Research Article

Suyeon Im*, José Ignacio Hualde and Jennifer Cole

On the salience of prenuclear accents: evidence from an imitation study

Abstract: Whereas some authors claim that the distribution of prenuclear accents in English largely follows from rhythmic and other non-informational considerations, other authors report a small but meaningful effect of prenuclear accents on the interpretation of sentences. In this paper we report on an experiment where native English speakers were asked to repeat stimulus sentences with one of three different accentual patterns on a word in sentence-initial prenuclear position: unaccented, with a high pitch accent on the syllable with primary stress or with a high accent on an earlier syllable with secondary stress. Participants were moderately successful in reproducing the intonational patterns. The early high accent pattern was reproduced particularly well. An automatic classification algorithm nevertheless produced four clusters of contours, instead of the three patterns present in the stimuli. Two distinct contours were used to signal the presence of a high tone before the syllable with primary stress. We conclude that the early high accent pattern is a strong attractor in imitations, but it was implemented with F0 trajectories that would be analyzed as phonologically different, suggesting an equivalence class of prenuclear contours. We also note a preference for rhythmic anchoring in the prenuclear position.

Keywords: pitch accent; prenuclear accent; secondary stress; stress shift; imitation

*Corresponding author: Suyeon Im, Department of English Language and Literature, Soongsil University, 369, Sangdo-ro, Dongjak-gu 06978, Seoul, Republic of Korea, E-mail: sim@ssu.ac.kr
José Ignacio Hualde, Department of Linguistics, University of Illinois at Urbana-Champaign, Urbana, IL, USA, E-mail: jihualde@illinois.edu
Jennifer Cole, Department of Linguistics, Northwestern University, Evanston, IL, USA, E-mail: jennifer.cole1@northwestern.edu

1 Introduction

In English, as in other languages without lexical tone, the pitch contours that extend over entire utterances vary in complexity, with some contours remaining relatively flat or gently sloping, and others containing one or more pitch excursions, each extending over one or more syllables. It is generally agreed that pitch patterning at the phrase level is meaningful and conveys pragmatic information (e.g., distinguishing new and given information, contrastive focus, and speech acts), but not every measurable characteristic of a pitch contour conveys linguistic meaning. Linguistic meaning conveyed through pitch is understood to reside in the sparse specification of intonational features.[1] In the Autosegmental-Metrical model of intonational analysis, these intonational features consist of tones (e.g., High, Low, or their combination) that mark prosodic boundaries, which frequently correspond to syntactic or discourse boundaries, and tones that mark prosodic prominence (Beckman and Pierrehumbert 1986; Ladd 2008; Pierrehumbert 1980). This paper focuses on the intonational features that mark prosodic prominence—*pitch accents*.

  In English, a pitch accent is assigned to the word that is the structural head, or nucleus, of the prosodic phrase (Chafe 1987; Pierrehumbert 1980; Selkirk 1995).[2] The default location for the nuclear pitch accent is on the rightmost content word of the sentence (1a, nuclear accent in CAPS), but earlier placement occurs on a word that has semantic focus (1b), or to avoid placing a pitch accent on a word that is lexically or referentially given (1c, adapted from Baumann and Riester 2013).[3] The nuclear pitch accent is, by default, the rightmost and therefore final pitch accent in the prosodic phrase. Consequently, the location of this pitch accent signals information about focus and the information status of the words in a sentence. Prenuclear pitch accents are optional on words preceding the nucleus (1d, prenuclear accents underlined),

---

1 See Cole (2015), Prieto (2015), Wagner and Watson (2010) for reviews of the extensive literature behind this claim.
2 A different view of accent assignment is that of Gussenhoven (2011, 2015), wherein pitch accents are assigned by default to every lexical stress, and a set of accent deletion rules remove accents e.g., in prenuclear position, or on words whose referents are "given" (in the common ground). What is important for our purpose in this paper is that in Gussenhoven's account, as in other analyses within the AM framework, pitch accents on words preceding the nucleus are optional. For Gussenhoven, the preservation of prenuclear accents is governed primarily by phonological constraints on accent clash, with a marginal role, if any, for information status.
3 Baumann and Riester's (2012) account of accenting follows a long line of work on the prosodic encoding of focus and information status, starting with Halliday (1967) and Bolinger (1972), and including later influential contributions by, among others, Beaver and Velleman (2011), Büring (2006), Gussenhoven (1984), Ladd (1980), Rooth (1992), and Selkirk (2007).

where they appear to serve primarily as an expression of phrase-level rhythmic stress (Calhoun 2010; Chodroff and Cole 2018).

(1) a. Sam was afraid of the DOG.
  b. {Speaker A: Were you afraid of the dog?}
      Speaker B: SAM was afraid of the dog.
  c. On his way home, a dog barked at Sam. He was AFRAID of the dog.
  d. <u>Sam</u> was <u>afraid</u> of the DOG.

One challenge to the notion that prenuclear accents are merely ornamental (Büring 2007)[4] is the finding that prenuclear accents are more likely, or considered more acceptable, in sentences with broad focus on the VP, and are generally unacceptable in sentences where the nuclear accent marks narrow focus on an object or adjunct (Gussenhoven 1983). Bishop (2017) finds a weak effect of prenuclear accent on verbs for the interpretation of focus projection. An accentual pattern as in *He bórrowed the SALT*, with prenuclear accent on *borrowed*, is unexpected for listeners if narrow focus on *salt* is intended. On the other hand, lack of an accent on the verb appears to be compatible with both narrow focus on the object and focus on the entire verb phrase. Bishop (2017) concludes that prenuclear accents play some role in constraining listeners' interpretation of focus. However, this is not a strong constraint and is subject to variation among listeners.

In this paper, we examine the status of American English prenuclear pitch accents (henceforth, PPA) in the cognitive representation of heard speech. Given that PPAs are optional in production and have at best a weak function in signaling linguistic meaning, they may be considered as part of the intonational phonetic "detail." The question then arises, do listeners attend to this aspect of the intonational patterning of an utterance and assign PPAs a representation that is encoded in memory? Evidence from prior studies using a speech imitation paradigm suggest they do, though with mixed evidence. In one study, Italian speakers are shown to imitate non-contrastive variation in the (early vs. late) alignment of an accentual peak, converging to the dialectal variant of the model speaker (D'Imperio et al. 2014), and another study shows American English speakers imitate continuous variation in the F0 range of high/rising pitch accents from continua that span two pitch accent categories, H* to L+H*, and H* to L*+H (Dilley 2010). In both studies, stimuli were very short—consisting of two words with a single pitch accent (e.g., Italian: *Ci veníva?* "would he come by?" and English: *some oREgano*)—and participants were explicitly instructed to imitate each phrase as closely as possible. These findings suggest that under at least some task conditions (i.e., with explicit instructions to listen or imitate closely) phonetic detail related to a *nuclear* pitch accent is imitated,

---

4 See Frota (2003), Petrone and Niebuhr (2014) for works on PPA in languages other than English.

but leave unanswered the status of prenuclear accents. Rather different results are reported from imitation studies using more complex stimuli (longer sentences with two or more pitch accented words), which show that imitated productions more reliably reproduce contrastive properties of intonation, such as the presence or absence of pitch accent or the pitch accent category, than the non-contrastive, variable phonetic detail (Braun et al. 2006; Cole and Shattuck-Hufnagel 2011; Cole et al. 2023; German 2012; Goodhue and Wagner 2018; Petrone et al. 2021; Torreira and Grice 2018; see also Zahner-Ritter et al. 2022). Yet these studies do not analyze prenuclear accents separately from nuclear accents. Thus, we do not yet have specific evidence from imitation, or any other experimental paradigm, about how listeners perceive and encode pitch accents in prenuclear position.

We address the status of PPAs in Mainstream American English (MAE) through an analysis of intonation imitated from sentences with multiple pitch accents. We ask if imitators reproduce the prenuclear intonational pattern of a heard utterance in (i) distinguishing the presence versus absence of PPAs in the prenuclear region, and (ii) reproducing the specific PPA pattern of the stimulus. Based on the findings from Braun et al. (2006) and Cole and Shattuck-Hufnagel (2011) for English (British and American, respectively), we predict that the MAE speakers in our study will be more accurate in reproducing the contrastive presence versus absence of PPAs, but may be less accurate in reproducing non-contrastive properties of PPAs. To the extent that imitations do in fact capture non-contrastive properties, the experimental findings will lend evidence to theories of intonation that specify a phonological representation of PPAs as "ornamental" features, despite their marginal status in conveying linguistic meaning.

2 Methods

2.1 Experimental materials

In order to have controlled variation in PPA patterning, we restrict our study to PPAs in English words that have two possible landing sites for a pitch accent, presented in non-final position of a complex noun phrase in a complete sentence. In English, words where the syllable with primary stress is preceded by another syllable with secondary stress (2–1 words, e.g., *èlevátion, òptimístic*) allow a less frequent pronunciation with reversal of prominence (e.g., *èlevátion → élevàtion*), in which case their pattern becomes identical to that of 1–2 words, which have only one prominence pattern (e.g., *élevàtor, súpermàrket*—which crucially cannot be pronounced **èlevátor, *sùpermárket*). This perceptually salient stress reversal phenomenon manifests in production through a High-tone pitch accent, H*or L+H* in the ToBI

annotation system (see Beckman et al. 2005), associated with the syllable that bears secondary stress in the unmarked pronunciation of the word. We refer to this accent pattern as "Early High."

