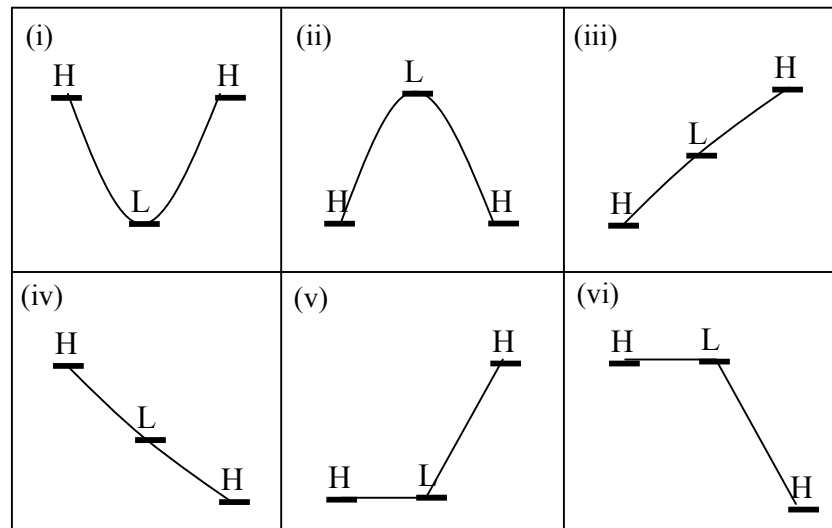


Figure 2. Some F0 contours which could result if the relative heights of adjacent tones are not sufficiently constrained in a theory of phonetics and phonology.

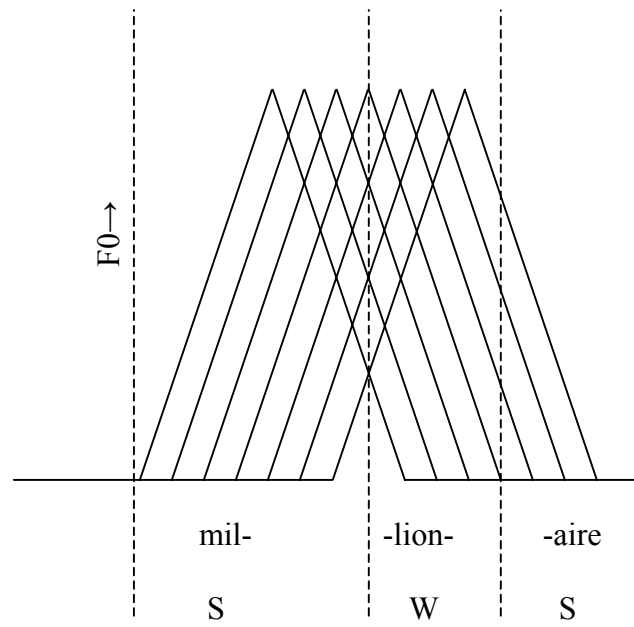


Figure 3. Schematic representation of stimuli used in Pierrehumbert and Steele (1989).

