

Phonological Aspects of Hesitation Disfluencies

Vered Silber-Varod

School of Jewish Studies, Tel Aviv University, Israel

vereds@openu.ac.il

Abstract

An effective approach to the study of prosody in spoken language seeks to identify prosodic patterns and their communicative values, and to subsequently find a correlation between these prosodic patterns and other layers of linguistic structure. The present research strives to define a single prosodic boundary pattern: the boundary tone of hesitation disfluencies in spontaneous Israeli Hebrew. This entails uncovering the phonological environments in which they occur. Results show two distinct domains for such disfluencies with regard to word-level phonology: word-final syllables and appended *e* vowels that are inserted after a word, but within the same intonation unit. Statistically significant relations were found between these domains and the phonological structures of the disfluent syllables.

Index Terms: prosody, spontaneous Hebrew, disfluencies

1. Introduction

The current research seeks to define the phonological environments of hesitation disfluencies. However, in order to establish the frequency of any aspect of prosody, it is first necessary to categorize the prosodic “world.” The first assumption is that the basic structural unit of spoken language analysis is the Intonation Unit (IU) [9]. IU segmentation has already been implemented on spontaneous Israeli Hebrew (IH, also known as Modern Hebrew) in The Corpus of Spoken Israeli Hebrew – CoSIH [10], which was the database utilized for the purposes of the present research.

1.1. The communicative value of disfluent boundary tones

I adopt here a conclusive vs. continuative binary classification of IU boundary tones. This dichotomy is commonly implemented when studying prosodic-discourse interfaces [12], and is based on Brazil's communicative value of intonation [5]. According to Clark and Wasow [6], hesitation disfluencies, which are the focus of this paper, exist when speakers suspend speaking after an initial commitment to it. This suspension is indicated on one of the prosodic boundary tones and signals the speaker's wish to continue. This phenomenon of disfluent speech was also mentioned in [6, 18] and is among other prosodic “techniques,” such as the *Rising* tone mentioned in [14, 16], employed to keep speech as cohesive as possible.

According to IH-ToBI [8], “a disfluent break long enough for an intermediate phrase boundary” is tagged as p3. The term that I will use to describe hesitation disfluency boundaries is *Continuous Elongation* (CE). This term was selected due to the fact that the most prominent phonetic cue of these disfluencies is segment elongation. The use of a prosodic term, i.e. *Continuous Elongation* to describe the focus of this research is of major importance as it encompasses a premise regarding this recognized prosodic pattern.

My research is presented from the point of view of the speaker, not the listener. This is because considering the perceptual aspect of disfluencies may lead to the conclusion

that, since they are interpreted by listeners as phenomena that interrupt the flow of speech and do not add propositional content to an utterance, they are therefore not useful to study. The terminology used was therefore adapted to this approach, and perceptual or psycholinguistic terms such as “hesitations” are replaced here with a prosodic characteristic term: elongations. The term “filled pauses,” mentioned in [7, 15, 17], was not adopted either, although it is well-known and frequently used in prosodic research. The reason for this is that the term was defined as an isolated *e* or *em* (the default IH filled pauses) and not as a pattern of prosody. However, the syntactic structures surrounding filled pauses [6, 15] have previously been researched, either for implementing them into Language Modeling [17], or for understanding the cognitive process of speech [18].

Indeed, prosodic elongations are speech-sound events that interface with other sounds of speech, namely segments, and “prosody cannot and should not be dissociated from analyses of variable speech phenomena.” [11] Thus the goal of this paper is to uncover the phonological environment of the *Continuous Elongation* boundary tone in IH. In more general terms, this paper seeks to illuminate phonological regularities in *Continuous Elongation* production as a phase in the process of studying the prosodic-syntactic interface in disfluent contexts, cf. [6, 15, 17, 18].

2. Research question

The hypothesis in this research is that *Continuous Elongation* boundaries in IH do not always occur in isolation, as *e* or *em*, but are parasitic to the final syllables of lexical items which are elongated in order to accommodate the disfluency. This assumption raises a key question: are open word-final syllables more likely to “attract” elongations than closed syllables, due to the [+continuant] feature of vowels?