This phenomenon of stress reversal in 2–1 words has been described as particularly common when there is a following accented word within the phrase, where it would be motivated by rhythmic reasons; e.g., *Mississíppi → Míssissippi Ríver; thirtéen → thírteen mén; Chinése → Chínese lánguage* (see e.g., Liberman and Prince 1977, among others). Thus, Early High in lexical 2–1 words has been described as a strategy for resolving stress clash (Horne 1990; Nespor and Vogel 1989), where locating a pitch accent on the initial syllable of a 2–1 word increases its distance from a pitch accent in the following word. Other researchers argue that the Early High pattern arises not by retracting stress from the later syllable, but rather from the placement of an additional pitch accent on the initial syllable (Ross et al. 1992; Shattuck-Hufnagel 1988, 1992, 1995). Grabe and Warren (1995) conclude that stress shift (Early High) is mostly a perceptual phenomenon, driven by listeners' expectation of rhythmic stress alternation.

We conducted an imitation experiment using constructed sentence stimuli where words with a lexical 2–1 stress pattern were produced in the prenuclear position of a prosodic phrase, with one of three accentual patterns: accent on the lexically specified **primary** stress (2–1), accent on the initial syllable in the **Early High** pattern (1–2), and **unaccented** (but retaining the lexically specified 2–1 stress pattern). By varying the presence/absence and location of the PPA within target words, we could compare the imitation of different, non-contrastive accent patterns on the same word, in the same location within the sentence and within the prosodic phrase. Twelve words with a canonical 2–1 stress pattern were selected as the target words for analysis. All target words had a secondary stress on the initial syllable and primary stress on the third syllable from the beginning of the word, e.g., *rèalístic*. The complete list of target words is given in (2):

(2)   Target words
        realistic, systematic, Unitarian, adversarial, professorial, inspirational,
        regulation, editorial, categorical, supplementary, automatic, disappointing

For each of the target words in (2), we constructed a sentence where the target word was the first content word in a noun phrase in sentence initial position (target phrase). The complete list of experimental sentences is given in (3). The phrase containing the target word is in brackets for clarity:[5]

---

5 The primary stress in the second content word is located on the initial syllable for the majority of experimental sentences, with the exception of sentences g, j, k, and l. This variation among the experimental sentences may potentially influence the tonal realization of the target words. A

(3)   Experimental sentences
  a.   [The realistic story] included a few untrue elements about George
        Clooney's hometown.
  b.   [The systematic tutors] always give clear instructions that even a
        beginner could follow.
  c.   [The Unitarian journalist] tried to be impartial in political disputes.
  d.   [The adversarial prosecutor] was not successful in making friends at the
        office.
  e.   [The professorial fashion] was never even noticed by most of the
        students.
  f.   [The inspirational speech] bored Alice out of her mind.
  g.   [The regulation of child labor**]** did not please everyone.
  h.   [The editorial column] reflected mainstream political views.
  i.   [His categorical stance] on protecting endangered animals admits no
        counter-arguments.
  j.   [The supplementary details] were unnecessary and made for a boring
        read.
  k.   [The automatic potato peeler] was too expensive for Johnny to buy.
  l.   [The disappointing performance] was depressing for Sue and the whole
        group.

An adult female native speaker of American English (one of the authors) recorded each of these twelve sentences three times, each time using a different prosodic pattern on the target word, as described in Table 1 and illustrated with the target phrase in (3a).

These three patterns are illustrated in Figure 1. Notice that the second content word in the target phrase was always accented, although its peak is downstepped when the preceding word also carries an accent, i.e., in the Primary and Early High patterns. Notice also the presence of a pitch rise on the final syllable of the head noun, marking a prosodic boundary at the end of the target phrase. The target word is thus always in prenuclear position.

The speech materials were recorded in a sound-attenuated booth using a head-mounted microphone. To ensure that the three sound files for each sentence were identical except for the prosody of the target phrase, the three recorded versions of the target phrase were spliced onto the same sentence continuation. For instance, the

supplementary examination confirms that this difference in stress location does not substantially affect the tonal realization of the target words in the imitated productions. More information can be found in the Supplementary material 2. Furthermore, to account for potential item effects, we incorporated a random intercept for the experimental (stimulus) sentences in our statistical models.

Table 1: Prenuclear accent patterns in experimental stimuli. Syllables produced with a pitch accent are bolded and the syllable with the nuclear prominence (and with another pitch accent) is underlined.

| Accentual pattern	Target phrase	Description
Primary	The rea**lis**tic **sto**ry	The prenuclear word has a main tonal prominence on the lexically specified primary stressed syllable (always the third syllable), conveyed by a high tone pitch accent (H* or L+H*) and falling pitch to the end of the word.
Early High	The **re**alistic **sto**ry	The prenuclear word has a main tonal prominence on the lexically specified secondary stressed syllable (always the initial syllable), conveyed by a high tone pitch accent (H* or L+H*) on this syllable and falling pitch to the end of the word.
Unaccented	The realistic **sto**ry	The prenuclear word is unaccented, with flat low pitch over the entire word.
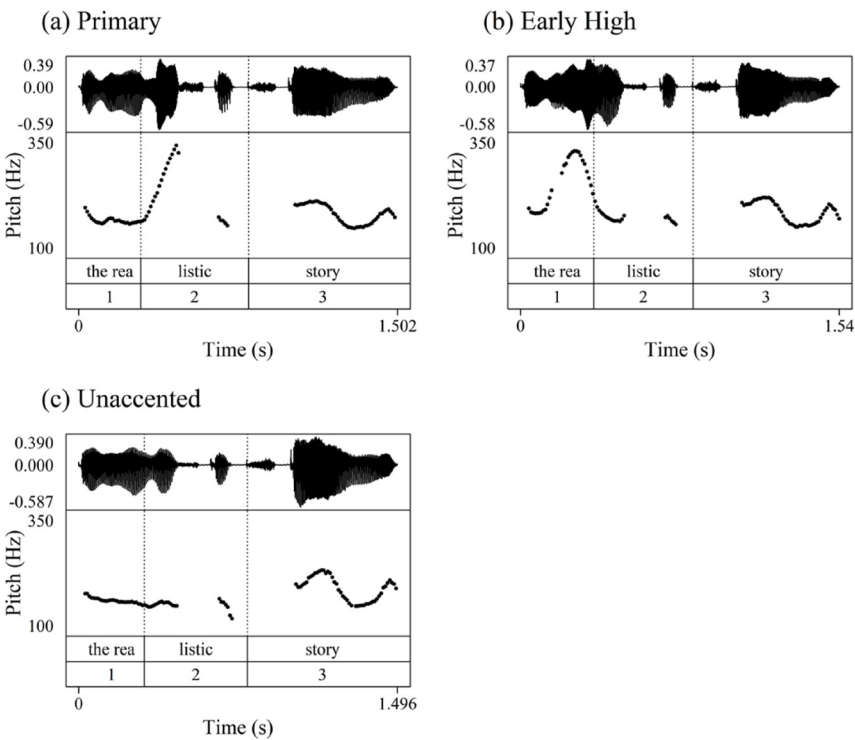
Figure 1: Examples of F0 trajectories over target phrases showing three accent patterns in stimuli: Primary (a), Early High (b), and Unaccented (c), as described in Table 1.

final set of experimental sound files contained three files for the stimulus sentences (3a), where each file started with a different recording of *the realistic story*, one pronounced with an Early High pattern (*the réalìstic stóry*), one with the Primary pattern (*the rèalístic stóry*) and a third one where the initial content word was unaccented (*the realistic stóry*). The remaining part of each file was the same single recording of the sentence continuation for all three files.

The final set contained 36 stimulus sound files (the 12 sentences in (3) each produced with the 3 prosodic patterns) and was divided into three lists of 12 sentences each, such that each subset contained all 12 sentences from (3) and comprised four sentences in each of the three prosodic patterns. Thus, each sentence was presented in only one prosodic pattern in each list.

2.2 Participants and experimental task

Thirty-three participants for this experiment were recruited from the student population of a large US public university. All participants were monolingual, native speakers of American English (23 female and 10 male). Participants were asked to imitate aural stimuli, with the instruction to repeat the sentence they heard "in the way the model speaker said it." The experiment took place in a phonetics laboratory. Stimuli were presented through headphones and participants also wore a head-mounted microphone. Each stimulus was presented two times in succession, in auditory format only (no text), after which participants were asked to repeat it. Participants were randomly assigned to one of the three stimulus lists, as defined in the preceding Section 2.1. The stimulus list was presented twice, resulting in two iterations of the imitation per participant. The initial iteration was subject to analysis, except in cases where speech errors occurred, in which case the second one was analyzed instead. The total duration of the experimental task was approximately 30 min per participant.