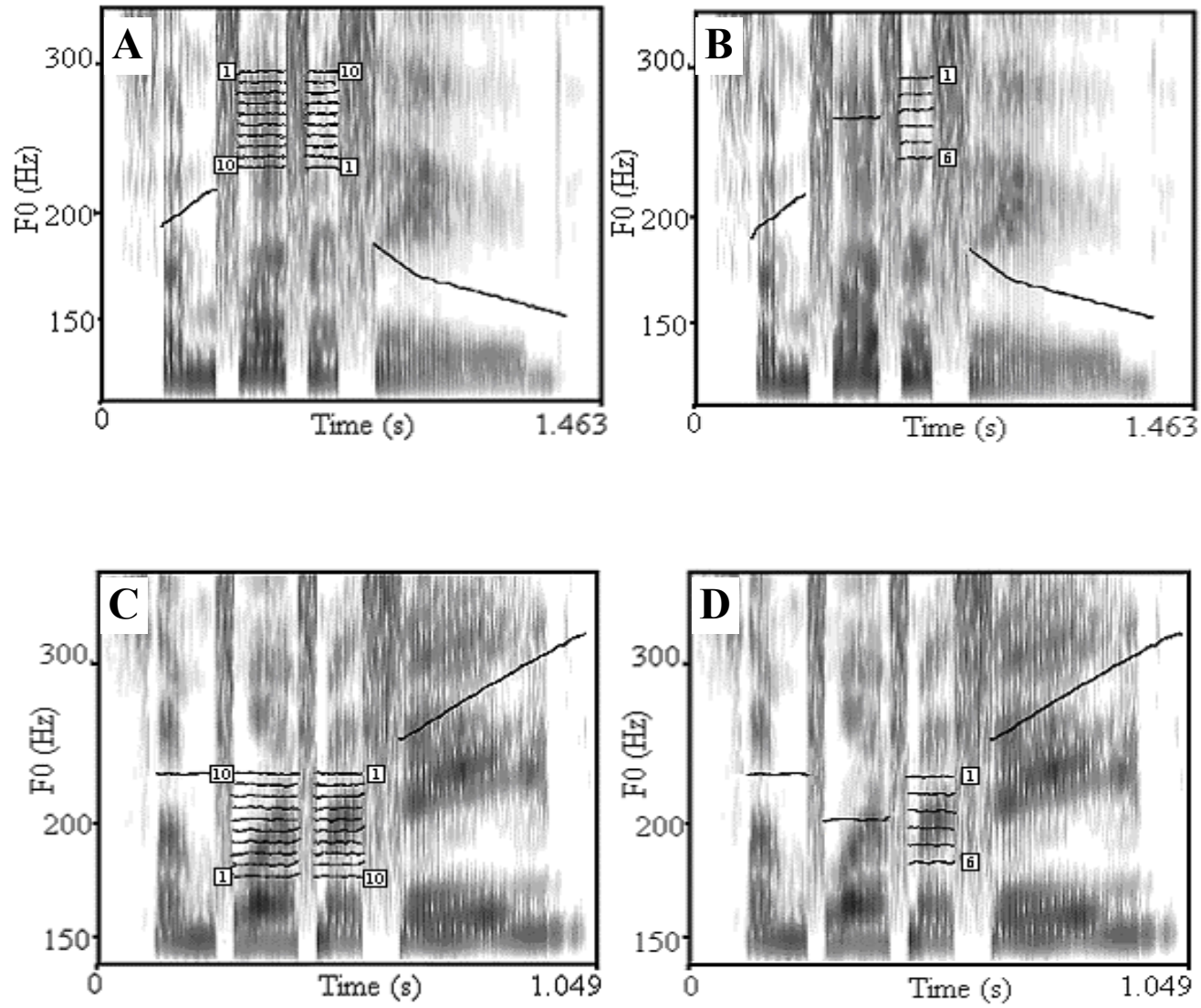


Figure 4. Stimulus series used in the imitation task.