3. Method

The perceptual segmentation into IUs was performed by the author, who is a native Hebrew speaker. 25% of the recordings were also annotated by two additional researchers (also native speakers). Each IU boundary was marked perceptually as either continuous, whenever the final tone of the IU signaled “more to come” or as terminal, when intonation signaled “nothing more to say.” Truncated IUs were also annotated.

3.1. Phonetic realizations in *Continuous Elongation* (CE) boundary tones

Alignment between the transcription and the sound wave was achieved using the Praat textgrid tool [2], with considerable reliance on acoustic features and measurements. Although the rhythmic aspect (the excessive lengthening of the final syllable) is the prominent cue of CE boundaries, two additional acoustic parameters, formants and F0, are also *typical* to this kind of boundary and are present in most of CE boundaries.

Table 1. *Phonological scheme of disfluent boundaries in IH.*

Final syllable type	Final segment	Elongated segment	Measure of duration	Threshold
Open	Any vowel	Vowel	Onset+nucleus	>230ms
Closed	Only <i>a, i, o, or u</i>	Appended <i>e</i>	N/A	Any length of appended <i>e</i>
	Consonant	Vowel+consonant	Nucleus+coda	>230ms
	Consonant	Appended <i>e</i>	N/A	Any length of appended <i>e</i>

3.1.1. Formants

Hesitation disfluencies in IH are typically produced via the insertion of an *e* vowel. This *e* vowel is realized as the mid-front vowel [ɛ]. *e* is the phonetically unmarked vowel in Israeli Hebrew and is the vowel most likely to break up undesirable consonant clusters and to be affected by casual vowel deletion [3]. This vowel is among four others, *a, i, o* and *u*, in the IH vowel system [4]. Formant measurements in the current corpus sampling show correlation with the *e* vowel formants found in [13]. The vowel quality of elongations is in all probability the most language dependent variable of elongations. In French, orthographic transcriptions of *ehh* are commonly found. These are pronounced, in SAMPA, "2" (bolded in (1)) or "9" but also as an E bolded in (2) (examples courtesy Mathew Avanzi):

e 2 la Sa~bR d2 d@ kO mERs (1)
'*et euh la Chamber de de Commerce*'

j a~ na e~ i fE ve~ s@mi s@mEn i fE (2)
'*y en a un il fait vingt semi semaines il fait*'

American-English transcriptions of *uh* and *um* indicate another vowel quality, [V] in SAMPA. A characteristic American-English *uh* insertion is demonstrated in (3).

And we don't always know **uh** all the facts that (3)
were involved
[Obama's interview on Al-Arabiya TV, 27/1/09.
http://www.youtube.com/watch?v=HO_ILttxrs.]

3.1.2. F0

The CE boundary tone is produced on a level-tone that is in the average key of the speaker. In sample measurements of a single recording from an adult female speaker in the corpus, the pitch of the elongated vowels was measured at 192Hz.

3.1.3. Duration

The most prominent perceptual cue of CE boundaries is length. The problem is that perceptual elongation will always be identified as such, relative to the adjacent syllables and not according to an absolute acoustic parameter. Therefore, in order to avoid bias during the labeling stage, a threshold of 230ms was set as a minimum duration for all CE boundaries. This decision was based on sample measurements of a single recording (2000 words, 680 IUs). Examples (4) and (5) demonstrate three cases of such excessive lengthening.

tRemp ze **Sta:** medabeR im anaSim ba eRev (4)
'*hitchhiking is when you are talking to people in the evening*'

ve **ze: ata: #** (5)
'*and this you*'

3.2. The phonological structure of *Continuous Elongation* (CE) boundary tones

Table 1 sets a scheme of the possible syllabic structures that can be found in word-final syllables and the corresponding expected elongation type. The far left column specifies open vs. closed syllable structures. The next column to the right

shows the possible *final segments* in the syllable. The *elongated segment* column details which segments are expected to undergo elongation: the *vowel* in the case of any open syllable, such as in *o:* 'or', *keilu:* 'as if', *be:* 'in'; or an *appended e* in the case of open syllables that do not end with an *e*, as in (6).