2.3 F0 measurement and modeling

A total of 396 imitated utterances were recorded (33 participants *12 utterances). Five imitations were excluded due to lexical or syntactic errors. The remaining 391 imitated utterances were subject to acoustic analysis using Praat (Boersma and Weenink 2020). The target phrase within each imitated utterance was manually segmented into three intervals. Interval 1 spanned from the start of the subject NP to the consonantal onset of the primary stressed syllable of the target word, interval 2 spanned from the end of interval 1 to the end of the prenuclear target word, and

interval 3 spanned from the end of interval 2 to the end of the head noun of the target phrase, as shown in (4) with numbers 1-3 marking the right edge of these intervals. Within each of the three intervals, we extracted the maximum and mean F0, manually checking for and avoiding consonantal perturbations:

(4)   Segmentation procedure
        [The rea $_1$| listic $_2$| story $_3$|] included …

The imitated utterances were classified as Early High, Unaccented or Primary based on the relative difference of F0 mean values[6] in each interval (1, 2, and 3), as in Table 2.[7] This classification was based only on the F0 pattern produced in the imitated utterance, regardless of the intonational pattern of the stimulus that was being imitated. The classification was done in three steps as follows. The imitation was classified as Early High if 1 > 2; otherwise, it was classified as Unaccented if 3 > 2; otherwise it was classified as Primary. By this method, an imitation was classified as Early High on the basis of the mean F0 in the first two intervals (1 and 2), with no restriction on the F0 mean in 3. Similarly, an imitation not identified as Early High in step 1 was classified as Unaccented based on the presence of an increase in mean F0 from the second to third interval, without further consideration of the first interval. The remaining contours were counted as Primary, which included imitations in which the F0 contour was rising over the sequence 1–2, in which case 2 ≥ 1, and those with a flat F0 over 1–2, in which case 1 = 2 (within 4 Hz, see Footnote 7).

After classifying the imitated utterances, the imitations of each prosodic pattern in the stimulus set were analyzed to determine how many were correctly imitated, and how many were imitated with either of the other two patterns in the classification scheme in Table 2. For example, this analysis calculated how many imitations of Early High stimuli were produced with an Early High pattern, a Primary pattern, and an Unaccented pattern based on the relative F0 mean (or max) values.[8]

The classification analysis just described has the obvious drawback that it assumes that the imitated productions are adequately classified as instantiating one

---

6 The classification analysis was repeated using the F0 maximum in each interval, instead of the mean F0, and the same classification criteria as shown in Table 2, with very similar results, as also noted in Section 3. The analysis is provided in the Supplementary material 2.

7 F0 measures had to differ by at least 4 Hz to be considered different for the purpose of this classification. The 4 Hz difference is slightly higher than published JND values for pitch (Zwicker and Fastl 2013) and was chosen based on the authors' auditory perception of noticeable differences in our data.

8 We feature the analysis using F0 mean values in this presentation, to allow a more comprehensive view of the overall tonal patterns (i.e., presence and position) while abstracting from localized pitch excursions that may arise as perturbation effects from adjacent voiceless consonants, which would not be representative of the broader prosodic structure under our examination.

Table 2: Classification of imitated F0 patterns, based on relative F0 maxima or mean values.

1 2 3	1 2 3	1 2 3 *or*	1 2 3
the rea | listic | story	the rea | listic | story	the rea | listic | story	the rea | listic | story
Early High	Unaccented	Primary	
1 > 2	3 > 2	2 ≥ 1	

of the three intended patterns in the auditory stimuli. It is possible, however, that the best classification of the results would group the imitated F0 contours using a different number of clusters. To explore the possibility that participants actually produced a smaller or greater number of distinct F0 patterns, and/or the possibility they produced distinctions that are not captured by our three predetermined classes, we performed two clustering analyses over the F0 trajectories in the imitated utterances, unlabeled for the accent pattern of the stimulus. The first clustering analysis was conducted over the series of three mean F0 measurements that were the basis for the classification analysis described above. These three F0 measures define a coarse-grained, simplified F0 contour. The second clustering analysis was conducted using the full F0 contours from each imitated production, time-normalized to 30 samples using ProsodyPro (Xu 2013), with 10 equidistant samples in each interval 1, 2, and 3 shown in (4). The total number of imitations entered into each of the analyses included 129 imitations of the Primary pattern, 131 imitations of the Early High pattern and 131 imitations of the Unaccented pattern (total = 391). All clustering analyses were conducted using the *k*-means clustering algorithm for longitudinal data with the kml package in R (Genolini et al. 2015).[9] As the kml analysis is based on the Euclidean distance between the mean of hypothesized clusters, we z-normalized

---

9 *K*-means clustering for longitudinal data is an exploratory technique based on a non-parametric algorithm that considers all the data points of each subject as a single trajectory and partitions the trajectories into the optimal number of clusters. The kml function in R initially assigns each trajectory to a cluster at random and computes the Euclidean distance (with Gower adjustment) between the center of each cluster, reassigning trajectories until no further change occurs in the clusters. kml tests 2, 3, 4, 5, and 6 clusters, 20 times each. The relative adequacy of each clustering is determined by the Calinski-Harabatz criterion, which optimizes for solutions with clusters that are both compact and well separated from one other.

subjects' F0 values using the *scale* function in R (R Core Team 2024) to reduce differences due to gender and individual variation in pitch range.

The clustering analyses report the optimal number of clusters for the imitated F0 trajectories (modeled as time-series data), and the mean values at each time point for each reported cluster. We examined the optimal clusters from each analysis in relation to the F0 contours of the three PPA patterns, comparing the number of clusters and the mean F0 contour shape of each cluster.

2.4 Predictions

To the extent that the presence of a PPA is potentially informative about information structure, pertaining either to the discourse referent of the prenuclear word or of a downstream word (as shown by Bishop [2017]), we predict that participants will reproduce a PPA when it occurs on the target prenuclear word in the stimulus. Specifically, we predict that imitations of Unaccented stimuli will have no PPA on the target word, while imitations of stimuli with Primary and Early High PPA patterns will realize a PPA on the target word. As for the distinction between the Primary and Early High accent patterns, there is a weaker prediction. Previous accounts of the Early High pattern (cited in Section 2.1) make no claim about a meaning distinction between the Early High and Primary accent patterns for 2–1 words, suggesting that this distinction may truly be "ornamental" intonational detail. In light of findings from prior imitation studies discussed in Section 1 showing that memory for noncontrastive acoustic detail of pitch accents is less reliable than memory for contrastive accent specification, we predict that the distinction between the Early High and Primary accent patterns will be less accurately reproduced in imitations than the distinction between either of these and the Unaccented pattern.

3 Results

We report first on the distribution of the imitated F0 contours based on their classification into one of the three pre-determined accent groups: Early High, Primary and Unaccented. This is followed by the results of the classification that results from the clustering analysis described in the previous Section 2.3, which does not assume the three pre-defined accent groups.

3.1 Classification analysis

As described above, F0 mean values were extracted from the three intervals 1–3 for each target phrase, as defined in (4).[10] The imitations were classified into three accent patterns based on the relative F0 mean values in the three intervals. Figure 2 shows the number of imitated utterances classified in each of the three accent patterns. A striking finding is the prevalence of the Early High pattern in the imitated productions, which accounts for fully 61 % of the imitated productions.

Next, we look at differences in F0 means between intervals 1 and 2 for each utterance produced by our subjects, classified according to the criteria in Table 2 above, regardless of stimulus pattern. As can be seen in Figure 3, in tokens classified as presenting a Primary pattern, the difference in F0 between intervals 1–2 tends to be negative; that is, the mean F0 in interval 2 is higher, indicating the presence of a high-tone pitch accent in the second foot of the target word. In production tokens that we have classified as Early High, instead, the mean F0 in interval 1 is higher than in interval 2. Finally, in tokens classified as Unaccented, differences in the F0 mean between the two feet of the target word cluster around zero.
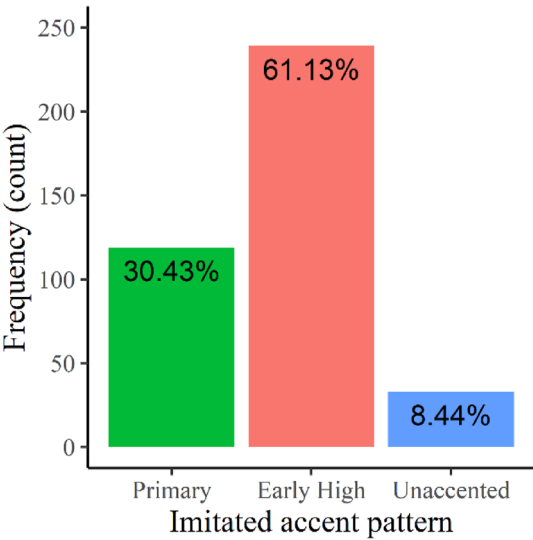

Figure 2: Classification of imitated productions into three accent patterns based on the relationship between F0 means in three measurement intervals as described in Table 2. (The term "imitated accent pattern," as depicted on the x-axis of Figure 2, will be employed consistently throughout the paper to denote the participants' imitations, which in this figure were categorized into one of three accent patterns in accordance with the criteria outlined in Table 2.)