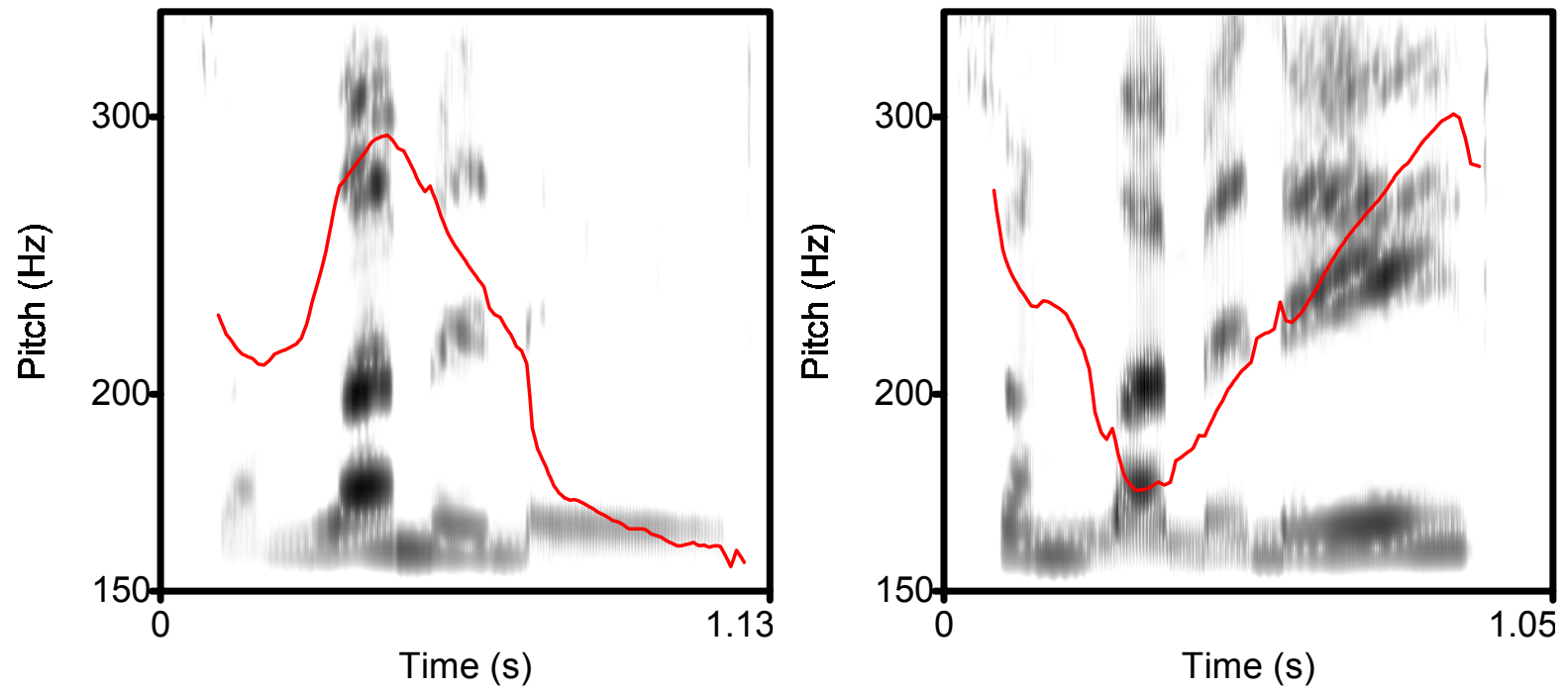


Figure 5. Typical F0 contours produced in response to Series A and B (left) and Series C and D (right).