ki lo haja la **e:** (6)
'*because she didn't have e*'

In the case of closed syllables in the last lexical word of an IU, either the vowel nucleus vowel or the final consonant, in the case of sonorant or continuant codas, may be lengthened. It is also possible for stops to be lengthened, as in the case of the Swedish disfluencies *att* 'that' and *och* 'and', which were documented and analyzed in [18]. Nevertheless, an appended *e* is generally expected following closed syllables, as in (7).

aval **e:** 'but e' (7)

This *e* was treated as a CE boundary cue whenever it was phonologically attached to the last word of an IU. Evidently, the elongated syllable is not always part of a lexical word and although no lexical meaning is attributed to this appended *e*, it is marked as a hesitation expression in spoken Hebrew and appears in IH slang dictionaries. The suggestion that this appended *e*, the default vowel in IH, is (part of) an elongated prosodic pattern is a new notion this paper would like to raise. In accordance with this prosodic perspective, isolated *es* or *ems* were also tagged with CE boundary tone.

In Table 1, the *Measure of duration* column indicates which segments were measured in each syllabic structure and the *Threshold* column details the cases where a threshold of 230ms was set. It should be noted that in a good many cases, perception of elongation was ruled out because of this 230ms threshold. In cases of appended *es*, no duration measurements were carried out and such cases were automatically marked as CE boundaries.

4. Data

The corpus contains 18 face-to-face spontaneous dialogues from CoSIH [10]. Each dialogue consists of conversations between one core speaker and various interlocutors the speaker interacted with on that day. It is already common knowledge that analyzing spontaneous speech allows us to identify unique patterns of prosody that cannot be established in read speech [17]. Sentences that are read by informants in a lab, or a lecture that is read from a manuscript establish a different spectrum of prosodic patterns than those found in spontaneous speech.

The total duration of the recordings is 189 minutes, with 23,932 word tokens and 3,714 word types (after clean-up). Among the 5,890 Intonation Units (IU) in the corpus, 3,712 (63%) were Terminal IUs. Terminal IUs refer to those intonation units that contain a terminal tone found in question IUs or otherwise perceptually determined terminal tones. This massive amount of Terminal IUs is reasonable considering the dialogue genre of the corpus which includes backchannel expressions (*ken* 'yes', *okej* 'OK' etc.) and other short utterances that are the result of rapid back-and-forth conversation. 296 IUs (5%) were truncated; and 1,882

(31.95%) contain various Continuous boundaries (either rising, level, falling, rise-fall or CE tones). 740 IUs were annotated with a CE boundary tone, making up 12.58% of all IUs in the corpus and 39.3% of IUs with continuative boundary tones. As mentioned above, it is these IUs that are the focus of this paper.

5. Results

Results show three distinct domains of elongation: isolated *e*; or *em* (i.e. filled pauses); word-final syllables; and an appended *e* following a lexical word but within the same IU. The distribution of these domains is as follows: 42 cases are of an *e* or *em* that constitute an isolated monosyllabic IU, i.e. before or after a pause or with a strong IU boundary (these 42 cases are not mentioned in Table 2); 698 can be divided into two classes: 502 with elongated word-final syllables; and 196 with appended *es*. Table 2 is a contingency table summarizing the phonological structure of three prosodic variables: two groups of disfluent speech (*Elongated Word-final syllables* and *Word-final syllables with an appended e*) and one group of fluent speech (*NON-elongated Word-final syllables*). Data were retrieved using AntConc 3.2.1 version [1]. The results of the contingency table are statistically significant ($\chi^2 = 430$; $p < 0.001$). This means that the two variables mentioned in Table 1, *syllable type* and *elongated segment* variables, show contingency. Another group of diphthongs is mentioned but as shown in Table 2, this type of word-final syllable constitutes only 2% of the corpus, and is therefore ignored in the following discussion.

