Characteristics of Intonation Unit Boundaries in Spontaneous Spoken Hebrew – Perception and Acoustic Correlates

Noam Amir∗, Vered Silber-Varod & Shlomo Izre'el∗∗

∗Department of Communication Disorders; ∗∗Department of Hebrew and Semitic Linguistics
Tel Aviv University
{noama; izreel}@post.tau.ac.il; vereds@openu.ac.il

Abstract

The notion of intonation units is very basic to the study of discourse. Nevertheless, a clear-cut definition of what comprises an intonation unit has not been forthcoming. In reality, it seems that the boundaries delineating intonation units are somewhat easier to define, though this is by no means a closed subject. In this preliminary study of spoken Israeli Hebrew, we took four common criteria for intonation unit boundaries (fast initial speech, slow terminating speech, pitch reset, pauses) and analyzed their occurrences in a segment taken from a spontaneous speech corpus, containing approximately 54 such units. This segment was parsed perceptually by four researchers, and the resultant boundaries were analyzed acoustically to determine which were present at each boundary. A number of interesting conclusions result: only a quarter of the boundaries conformed to all cues, while two boundaries that were agreed upon by all the listeners conformed to none. Final lengthening was most prevalent, followed by pitch reset, then pauses, and finally fast initial speech. A larger study, involving many more units and more speakers is in progress.

1. Introduction

Hebrew has one of the longest recorded histories among languages of the world. The earliest recorded texts go back to the beginning of the first millennium BCE. After over a millennium during which Hebrew was spoken and written, the language ceased to be used as a vernacular, and was used mainly as a literary and liturgical language. At the turn of the twentieth century Hebrew was transformed into a full-fledged vernacular and became the national language of the Jews in Israel when it was established in 1948.

Shaped mainly under the influence of European strata at the beginning of the twentieth century, Israeli Hebrew (henceforth: IH), as compared with earlier stages of Hebrew, has a significantly different structure. IH is still undergoing rapid change because of massive waves of immigration and swift changes in Israeli society, which bring about constant changes in its sociolinguistic and linguistic structure. Having been shaped mainly by the speech of immigrants from eastern Europe, IH shows many structural affinities to Yiddish and some Slavic languages. Prosody is an obvious candidate for such an influence.

The aim of this paper is a rather modest one, and will not endeavor any comparative or explanatory observations. The time is far from ripe for any such research, as IH prosody has not drawn much attention. We seek here rather to make a fresh start in empirical research regarding the basic prosodic units of natural, spontaneous IH, with the hope that it will also draw some more general, theoretical, interest.

2. Premises

The initial premise in our research is that the basic structural unit of spoken language is the intonation unit. An intonation unit (henceforth: IU) is the term used by Chafe for his cognitive model of information flow in discourse, where IU's are constrained by the cognitive processes occurring during verbalization in the minds of both speaker and hearer [4]. The intonation unit, as well as its underlying concept, is not alien to researchers of intonation from other theoretical backgrounds, in this or other terms (e.g., 'intonation group', 'tone unit'), and different descriptions have been used to explicate the concept [2], [5], [9], [10]. It is our stand that the IU is the basic prosodic-syntactical unit of language in its spoken medium, and studying it should be disassociated from its relationship with (traditional) syntax, at least in the first stages of research.

According to Chafe's theoretical stand, the IU is a speech unit that has close association with a "coherent intonation contour", an association which, *mutatis mutandis*, can be found in most descriptions of units of intonation. The problem with this kind of description is that "a coherent intonation contour" is very hard to define, neither is it easy to define an IU by any other internal criteria [5], [14]. Therefore, a commonly held procedure to parse an utterance into IUs is according to their boundaries [5], [6], [7].