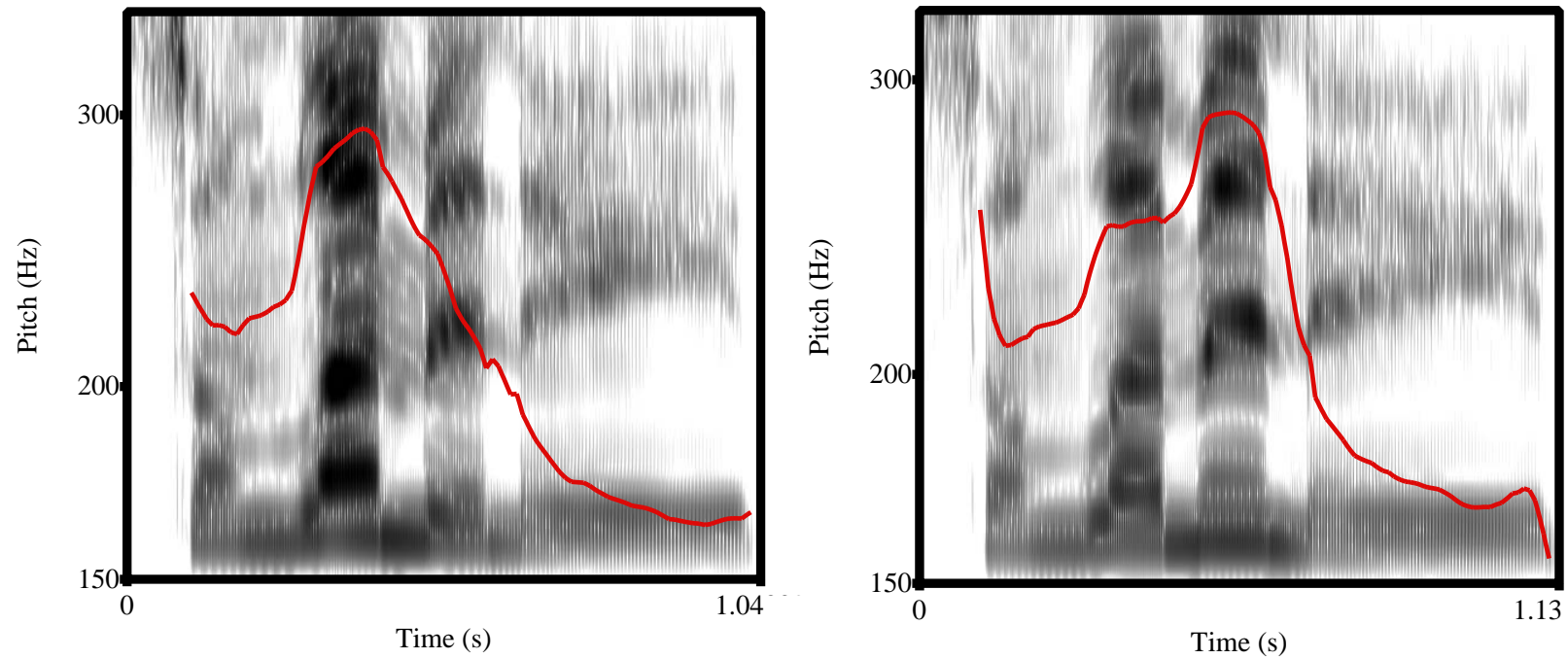


Figure 6. F0 contours produced by subject EL illustrating early peak alignment (left) and late peak alignment (right), in response to Series A stimuli 1 and 10, respectively.