---

10 A parallel classification analysis using F0 maxima values from each interval was also conducted, with results that were substantially the same as those reported here with classification based on F0 mean values in each interval. The analysis using the F0 maximum can be found in the Supplementary material 2.
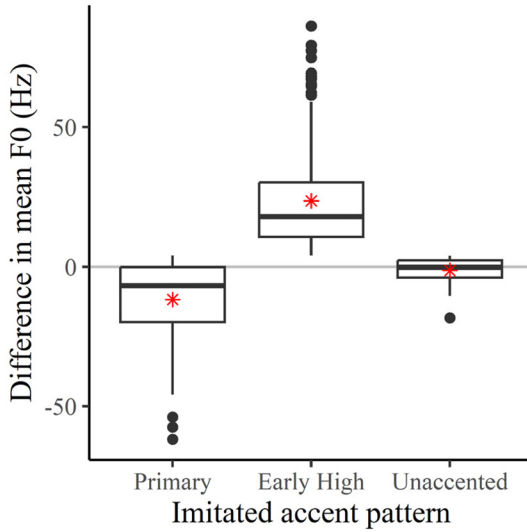
Figure 3: Difference in mean F0 between intervals 1 and 2 for imitated productions classified according to the criteria in Table 2 (*x*-axis). The red asterisk marks the mean of each imitated productions classified by the criteria in Table 2.

A linear mixed-effects regression was run in R (R Core Team 2024) with the packages *lme4* (Bates et al. 2015) modeling variation in the difference in mean F0 between intervals 1 and 2 (1–2), with Accent pattern classified as in Table 2 (3 levels: Primary, Early High, Unaccented), participant self-reported Gender (2 levels: female, male) and participant Group (3 levels: A, B, C) as fixed factors, and random intercepts for Participant and Utterance. Significant effects were evaluated based on *p*-values calculated using Satterthwaite's method for degrees of freedom. Significant differences were found for Accent pattern when Early High was compared to the other two patterns (Primary: $\beta = -27.81$, $SE = 4.33$, t [268.55] = -6.42, $p < 0.001$; Unaccented: $\beta = -21.63$, $SE = 7.76$, t [360.61] = -2.79, $p < 0.01$). No significant effects of Gender (F [1, 79.07] = 0.01, *n.s.*) or Group (F [2, 79.79] = 0.05, *n.s.*) were found.

Turning now to the difference in mean F0 between intervals 2 and 3, this comparison clearly distinguishes Unaccented productions from the other two, as can be seen in Figure 4. In utterances where the target word is unaccented, the pitch accent in interval 3 is not downstepped, since there is no preceding pitch accent, and the F0 mean of interval 3 is substantially greater than that of interval 2 (yielding a negative difference for the subtraction 2–3). A linear mixed-effects regression on the difference in mean F0 between regions 2 and 3, with the same factor structure as above, returns a significant difference between Unaccented and each of the other two Accent patterns (Primary: $\beta = 30.65$, $SE = 8.54$, t [355.85] = 3.59, $p < 0.001$; Early High: $\beta = 17.28$, $SE = 8.44$, t [346.36] = 2.05, $p < 0.05$). There were no significant effects of Gender (F [1, 40.86] = 1.72, *n.s.*) and Group (F [2, 41.19] = 0.12, *n.s.*).
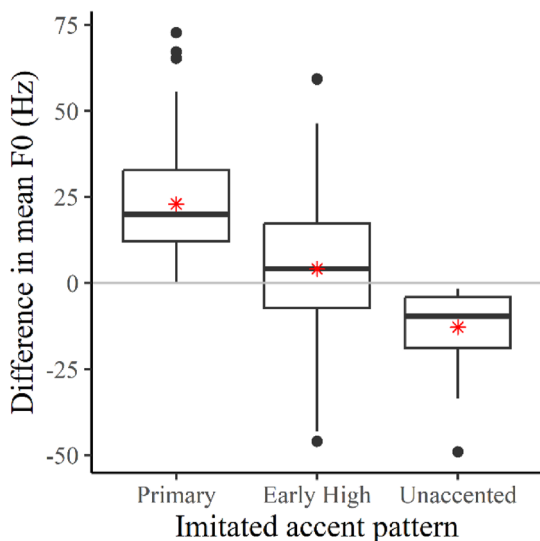
Figure 4: Difference in mean F0 in intervals 2 and 3 (see Table 2) for imitated productions (*x*-axis). The red asterisk indicates the mean of each imitated productions classified by the criteria in Table 2.

The classification of participants' productions in three groups using the criteria in Table 2 thus results in three statistically distinguishable groups of data, differentiated by the F0 pattern over the prenuclear and nuclear accents of the complex subject noun phrase. This is unsurprising, since the classification of imitations into the three accent pattern categories was based on observed differences in F0 across the three intervals in the target region.

Next, we compare the classification of each imitated production with the accent pattern of its corresponding stimulus. What we find is that Early High stimuli were imitated with high accuracy across all productions (85 % accuracy), with lower accuracy for Primary (60 %) and Unaccented (15 %) stimuli (Figure 5). Figure 5 also shows the breakdown of incorrect imitations into the three designated accent patterns. As can be seen, Early High was also the most frequently produced "error" for both Primary and Unaccented stimuli. These errors are the basis for the high proportion of Early High patterns in the imitated productions, as shown in Figure 2. In other words, our participants' clear preference for an early prenuclear accent resulted in most Early High stimuli being accurately imitated, but also in most Unaccented stimuli being imitated with Early High and a sizable proportion of Primary stimuli also being produced with Early High.

The analysis of the results that we have offered in this section has the drawback that it focuses exclusively on the relative mean F0 across three intervals, necessarily classifying all production in one of three groups. In the next section, we present the results of an automatic procedure that takes into account additional aspects of the intonational contour and allows for classification based on more detailed F0 trajectories.
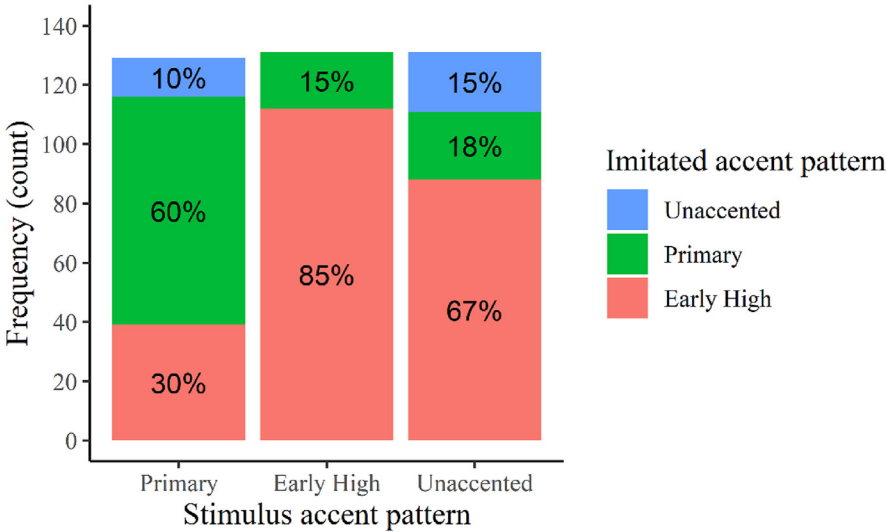
Figure 5: Imitations grouped by stimulus accent pattern (*x*-axis). Colored bars indicate the accent patterns of the imitated production according to the relative F0 mean values in intervals 1–3 as described in Table 2. (The term "stimulus accent pattern" will be consistently employed throughout the remainder of this paper to designate the accent patterns of the auditorily presented stimulus sentences that participants were instructed to imitate. Figure 5 plots imitated productions grouped by the accent pattern of the stimulus that was the intended target of imitation.)

3.2 Cluster analysis

Cluster analysis offers a data-driven exploration of the distinctions in F0 trajectories that participants produced in their imitations, without invoking *a priori* assumptions on the number or type of distinctions. In the first cluster analysis that we conducted, the mean F0 values in the three intervals of each production were submitted as a 3-point trajectory to *k*-means clustering.