The most prominent criteria suggested for the delimitation of an IU are: (1) pause; (2) final syllable lengthening or slow speech rate at the end of an IU, and a following (3) fast speech rate at the beginning of the next IU; (4) pitch reset. While different languages differ in their most prominent cue for delimitation of IUs [11], the following hierarchy has been suggested for IH: (1) pitch reset; (2) cross-boundary change of speech rate; (3) pause [15], [16] (aside from the existence of one of several nuclear tones possible in the language, which, in fact, can be regarded as an internal criterion). However, this suggestion has been based on elicited speech and not on naturally occurring spontaneous speech. Initial findings in IH spontaneous speech give the impression that speech rate, particularly final syllable lengthening, may be higher in hierarchy than the other cues [12]. Transcription of IH spontaneous speech carried out in preparation for the compilation of the Corpus of Spoken Israeli Hebrew (CoSIH) [13] has been hitherto performed perceptually, following, by and large, procedures set forth for transcribing the Corpus of Spoken American English [6], [7]. It has been found, though, that transcribers may differ in their perception of IU boundaries, and it also often occurs that the same transcribers

change their mind when going over the same text after some time lapse. Therefore, we aim at examining the accuracy of IU parsing with quantitative acoustic tools. In this paper we will present our initial findings regarding correlation between acoustic analysis of IU boundaries set against human perception.

3.   Textual Data and Methodology

The analyzed speech is a short narrative of a 27 year old male college student, a native speaker of IH, taken from a free conversation held between him and his father, also a native speaker of IH. We chose to start with a narrative rather than with a conversation, as it includes more sequences of IUs uttered by the same speaker, entailing substantially more continuing IUs than in a common conversation. It is with continuing tones that the majority of perceptual problems occur. The choice of a narrative further facilitates the comparison between perceptual boundaries and their acoustic correlates, since free conversation usually contains numerous instances of broken speech, truncated units, overlaps and is often conducted in a noisy environment. The sound sample with a glossed transcription and a translation can be found in: <http://www.tau.ac.il/humanities/semitic/sp2004.html>.

This sample was given to four native speakers of IH. As the notion of IU is rather hard to be explained to a layperson, these informants were people already familiar, at least to some extent, with this task. The informants were asked to make their parsing without regard to syntactic cues. Their perceptual segmentation was then analyzed acoustically using PRAAT software.

4.   Findings and Analysis

Some 54 units were analyzed. The average of (written) words in an IU is three, ranging in length from one to six words. (Note that IH is a synthetic language, so that the number of written words in a syntactic unit is relatively small.) The average number of syllables per IU is six, ranging from one to fourteen. The average duration of an IU is 0.95", ranging from 0.31" to 1.78".

4.1. Perceptual Findings

The given number of 54 IUs was reduced to 47 IUs, on which there was full agreement among the 4 informants regarding their parsing. Three of the original IUs were excluded from the analytical statistics, since — although the perceptual segmentation was in complete agreement among the informants — they included unclear speech which made the acoustic analysis impossible. Four more strings were also excluded from the statistics, because there was no consent among the informants as regards their parsing into IUs. These are: (1) two truncated units; (2) a string which included in its midst a long hesitation sound, and, in any case, was semantically empty; (3) the string shown in Figure 1.
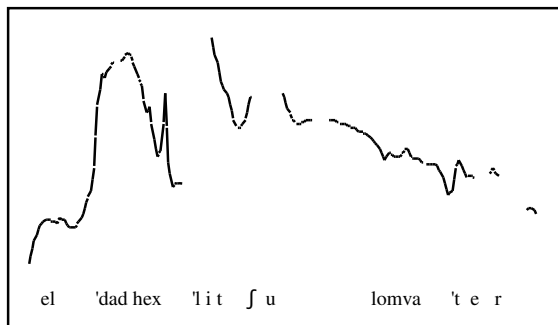
eldad hexlit ʃehu lo mevater

Eldad decided that-he not give-up
"Eldad decided that he would not give up."

Figure 1: *Pitch of IU with different perceptual segmentations.*

The syllable lengthening of *hu* and the fast speed rate of *lo mvater* made two informants interpret this string as two IUs, while the seemingly coherent contour made the other two interpret it as a single IU. We suggest two possibilities of analysis: either it is two IUs joined by sandhi, or one IU with two prominent syllables, viz., *dad* and *lo* respectively. While in the case of *dad* prominence is produced by pitch, duration and intensity together, in the case of *lo* prominence is produced by duration alone (note that the preceding segment includes a filled pause). Either of these two analytical options is theory dependent. Needless to say, the basis for a comprehensive theory regarding IH IUs must be left for future research.