Table 2. Frequency and percent of word-final syllable types per prosodic variable.

Final Syllable type	Freq. in NON-elongated Word-final syllables and (%)	Freq. in elongated Word-final syllables and (%)	Freq. in word-final syllables with an appended <i>e</i> and (%)	Total Freq. and (%)
open	12813(55)	452(90)	14(7)	13279(55)
diphthong	384(2)	3(1)	6(3)	393(2)
closed	10037(43)	47(9)	176(90)	10260(43)
Total	23234(100)	502(100)	196(100)	23932(100)

The following sections discuss the featural distinctions between word-final segments of the two major syllabic groups, open syllables and closed syllables. With regards to the open syllable group, differences between the 5 IH vowels at the nuclei will be examined in relation to the three prosodic variables (§5.1). With regards to the closed syllable group, the manner of articulation of the coda segment will be examined in relation to the three prosodic variables (§5.2).

5.1. Elongation of open syllables

452 occurrences (90%) of elongated word-final syllables occurred in open syllables. This accounts for 64% of 698 CE boundaries, cf. Table 2. Moreover, the most frequent elongated word-types are open monosyllabic words: 82 occurrences of *ha* 'the', 39 occurrences of *ve* 'and', 30 occurrences of *ze* 'this', 28 occurrences of *ba* 'in the/at the', 25 occurrences of *Se* 'that', etc. Worth noting is the fact that these are function words. Elongated word-final syllables also occur in disyllabic words such as *ha'ze* 'that' or in monosyllabic CCV structures such as in *kSe* 'when' and *kmo* 'as'. As shown in Table 2, 90% of the appended *es* were added to closed syllables while only 7% were added to open syllables, such as *hi* 'she' and *ani* 'I am' (see also (6) above).

Table 3 allows a closer look at the vowel quality of the word-final nuclei and its relation to the three prosodic variables. Elongation takes place in the case of each of the five vowels, and the rate of occurrence is comparable to the relative frequency of the vowel overall. In the case of word-final /e/, 153 elongated open Ce syllables account for 34% of 452 elongated open syllables while their ratio in non-elongated word-final syllables is only 27%; the ratios of Co and Ci are lower in elongated word-final syllables than in non-elongated word-final syllables; while Cu and Ca have more or less the same ratio in the elongated group, the non-elongated group and in the entire corpus. The results of the contingency test in Table 3 were found to be statistically significant ($\chi^2 = 36.8$; $p < 0.001$), i.e. there is a significant contingency between the two variables: final vowel and prosodic variables.

Table 3. Frequency and percent of vowel quality types in open word-final syllables.

Final vowel type	Freq. in NON-elongated Word-final syllables and (%)	Freq. in elongated Word-final syllables and (%)	Freq. in word-final syllables with an appended <i>e</i> and (%)	Total Freq. and (%)
a	4634(36)	161(36)	4(29)	4799(36)
e	3433(27)	153(34)	0(0)	3586(27)
i	2195(17)	61(13)	8(57)	2264(17)
o	1608(13)	38(8)	2(14)	1648(12)
u	943(7)	39(9)	0(0)	982(7)
Total	12813(100)	452(100)	14(100)	13279(100)

5.2. Elongation of closed syllables

Elongation of closed syllables occurs in 47 cases (9% of elongated word-final syllables). This meets the low expectation of elongated closed word-final syllables that was mentioned in §3.2. As opposed to open syllables which exhibited elongation effects on the word-final segment, closed syllables tend to be elongated by appending an additional *e* vowel to the end of the word (cf. table 2). Thus, the data show that all elongated closed syllables occur between 1-3 times and the most frequently elongated words in this group are *aval* 'but' and *hem* 'they' which occur four times each.