Figure 6 shows that the 3-point F0 trajectories of the imitations are optimally categorized into four clusters using this algorithm. Three of the clusters obtained from this analysis can be identified with the accent patterns present in the stimuli: Early High (cluster A in Figure 6), Primary (cluster C), and Unaccented (cluster D). The analysis also returns a cluster where the F0 mean decreases, suggesting a pattern of downstep across peaks in the first and second interval (cluster B), a pattern that has characteristics of both the Early High and Primary accent patterns. Among these four clusters, cluster A (Early High) is the most common, accounting for 30 % of the data, followed by cluster B (Early High + Primary), which accounts for an additional 29 % of the data. Cluster C (Primary) and cluster D (Unaccented) represent 23 % and 18 % of the
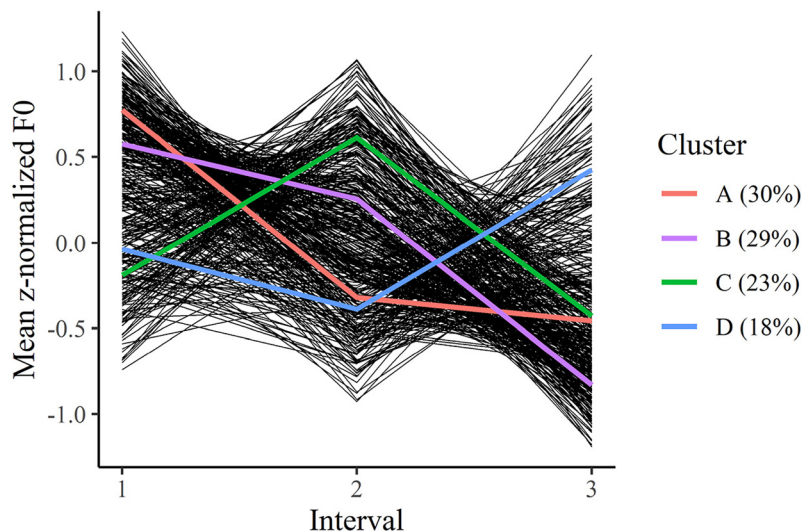
Figure 6: Output of kml cluster analysis over trajectories of three F0 mean values from intervals 1–3 of the imitated target phrases. Colored lines show the mean trajectories for each of the four clusters in the optimal clustering solution. Black lines are the input trajectories.

data, respectively. Combined, clusters A and B, where the highest mean F0 patterns occurs in the first interval, account for almost 60 % of all patterns produced in our imitation experiment, which is consistent with the results above in Figure 2.

Additional cluster analyses further explore the distinctions in the imitated F0 trajectories, now with each trajectory submitted as a time series of 30 F0 samples drawn from the time-normalized F0 trajectory of the target phrase. The purpose of these analyses was to examine variation in F0 contour shape beyond the coarse-grained analysis based on the mean F0 in three intervals, presented above. We report first on a cluster analysis performed over the F0 trajectories in the target region of all the imitated utterances, with no grouping imposed. The optimal solution has only two clusters, shown in Figure 7. The mean contour of cluster A has an early peak (between time points 5–10, in the first interval) followed by a fall in F0 to a lower peak in the third interval (around time point 25), corresponding to the head noun of the target phrase, while the mean contour of cluster B has a peak in the second interval (between time points 10–15) followed by a rise to a higher peak near the end of the phrase (around time point 25, in interval 3). The mean contour of cluster A, which includes 56 % of the imitated contours, resembles the Early High contour of the stimuli (Figure 1b), while that of cluster B, with 44 % of the imitated contours, resembles a combination of the Primary and Unaccented contours of the stimuli (Figure 1a and 1c). These proportions are similar to the proportions from the classification analysis presented above (Figure 2)
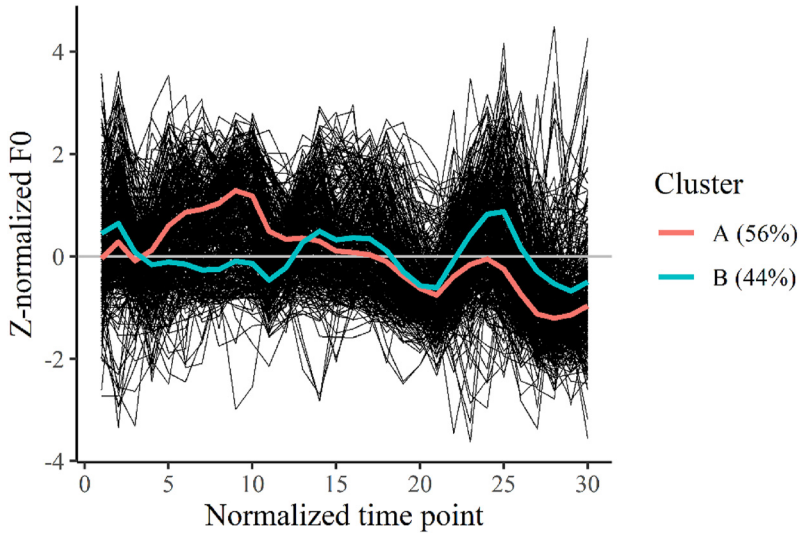
Figure 7: Output of kml cluster analysis over trajectories of 30 time-normalized F0 values from target phrases of all imitated productions. Colored lines show the mean trajectories for the two clusters in the optimal clustering solution. Black lines are the input trajectories.

using the three pre-determined accent patterns of the stimuli, which yielded 61 % of imitations classified as Early High and the remaining 39 % split between Primary and Unaccented.

Three further cluster analyses were performed using the same 30-point F0 trajectories on subsets of the imitated productions grouped by stimulus accent pattern. The purpose of these analyses was to determine if there are regular patterns of variation in the imitated F0 contours for each of the stimulus accent patterns. The first observation from these subset analyses is that, for each stimulus accent type, the optimal clustering solution had two clusters, shown side-by-side in each row of Figure 8, of which one, arbitrarily labeled "A," has the expected F0 trajectory shape corresponding to the stimulus accent type. For imitations of the Primary accent pattern, those in cluster A have the largest F0 excursion and highest peak in the second interval and a second, downstepped peak in the third interval (e.g., "*rea | **listic** | **story***," with F0 peaks in bolded intervals, and underline marking the highest such peak). For imitations of the Early High accent pattern, the imitations in cluster A have a peak in the first interval and a much weaker peak in the third (e.g., "**_rea_** *| listic | **story***"). Finally, imitations of the Unaccented pattern in cluster A have only one peak, in the final interval (e.g., "*rea | listic | **story***"). The A cluster includes a slim majority of the imitations for each accent type: 56 % of Primary, 53 % of Early High, 53 % of Unaccented.
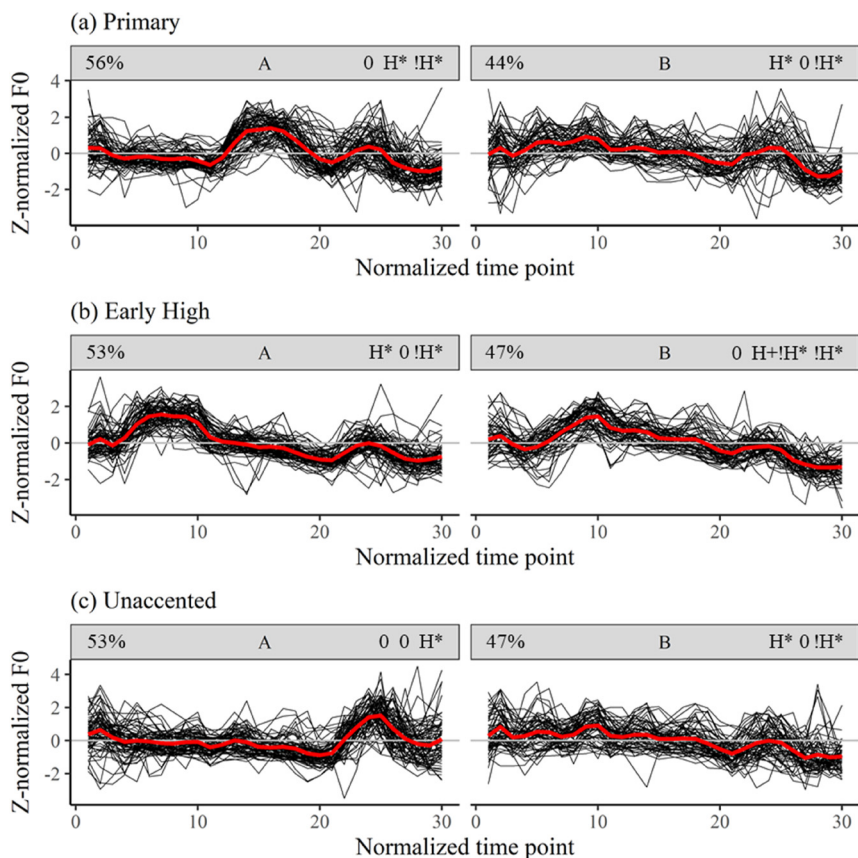
Figure 8: Output of kml cluster analyses over trajectories of 30 time-normalized F0 values from target phrases of imitated productions subsetted by stimulus accent pattern: Primary (a), Early High (b), and Unaccented (c). The optimal clustering of F0 trajectories in each group has two clusters, displayed separately in side-by-side graphs in each row. The red line marks the mean trajectory of each cluster, and black lines are the input trajectories.