4.2. Acoustic Findings

The four criteria, or cues, mentioned above for detecting IU boundaries were examined quantitatively: (1) fast speech rate at the beginning of an IU (henceforth FSR); (2) final syllable lengthening or slow speech rate at the end of an IU (henceforth SSR); (3) pitch reset; (4) pause.

Measurements were verified independently by two researchers. Among the 47 IUs with perceptual consent among the informants, 11 IUs (24%) conformed to all four cues. On the other hand, two IUs did not conform to even one acoustic criterion. This finding was surprising in view of the unequivocal agreement in the perceptual segmentation. We explain this by the fact that these two IUs have continuing tone and thus form part in a clause complex [9], [10], [17]. Therefore, internal cues might be at play here and should be taken into consideration. Among the remaining IUs, nine IUs had three boundary cues each, among which SSR was found in all nine, and pitch reset was found in eight out of the nine IUs. 17 IUs showed two of the expected cues, among which SSR was attested in 13 IUs and the same number included pitch reset, though not necessarily the same units. A single cue was attested in eight IUs, where SSR was attested in five of them, pitch reset in one, pause in two, while FSR was not found in any IU of this latter group. Table 1 presents the type and number of cues along the sample.

Table 1: *Type and Number of Cues within an IU.*

# of cues	SSR	pitch reset	pause	FSR	total
1	5	1	2	-	8
2	13	13	7	1	17
3	9	8	6	4	9
4	11	11	11	11	11
total	**38**	**33**	**26**	**16**	**45**

Among the six IUs excluded from the statistical analysis only 0-2 of the expected acoustic cues for IU boundaries were found in each unit, notably pitch reset.

We now discuss the individual criteria in detail.

4.2.1. SSR

The criterion of final syllable lengthening or **slow speech rate** at the end of an IU was found in 38 IUs (81% of the entire sample). The calculation of SSR is the ratio between the average of duration of the final syllable of an IU and the average syllable duration in that IU. We indicated lengthening whenever the ratio was >1.1 (i.e., duration of final syllable is more then 10% of the average of syllable duration in IU). The average ratio was 1.95 (SD=0.67). Word stress in IH can be either on the ultima or on the penultima, and, mostly in borrowed words, also on the antepenultima. In six IUs we measured two final syllables due to penultimate stress word. This sample of six IUs was too diverse in data and is obviously too small as to enable any conclusions drawn from it at this point. The average duration of final syllable in the entire sample is 0.24".

4.2.2. Pitch reset

The criterion of **pitch reset** was found in 33 IUs (70%). Pitch falls and rises are basic cues of intonation, but they also form basic cues of word stress [5], [11]. IH has phonemic stress, with overt rhythmic play between stressed and unstressed syllables. An IH word has only a single primary stress, and may have a secondary stress. Function words usually do not carry stress. The place of stress has significant implications on the syllable structure of a lexical or prosodic word [1]. According to a preliminary research conducted in the 1960s [8], all three factors were active in making up word stress in IH, with the following hierarchy: (1) pitch, (2) duration, (3) intensity. Given the rapid and significant changes that IH has gone since the 1960s, and given the difference in research focus, a fresh look at the interrelationship between these correlates is strongly needed. For this reason, our decision on whether to mark pitch reset or not was made only after conducting a random check of frequency differences between stressed and unstressed syllables in the recorded sample. Having found that the significant difference is less then 15 Hz, we decided to indicate pitch reset whenever the difference was >15 Hz.

The average pitch reset (in absolute value) is 51 Hz. Reset downwards was found in 25 IUs (75% among the significant occurrences of this cue). Reset upwards was found in only eight IUs (24%). (We could not detect any significant difference in behavior between falling terminal tones and rising terminal tones regarding pitch reset.)

4.2.3. Pauses

Pauses were measured at the end of each IU. A pause was considered significant if it was at lease 0.02" in duration. This criterion was found in 55% of the sample's IUs, i.e., 26 IUs.

4.2.4. FSR

The criterion of **fast speech rate** at the beginning of an IU was found in 16 IUs (34%; we did not take into account short units of less than four syllables). The calculation of FSR is the ratio between the average of syllable duration preceding the prominent syllable and the average syllable duration in that IU. We regarded an IU with FSR whenever the ratio was < 0.9 (i.e., the average syllable duration before the prominent syllable is less then 90% of the average syllable duration in that IU). The average duration of syllables in the entire sample is 0.16", while the average duration of an FSR syllable is 0.11". The average ratio calculated is 0.76 (SD=0.11).