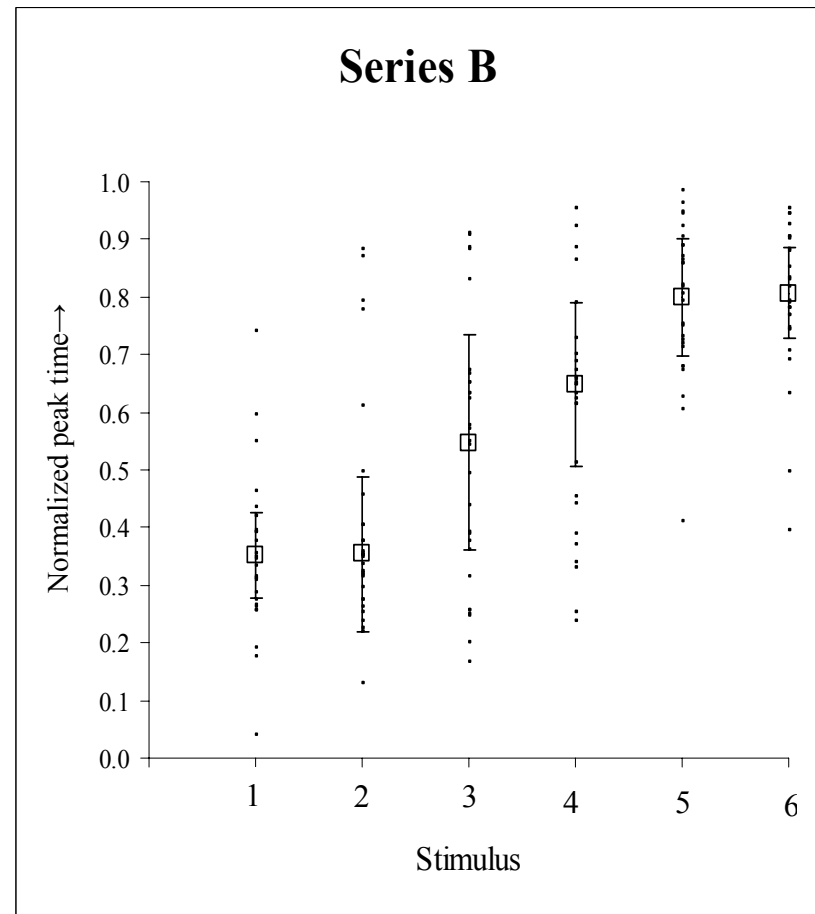
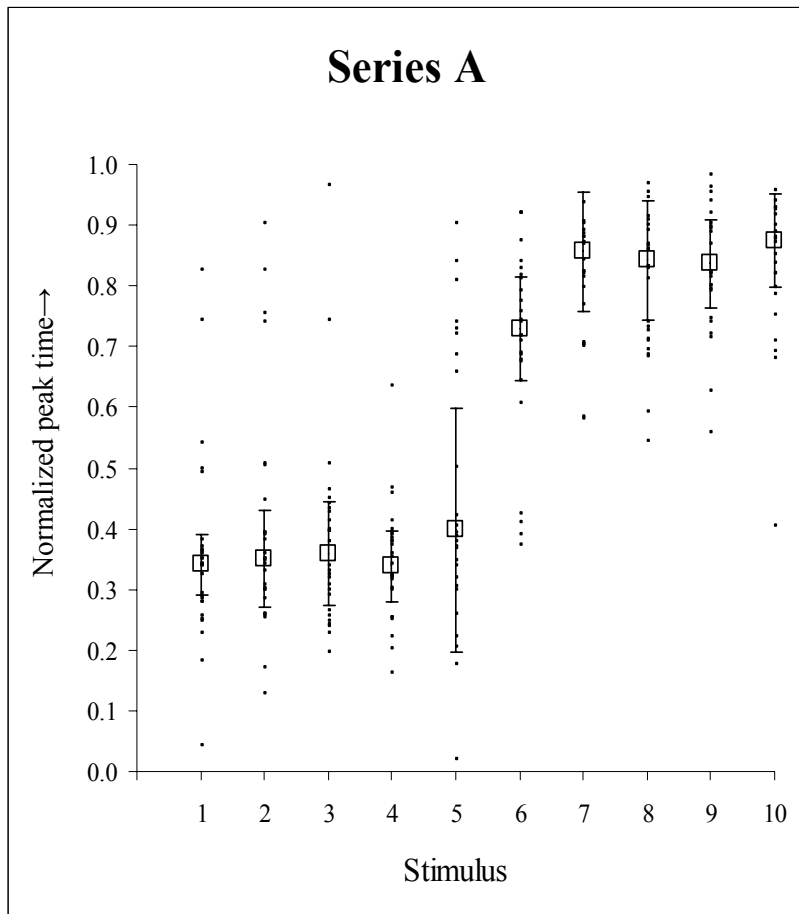


Figure 7. Normalized F0 peak times produced in response to Series A and B by subjects in the imitation task. Boxes are median values, and whiskers are semi-interquartile ranges.