The imitations grouped in the "B" cluster for each accent type do not conform to the expected F0 trajectories. Instead, considering the mean trajectory of each B cluster, it appears that the imitations in the B clusters don't uniquely resemble any one of the three stimulus accent patterns. All three B cluster patterns show a weak F0 peak in the third interval. In addition, F0 is slightly elevated across the first and second intervals in the Primary and Unaccented data, while the B cluster pattern for the Early High data show a noticeable late peak at the very end of the first interval. An example pitch track of the B cluster pattern for the Early High in imitated productions is shown in Figure 9. For all three B cluster mean trajectories, the mean
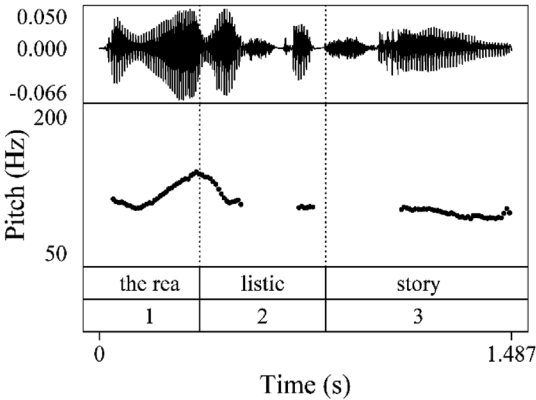
Figure 9: Example pitch track of cluster B for Early High in imitated productions.

F0 of the first interval is higher than that of the second interval, which is the likely source of the high number of imitations classified as Early High based on the criterion or relative mean F0 (shown in Figure 6). Qualitatively, the mean trajectories of the B clusters can be described with reference to the Early High pattern (as shown in cluster A of the Early High data). Imitations in the B clusters of the Primary and Unaccented data resemble the Early High pattern realized with a damped peak in the first interval. Imitations in the B cluster of the Early High data suggest something different—where the peak at the end of the first interval may be interpreted as the leading tone of a falling accent (H+!H*) on the primary stressed syllable. In other words, the imitations in cluster B of Early High resemble the Primary accent pattern in having a pitch accent on the primary stressed syllable (e.g., *"rea | lis̲t̲i̲c̲ | story*," with falling pitch over *"lis"*) followed by a downstepped !H* accent (*"story"*). Henceforth we draw on this qualitative interpretation of the cluster B pattern of the Early High data and refer to this as the "H+!H*" pattern.

The observation in Figure 8 of two F0 patterns in imitations of Early High stimuli was based on the clustering analysis of 30 time-normalized F0 trajectories from the target word. Above we suggested an interpretation of the two F0 patterns in terms of the location of the pitch accent on the secondary-stress syllable for Cluster A, and on the primary-stress syllable in Cluster B. In principle, this difference in pitch accent location could also manifest in durational cues of the target word in imitation. We thus carried out further analysis to examine duration (milliseconds) of two intervals 1–2 from the target phrases of imitated productions. Figure 10 shows subsets of imitated productions categorized by stimulus accent pattern. Two clusters of duration for each subset of imitated productions are grouped by the output of cluster analysis of F0 contours described in Figure 8 above.

From the subsets of imitated productions grouped by Early High (Figure 10b), we see that the relationship between intervals 1 and 2 is clearly different between
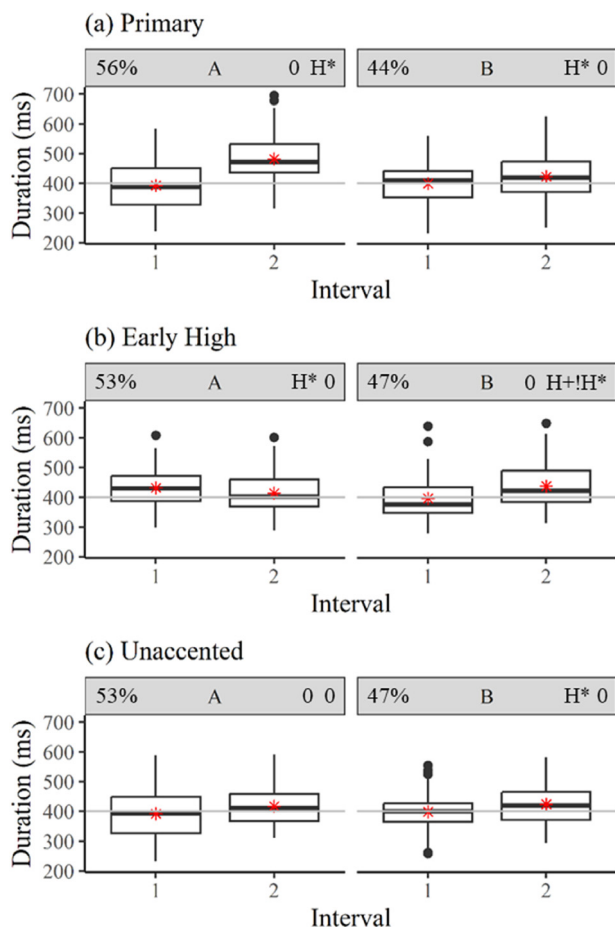
Figure 10: Duration of the first two intervals (1–2) from target phrases of imitated productions, grouped by stimulus accent pattern: Primary (a), Early High (b), and Unaccented (c). This is further categorized by the output of kml cluster analysis presented in Figure 8. The red asterisk marks the mean duration of each interval.

clusters A and B, which correspond to the contour with H* tone on the secondarily-stressed syllable and the contour with H+!H* on the primarily-stressed syllable in Figure 8, respectively. Whereas interval 1 was longer than interval 2 for A cluster, interval 2 was longer than interval 1 for B cluster. This visual impression was further confirmed by statistical analyses using mancova in R (R Core Team 2024). There was a significant difference between clusters A and B in the Early High pattern (holding constant participants and utterances) on the combined intervals 1, 2, and 3 (F [3, 86] = 15.95, $p < 0.001$). Pairwise comparisons between the clusters were significant for

intervals 1 ($p < 0.001$) and 2 ($p < 0.005$), but not for 3 ($p = 0.2$), suggesting that the duration of two clusters differed in intervals 1 and 2, but not in 3. Overall, these results are in alignment with those from the cluster analysis of F0 contours that imitation of Early High had two distinct patterns, i.e., the contour with H* on the secondarily-stressed syllable and a contour with an early F0 peak, which we suggest may be analyzed as H+!H* on the primary-stressed syllable.

A detailed examination of the F0 trajectories in the B clusters reveals additional features. As described above, visual assessment of the mean F0 contour of the B clusters in the Primary and Unaccented subsets finds a first peak in the end of interval 1 (Figure 8). These patterns for Primary and Unaccented trajectories in the B cluster differ from the cluster B pattern of the Early High subset, where there is a more pronounced F0 peak (i.e., higher peak, aligned later) in the end of interval 1, which may be interpreted as H+!H* anchored on the primarily-stressed syllable (in interval 2). Our duration analysis in Figure 10 supports this interpretation. The cluster B imitations for all three accent types have longer duration for interval 2 than interval 1, but this tendency is greater for the Early High subset (Figure 10b) compared to the Primary (Figure 10a) and Unaccented (Figure 10c) subsets. This visual observation was confirmed by statistical analysis using mancova with the same structure described above. Pairwise comparisons found significant differences between clusters B in Early High and the other two accent patterns on interval 2 (Primary: $p < 0.05$, Unaccented: $p = 0.05$), but not on interval 1 (Primary: $p = 0.61$, Unaccented: $p = 0.77$), suggesting that Early High differed from the other two accent patterns in the duration of interval 2, but not interval 1, for imitations grouped in cluster B. Pairwise comparisons also revealed non-significant differences between the B clusters of Primary and Unaccented on intervals 1 ($p = 0.85$) and 2 ($p = 0.87$), showing that Primary and Unaccented have similar duration in intervals 1 and 2 in these data.

Together with the results from cluster analysis of F0 contours, the duration results suggest that the imitations grouped into B clusters have a similar prominence pattern, with the exception that for the B cluster imitations of Early High, the primary stressed syllable in interval 2 has greater pitch and durational prominence than the corresponding syllables in interval 2 for the Primary and Unaccented data. The overall similarity across all B cluster imitations suggests that Early High is persistently activated in all imitated productions. We suggest that the differences between the B cluster imitations of each accent pattern is that, while all three B cluster patterns show an influence of the Early High pattern (i.e., a prominence asymmetry between intervals, 1 > 2), that influence appears stronger in the B cluster imitations of Early High stimuli. In that panel (8b, right), we observe the expected Early High asymmetry in F0 prominence (1 > 2), though now reduced due to the elevated F0 in interval 2 (compare to panel 8b, left). The influence of the Early High pattern on "incorrect" imitations of Primary and Unaccented stimuli (panels 8a and

8c, right), show little if any asymmetry in the relative prominence of intervals 1 and 2, considering both duration and F0. In other words, the "incorrect" imitations in the B clusters appear similar in being somewhat weak versions of Early High, i.e., with a weaker 1 > 2 prominence asymmetry that is characteristic of Early High.