Table 4 allows us a closer look at the segmental features of the codas and their relation to the three prosodic variables. In the case of stop codas, five elongated cases represent the unlikely tendency of $\sigma[+\text{obstruent}]$ to be elongated. This comes to a ratio of (11%), far less than the ratio of stops in the corpus (28%). The 18 elongated nasal codas make up a far greater percentage (38%) than their ratio in the corpus (28%). Lateral [l] codas are most likely to be elongated, either by lengthening or by an appended *e*, while fricative codas are less likely to be elongated compared to their ratio in the corpus. In summary, the results of the contingency test in Table 4 were found to be statistically significant ($\chi^2 = 54.6$; $p < 0.001$).

Table 4. Frequency and percent of coda types at closed word-final syllables.

Final coda feature	Freq. in NON-elongated Word-final syllables and (%)	Freq. in elongated Word-final syllables and (%)	Freq. in word-final syllables with an appended <i>e</i> and (%)	Total Freq. and (%)
stops	2802(28)	5(11)	37(21)	2844(28)
nasals	2903(29)	18(38)	48(27)	2969(29)
fricatives	3271(33)	13(28)	46(26)	3330(32)
lateral	1061(11)	11(23)	45(26)	1117(11)
Total	10037(100)	47(100)	176(100)	10260(100)

Lengthened nasal codas, either [m] or [n], occur in words like the pronoun *hem* 'they', and the adverb *en* 'there isn't'. Lengthened fricatives, among others, are [z] in *az* 'so' and [X] in the Question word *eX* 'how'. Worth noting is that the most frequent words with an appended *e* are the function words: *az e* 'so eh' (20 cases) *aval e* 'but eh' (16 cases) and *Sel e* 'of eh' (14 cases). These findings remain for future research regarding the prosodic-syntactic interface.

5.3. Silent pauses

In contrast to the term *filled pauses*, which implies an alternation between unfilled and filled pauses, silent pauses do occur following CE boundaries. In 203 cases (27.5% of CEs), a CE boundary is followed by a pause (#). In 64 cases, the pause follows an appended *e* and in 8 cases it is a pause after an *e* that constitutes an isolated IU, as in (9).

ze biRat	'it's the capital of'	(9)
#		
e:	'eh'	
#		
ha:	'the'	
ha:	'the'	
soT gObi	'South Gobi' [Omer recording]	

6. Discussion

This data driven research has been undertaken in order to put forward the notion of the *Continuous Elongation* boundary tone in IH. The interface between prosody and the segmental context is interesting in the case of CE boundaries since production of elongation is parasitic to word-level segments. The research presented here has attempted to categorize the segments that undergo elongation and identify how prosody compensates when segments are neither sonorants nor continuants.

CE in IH is noted on word-final sonorants, both vowels and consonants, as well as on words that end in obstruent consonants with an appended *e* vowel in final position, which itself appears to undergo elongation, i.e. CE occurs in any segmental context, and not only in the expected open vowels. In other words, given a phonological environment, prosody "finds its way" to articulate itself. This was found to be true in the present research and also in [16], where stressed vs. unstressed word-final syllables were examined in Continuous Rise boundaries.

The suggestion that the appended *e*, the default vowel in IH, is (part of) an elongated prosodic pattern is a new notion this paper has raised. This notion is based on two arguments: First, although lengthened CV syllables containing an *e* (e.g. *Se* 'that') cannot be divided into two *e*-components (i.e., *Ce+e*), other vowels with successive *es* are evidence to this appended increment. Furthermore, this (typically long) *e* is prosodically attached to the preceding word, which most likely would not be elongated were it not for this appended *e* (cf. §5.2).

It should be emphasized that, although the present research was made on perceptual grounds of IU segmentation, only data on CE boundaries, i.e. hesitation disfluencies, were examined. Although in some respects, realization of CE boundary tones is language dependent, the present study and the studies of hesitation disfluencies in other languages [15, 17, 18] show that this prosodic pattern can be identified by its communicative value (the speaker wants to continue his/her speech and does so by substantial elongation). A cross-linguistic study may illuminate the prosodic features common, and possibly unique, to what this paper refers to as CE boundaries.