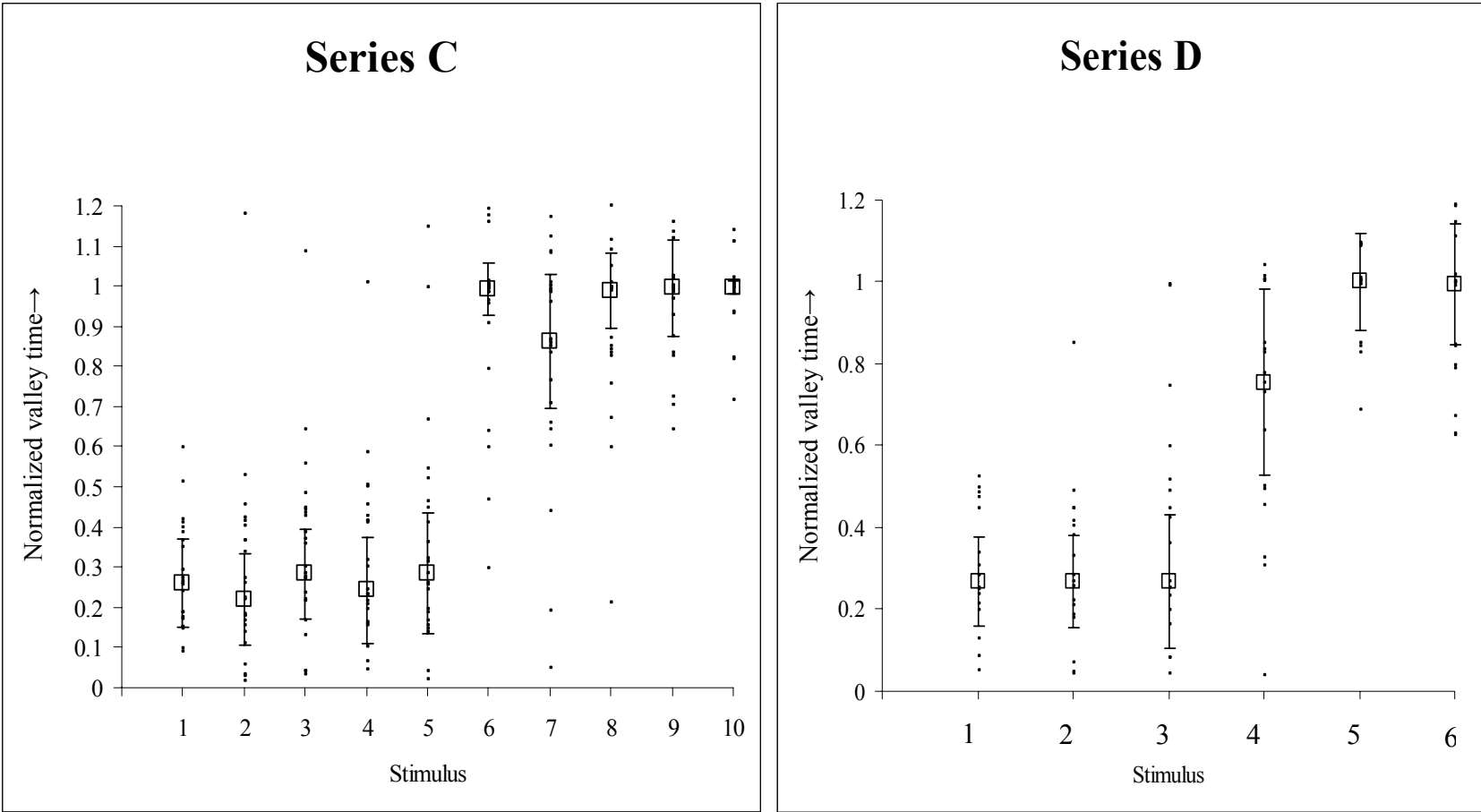


Figure 8. Normalized F0 valley times produced in response to Series C and D by subjects in the imitation task. Boxes are median values, and whiskers are semi-interquartile ranges.