To summarize, the subset clustering analyses for imitations of each stimulus accent type explains several key findings. First, the numerous imitations classified as Early High based on the relative F0 values over the three analysis intervals (Figure 5) have a likely source in the "inaccurate" (B cluster) imitations of the three stimulus accent types. Second, the appearance of a fourth F0 trajectory in the clustering analysis over the aggregated three-interval mean F0 values (cluster B in Figure 6), with downstep over the first and second intervals and lower F0 in the third interval is compatible with imitations in cluster B of the Early High subset (Figure 8b). Third, the B clusters of the subset analyses show two variants of the Early High—one where the "early" pitch accent in the first interval has a damped peak, and another where the F0 peak in the first interval seems to mark the start of a falling pitch accent in the second interval. Viewed in this way, we suggest that these imitations may have undergone a phonological re-analysis, mapping the Early High pattern of the stimulus onto a Primary pattern, but using a falling pitch accent rather than the rising accent of the Primary and Early High stimuli. Notably, this falling accent pattern with a late peak in the first interval was not represented in the stimuli.

4 Discussion

The question that we wanted to address in this study is to what extent listeners pay attention to the presence and position of prenuclear pitch accents. This issue is of interest because, whereas differences in the position and shape of the nuclear accent of the phrase have been associated with different pragmatic meanings, the distribution of prenuclear accents is less clearly meaningful. We reasoned that if speakers assign meaning to the distribution of prenuclear accents in the phrase, they might then imitate contours with different distributions of prenuclear accents, and conversely, if prenuclear accents are not considered meaningful, speakers may pay them less attention and less reliably include them in a linguistic representation of the heard utterance. For comparison, we expect English speakers to successfully perceive and be able to reproduce the meaningful contrast in VOT between word-initial /p/ and /b/ found in their language in words like *pit* and *bit* and to be less accurate in perceiving and reproducing a subphonemic contrast in this language such as the difference between prevoiced and non-prevoiced tokens of *bit*.

In our experiment, we used words with two possible anchors for a pitch accent in prenuclear position in a prosodic phrase, and participants were presented stimuli

with three distinct patterns of prenuclear pitch accents. They were moderately successful in distinguishing the three contours, although not equally for all three accent patterns. Our first prediction was that participants would be accurate in distinguishing the presence versus absence of a prenuclear accent in their imitation of the stimulus accent patterns. Specifically, we expected imitations of the Unaccented pattern would have no PPA on the target word, while imitations of stimuli with Primary and Early High PPA patterns would realize a PPA on one of the two possible locations of the target word. As expected, Primary and Early High patterns were mostly, if not always, reproduced with a PPA on the target word. Some imitations of the Primary accent pattern were classified as Unaccented based on an analysis of the relative mean F0 in three intervals spanning the prenuclear and nuclear region (10 % in Figure 5), but the subset clustering analysis of Primary imitations calls into question that finding, suggesting instead that "incorrect" imitations of the Primary pattern represent a damped version of Early High. On the other hand, all imitations of Early High had a notable F0 peak in the target word, and more specifically, in the first interval. Thus, we have partial confirmation of our first prediction, that imitations of Primary and Early High accent patterns have a prenuclear accent on the target word. Somewhat surprisingly, this prediction is not fully confirmed in imitations of the Unaccented pattern, where "incorrect" imitations with a PPA on the target word were observed in 85 % of the Unaccented data based on the classification analysis (Figure 5), which was lowered to 47 % of the Unaccented data in the subset cluster analysis (B panel in Figure 8c). The Unaccented pattern was thus dispreferred to some extent.

Our second prediction was that the distinction between the Early High and Primary accent patterns would be less accurately reproduced in imitations than the distinction between either of these and the Unaccented pattern. Contours with an early accentual peak, on a syllable with secondary lexical stress, were imitated particularly well and variants of this contour were also frequently found in the "incorrect" (cluster B in Figure 8) imitations of the other two accents. The prevalence of a (weak or strong) Early High accent pattern across all imitations may be interpreted as an indication that this eurhythmic pattern, which has been described in the phonological literature as involving accent retraction, is in fact the preferred pattern in this English variety.

On the other hand, however, further analysis revealed that many of the Early High reproductions actually show pitch falling across the syllable with primary stress—a pattern that was not present in the stimuli. It thus appears that contours with a H* tone on the initial, secondary-stressed syllable of words like **_adversarial_**, **_university_**, etc. and contours with a falling (H+!H*) accent on their third, primary-stressed syllable, _adversarial_, _university_, form a phonological equivalence class. The distinctive feature of this broader pattern is the presence of a pitch peak in the region

before the syllable with primary stress. That is, it is conceivable that the two clusters of Early High imitations, the contour with H* on the secondary-stressed syllable and the contour with H+!H* on the primary-stressed syllable, have the same general prominence pattern marked by higher F0 in the first foot than the second foot in the target word (1–2 words). As this study did not set out to test allophony in prenuclear accent patterns, we offer this as a *post hoc* interpretation of our findings that should be further examined in future research specifically designed to test the distinction between H* and H+!H* in prenuclear position.

Whereas we expected the Unaccented pattern to be perceived and imitated as clearly distinct from the other two (which did not turn out to be the case), the distinction between the Primary and Early High patterns was expected to be less accurately reproduced, since no meaning distinction between these patterns has been reported. In both cases, the target word receives a prenuclear accent. Our classification analysis (Figure 5) showed that the stimulus Primary pattern was reproduced as Primary (60 %) followed by Early High (30 %). Our cluster analysis (Figure 8a) also found that the stimulus Primary pattern was reproduced as Primary (56 % in Panel A) and a hybrid pattern of Early High and Unaccented (44 % in Panel B).

On the other hand, our participants showed a preference from the Early High pattern that would be predicted from rhythmic considerations as discussed on the metrical literature on English phrasal stress, but not from a perspective where what matters is whether or not a word carries a pitch accent. If Primary and Early High are not clearly distinctive, both patterns would be activated in imitation, thus, they would be unfaithfully reproduced in imitation at a similar rate. However, this was not the case. Our classification analysis (Figure 5) showed correct imitations were 60 % for Primary pattern and 85 % for Early High pattern. Further cluster analysis (Figure 8) revealed that "incorrect" imitations resembled the Early High pattern for both Primary and Unaccented patterns. In sum, it seems that Early High was more likely to be activated and faithfully reproduced than Primary in imitation. The rhythmic stress reversal rule (*thirtéen → thírteen mén*) seems to be indeed the preferred pattern for our participants, in agreement with the predictions of the metrical literature on English phrasal stress. Whereas Grabe and Warren (1995) failed to find robust acoustic evidence for stress shift, our analysis of pitch contours has allowed us to distinguish shifted (Early High) from unshifted (Primary) and Unaccented patterns in production.

5 Conclusions

In this study, we examined whether the presence and position of prenuclear pitch accents can be imitated by listeners. In our stimuli, the first word in a prosodic

phrase could be unaccented, bear an accent on the syllable with primary stress or bear an early accent anchored on a syllable with secondary stress. A following word in the phrase always bore a (nuclear) accent. The results of our imitation study showed an overwhelming preference for a rhythmic pattern with an early pre-nuclear accent, whether or not this was actually present in the stimulus. This early accent was produced in one of two ways: either as a H* accent on the syllable with secondary stress or as a H+!H* configuration on the syllable with primary stress. When the stimuli contained an early high accent on the first word of the phrase, this was almost always reproduced as such (in one of the two ways just described). In imitating stimuli where there was no prenuclear accent as well, the majority tendency was to introduce an early prenuclear accent, with relatively few faithful reproductions of the unaccented pattern. Finally, when the first word in the phrase contained a H* accent anchored on the syllable with primary stress, this pattern was faithfully imitated most of the time, but again with a sizable number of imitations with an earlier accentual peak.

We conclude that our participants showed a preference for the rhythmic rule of English that has been described in the phonological literature (e.g., *Míssissippi Ríver*) and a clear dispreference for the least rhythmic pattern, where the first word in the phrase is realized as unaccented. These phonological preferences led them to disregard in many cases the actual contours present in the stimuli that they were asked to imitate.

The results of our experiment are consistent with the view that the placement of prenuclear accents in US English is driven by rhythmic rather than pragmatic reasons. That is, prenuclear accents are "ornamental" in the sense of Büring (2007). A limitation for any broader conclusions is, of course, that we have examined only one context.

Finally, what we are calling Early High was in fact produced in two phonetically distinct ways, which we suggest correspond to two different phonological analysis in the ToBI framework (Beckman et al. 2005). This points to the possibility that these two patterns may in fact be allophonic.

Research ethics: This research was approved by the UIUC Office of the Vice Chancellor for Research (IRB Protocol 13473). Informed consent was obtained from all participants.
Conflict of interest: The authors have no conflicts of interest to declare.
CRediT author statement: Suyeon Im: Methodology, Formal Analysis, Investigation, Data Curation, Writing - Review & Editing, Visualization. **José Ignacio Hualde**: Conceptualization, Methodology, Formal Analysis, Writing - Original Draft, Writing - Review & Editing. **Jennifer Cole**: Conceptualization, Methodology, Formal Analysis,

Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Supervision, Project Administration, Funding Acquisition.