Another linguistic aspect that was revealed during this study concerns the lexical words that were elongated and the

parts of speech (POS) they belong to. It seems that discrepancies in elongation frequency between the five vowels and the four manners of articulation can also be attributed to the frequency of specific *words* that are more commonly elongated and not only due to phonological features. However, in order to complete a comprehensive linguistic description of hesitation disfluencies in IH, interface with other layers of language should be considered and studied in detail.

7. Acknowledgements

I would like to thank Prof. Shlomo Izre'el for his comments and suggestions and Mrs. Michal Zuker for her help with the statistical tests. Special thanks to the three anonymous reviewers, who addressed their critical comments with a supportive attitude.

8. References

- [1] Anthony, L. 2008. *AntConc version 3.2.1w*. Center for English language education, Waseda University. Retrieved from <http://www.antlab.sci.waseda.ac.jp/index.html>
- [2] Boersma, P. & Weenink, D. 2009. Praat: Doing phonetics by computer (Version 5.1.10) [Computer program]. Retrieved July 8, 2009, from <http://www.praat.org/>
- [3] Bolozky, S. 1999. On the special status of the vowels *a* and *e* in Israeli Hebrew. *Hebrew Studies* 40. 233-250.
- [4] Bolozky, S. 1997. Israeli Hebrew Phonology. In Kaye, A. S. & P. Daniels (eds.), *Phonologies of Asia and Africa* 1. 287-311. Winona lake: Eisenbauns.
- [5] Brazil, D. 1997. *The communicative value of intonation*. Cambridge University Press.
- [6] Clark H. H. and Wasow T. 1998. Repeating words in spontaneous speech. *Cognitive Psychology* 37. 201-242.
- [7] Cruttenden, A. 1997. *Intonation*. 2nd edn. Cambridge University Press.
- [8] Green H. & Tobin Y. 2009. Prosodic analysis is difficult... but worth it: A study in high functioning autism. *International Journal of Speech-Language Pathology* 11(4). 308-315.
- [9] Izre'el S. 2005. Intonation units and the structure of spontaneous spoken language: A view from Hebrew. In Auran, C, R. Bertrand, C. Chanet, A. Colas, A. Di Cristo, C. Portes, A. Reynier & M. Vion (eds.), *IDP05 International Symposium on Discourse-Prosody Interfaces*. CD ROM. <http://aune.lpl.univ-aix.fr/~prodige/idp05/actes/izreel.pdf>.
- [10] Izre'el, S. *The Corpus of Spoken Israeli Hebrew (CoSIH)*. <http://www.tau.ac.il/humanities/semitic/cosih.html> (19.10.2009)
- [11] Lacheret A. & Lyche C. 2008. Looking at French schwa in initial position through the glasses of prosody. *Speech Prosody 2008*. 523-526. Campinas, Brazil.
- [12] Lacheret-Dujour, A. 2007. Prosodie-discours: Une interface à multiples facettes. *Nouveaux Cahiers de Linguistique Française* 28. 7-40.
- [13] Most, T., Amir, O. & Tobin, Y. 2000. The Hebrew vowel system: Raw and normalized acoustic data. *Language and Speech* 43(3). 295-308.
- [14] Portes, C.; Bertrand, R.; Espesser, R. 2007. Contribution to a grammar of intonation in French form and function of three rising patterns. *Nouveaux Cahier de Linguistique Francaise* 28. 155-162.
- [15] Shriberg, E. E. 1994. Preliminaries to a theory of speech disfluencies. PhD thesis, University of California at Berkeley.
- [16] Silber-Varod, V. & Kessous, L. 2008. Prosodic boundary patterns in Hebrew: A case study of continuous IUs in weather forecasts. *Speech Prosody 2008*. 265-268. Campinas, Brazil.
- [17] Stolcke, A. & Shriberg, E. E. 1996. Statistical language modeling for speech disfluencies. *ICASSP-96*, May 7-10, Atlanta, GA: IEEE.
- [18] Roll, M., Frid, J., & Horne, M. 2007. Measuring syntactic complexity in spontaneous spoken Swedish. *Language and Speech* 50(2). 227-245.