Supplementary materials

The statistical results from the linear mixed-effects regression and the Analysis of Variance discussed in 3.1 (Supplementary material 1), and the supplementary materials from the analysis of experimental sentences in 2.1 as well as the analyses of classification and cluster using F0 maximum (Supplementary material 2) can be found in the online repository, https://osf.io/3e5sh/.

References

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Baumann, Stefan & Arndt Riester. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In Gorka Elordieta & Pilar Prieto (eds.), *Prosody and meaning*, 119–162. Berlin & New York: Mouton De Gruyter.

Baumann, Stefan & Arndt Riester. 2013. Coreference, lexical givenness and prosody in German. *Lingua* 136. 16–37.

Beaver, David & Dan Velleman. 2011. The communicative significance of primary and secondary accents. *Lingua* 121(11). 1671–1692.

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.

Beckman, Mary E., Hirschberg Julia & Stefanie Shattuck-Hufnagel. 2005. The original ToBI system and the evolution of the ToBI framework. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 9–54. Oxford, UK: Oxford University Press.

Bishop, Jason. 2017. Focus projection and prenuclear accents: Evidence from lexical processing. *Language, Cognition and Neuroscience* 32(2). 236–253.

Boersma, Paul & David Weenink. 2020. Praat: Doing phonetics by computer. [Computer program]. Available at: http://www.praat.org/.

Bolinger, Dwight. 1972. Accent is predictable (if you're a mind reader). *Language* 48(3). 633–644.

Braun, Bettina, Greg Kochanski, Esther Grabe & Burton S. Rosner. 2006. Evidence for attractors in English intonation. *The Journal of the Acoustical Society of America* 119(6). 4006–4015.

Büring, Daniel. 2006. Focus projection and default prominence. In Valéria Molnár & Susanne Winkler (eds.), *The architecture of focus*, 321–346. Berlin, Germany: Mouton De Gruyter.

Büring, Daniel. 2007. Intonation, semantics, and information structure. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces*, 445–473. Oxford, UK: Oxford University Press.

Calhoun, Sasha. 2010. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language* 86(1). 1–42.

Chafe, Wallace. 1987. Cognitive constraints on information flow. In Russell S. Tomlin (ed.), *Coherence and grounding in discourse*, 21–51. Amsterdam, Netherlands: John Benjamins Publishing Company.

Chodroff, Eleanor & Jennifer Cole. 2018. Information structure, affect, and prenuclear prominence in American English. *Proceedings of Interspeech 2018*. 1848–1852. https://doi.org/10.21437/Interspeech. 2018-1529.

Cole, Jennifer. 2015. Prosody in context: A review. *Language, Cognition and Neuroscience* 30(1–2). 1–31.

Cole, Jennifer & Stefanie Shattuck-Hufnagel. 2011. The phonology and phonetics of perceived prosody: What do listeners imitate? *Proceedings of Interspeech 2011*. 969–972. https://doi.org/10.21437/ interspeech.2011-395.

Cole, Jennifer, Jeremy Steffman, Stefanie Shattuck-Hufnagel & Sam Tilsen. 2023. Hierarchical distinctions in the production and perception of nuclear tunes in American English. *Laboratory Phonology* 14(1). https://doi.org/10.16995/labphon.9437.

Dilley, Laura C. 2010. Pitch range variation in English tonal contrasts: Continuous or categorical? *Phonetica* 67(1-2). 63–81.

D'Imperio, Mariapaola, Rossana Cavone & Caterina Petrone. 2014. Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology* 5. 1226.

Frota, Sónia. 2003. The phonological status of initial peaks in European Portuguese. *Catalan Journal of Linguistics* 2. 133–152. https://www.raco.cat/index.php/CatalanJournal/article/view/308975.

Genolini, Christophe, Xavier Alacoque, Mariane Sentenac & Catherine Arnaud. 2015. kml and kml3d: R packages to cluster longitudinal data. *Journal of Statistical Software* 65(4). 1–34.

German, James Sneed. 2012. Dialect adaptation and two dimensions of tune. In Qiuwu Ma, Hongwei Ding & Daniel Hirst (eds.), *Proceedings of the international conference on speech prosody*, 430–433. Shanghai, China: Tongji University Press.

Goodhue, Daniel & Michael Wagner. 2018. Intonation, yes and no. *Glossa: A Journal of General Linguistics* 3(1). 5. 1–45.

Grabe, Esther & Paul Warren. 1995. Stress shift: Do speakers do it or do listeners hear it? In Bruce Connell & Amalia Arvaniti (eds.), *Papers in laboratory phonology IV*, 95–110. Cambridge, UK: Cambridge University Press.

Gussenhoven, Carlos. 1983. *A semantic analysis of the nuclear tones of English*. Bloomington, IN: Indiana University Linguistics Club.

Gussenhoven, Carlos. 1984. *On the grammar and semantics of sentence accents*. Dordrecht, Netherlands: Foris Publications.

Gussenhoven, Carlos. 2011. Sentential prominence in English. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, Vol. 5, 2778–2806. Chichester, UK: Wiley-Blackwell.

Gussenhoven, Carlos. 2015. Does phonological prominence exist? *Lingue e Linguaggio* 14(1). 7–24.

Halliday, Michael A. K. 1967. *Intonation and grammar in British English*. The Hague, Netherlands: Mouton.

Horne, Merle. 1990. Empirical evidence for a deletion formulation of the rhythm rule in English. *Linguistics* 28. 959–981.

Ladd, D. Robert. 1980. *The structure of intonational meaning: Evidence from English*. Bloomington, IN: Indiana University Press.

Ladd, D. Robert. 2008. *Intonational phonology*, 2nd edn.; 1st edn. 1996. Cambridge, UK: Cambridge University Press.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2). 249–336. https://www.jstor.org/stable/4177987.

Nespor, Marina & Irene Vogel. 1989. On clashes and lapses. *Phonology* 6(1). 69–116.

Petrone, Caterina & Oliver Niebuhr. 2014. On the intonation of German intonation questions: The role of the prenuclear region. *Language and Speech* 57(1). 108–146.

Petrone, Caterina, Daria D'Alessandro & Simone Falk. 2021. Working memory differences inprosodic imitation. *Journal of Phonetics* 89. 101100.

Pierrehumbert, Janet B. 1980. *The phonetics and phonology of English intonationI*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Prieto, Pilar. 2015. Intonational meaning. *Wiley Interdisciplinary Reviews. Cognitive Science* 6(4). 371–381.

R Core Team. 2024. *R: A language and environment for statistical computing [Computer software]*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116.

Ross, Kenneth, Mari Ostendorf & Stefanie Shattuck-Hufnagel. 1992. Factors affecting pitch accent placement. In Ross, Kenneth, Mari Ostendorf & Stefanie Shattuck-Hufnagel. 1992. Factors affecting pitch accent placement. In *Proceedings of the international conference on spoken language processing* 1, 365–368. Edmonton, Canada: University of Alberta.

Selkirk, Elisabeth. 1995. Sentence prosody: Intonation, stress and phrasing. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 550–569. Cambridge, MA: Blackwell.

Selkirk, Elisabeth. 2007. Contrastive focus, givenness, and the unmarked status of "discourse-new". In Caroline Féry, Gisbert Fanselow & Manfred Krifka (eds.), *Interdisciplinary studies on information structure 6*, 125–145. Potsdam, Germany: Universitätsverlag Potsdam.

Shattuck-Hufnagel, Stefanie. 1988. Acoustic-phonetic correlates of stress shift. *The Journal of the Acoustical Society of America* 84. S98.

Shattuck-Hufnagel, Stefanie. 1992. The role of word structure in segmental serial ordering. *Cognition* 42(1–3). 213–259.

Shattuck-Hufnagel, Stefanie. 1995. The importance of phonological transcription in empirical approaches to "stress shift" versus "early accent": Comments on Grabe and Warren, and Vogel, Bunnell, and Hoskins. In Bruce Connell & Amalia Arvaniti (eds.), *Papers in laboratory phonology IV*, 128–140. Cambridge, UK: Cambridge University Press.

Torreira, Francisco & Martine Grice. 2018. Melodic constructions in Spanish: Metrical structure determines the association properties of intonational tones. *Journal of the International Phonetic Association* 48(1). 9–32.

Wagner, Michael & Duane G. Watson. 2010. Experimental and theoretical advances in prosody: A review. *Language & Cognitive Processes* 25(7–9). 905–945.

Xu, Yi. 2013. ProsodyPro – A tool for large-scale systematic prosody analysis. In Brigitte Bigi & Daniel Hirst (eds.), *Proceedings of tools and resources for the analysis of speech prosody (TRASP 2013)*, 7–10. Aix en Provence, France: Labratoire Parole et Langage.

Zahner-Ritter, Katharina, Marieke Einfeldt, Daniela Wochner, Angela James, Nicole Dehé & Bettina Braun. 2022. Three kinds of rising-falling contours in German wh-questions: Evidence from form and function. *Frontiers in Communication* 7. 58.

Zwicker, Eberhard & Hugo Fastl. 2013. *Psychoacoustics: Facts and models*. Berlin, Germany: Springer Science & Business Media.