It must be noted, that anacrusis (in its narrow definition as the number of unstressed syllables at the beginning of an utterance [5]) cannot be regarded — at least at this stage of research — as a significant factor in checking perceptual-acoustic correlates in IU parsing in IH.

In none of its appearances was FSR found to be the only cue. Moreover, FSR appears only as one of 3-4 cues, except for one IU in which it appears with just one other cue, viz., pitch reset. 94% of FSR attestations co-occur with SSR, and, as mentioned above, along with at least one other cue.

When looking at the appearance of SSR at the end of an IU and FSR at the beginning of the following IU, i.e., looking at speech rate as a cross-boundary cue, we found only nine such mutual occurrences (56 % of FSR attestations).

4.2.5. Pitch, Duration and Intensity

A last word is due regarding intensity. We did not examine intensity in this preliminary research, although this will be called for at a later stage when the internal structure of an IU will be studied, notably in order to look for more accurate means for IU parsing. Intensity has been mentioned above as a correlate to both pitch and duration regarding both prominence and stress. Figure 2 illustrates another instance of the correlation between the three prosodic constituents in the internal structure of an IU.
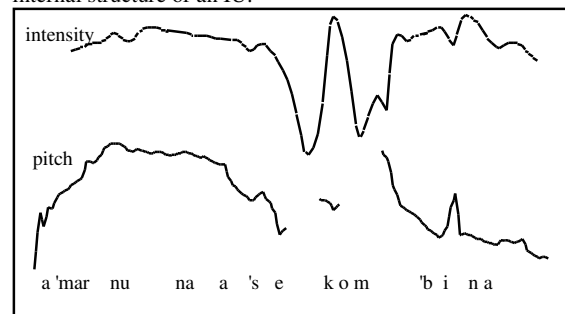
amarnu naase kombina

we-said we-will-do stratagem
"We said, 'Let us plot an intrigue.'"

Figure 2: *Pitch and Intensity correlates in an IU*

The prominent syllable in this IU has been taken by the four informants to be the penultimate syllable, viz., *bi*. Obviously, the pitch contour does not conform to this perceptual judgment. Moreover, the pitch curve as shown may suggest two IUs, as opposed to the perceptual judgements. In contrast, the intensity curve may well add some insight to the perceptual judgment, less to the unity of the IU than to the place of accent. Since the final word in this IU has a penultimate stress, duration of the prominent syllable is less significant in this analysis, especially since its measured length was close to the average syllable duration and therefore seems not to be valid cue neither for an IU boundary nor for the accented syllable. Still, much further research, as initiated by Mixdorff and Amir [18], is needed in order to draw any conclusions regarding the relationship between duration, intensity and pitch and their prosodic roles in IH.

5. Conclusions

Several remarks can be made in conclusion. On one hand, in the three ambiguous IUs , no more than two of the acoustic cues examined here were present. On the other hand, in the 47 agreed IUs, only two were found to lack any cue at all.

The three acoustic cues: final lengthening or SSR, pitch reset and pause were found in more then 50% of the units. SSR and pitch reset together were found in more then 70%, leading us to the conclusion that the perceptual IU segmentation is indeed influenced mostly by its boundaries. Still, much work is to be done comparing duration of stressed syllables and accented (prominent) syllables and assessing their relationship with SSR (cf. [3]). Nevertheless, we cannot avoid the conclusion that research on internal acoustic cues is due, which may explain the remaining 30% of the units.

FSR – albeit its lowest hierarchical status – may suggest its consideration together with SSR as an internal criterion and not an external one.

The frequency of occurrence of the cues studied here compels us to reconsider the hierarchy of cues given in previous intonation research on IH [15], [16].

6. References

[1] Bolozky, Sh., 2003. Phonological and Morphological Variation in Spoken Hebrew. In *Corpus Linguistics and Modern Hebrew: Towards the Compilation of The Corpus of Spoken Israeli Hebrew (CoSIH)*, B. Hary (ed.). Tel Aviv: Tel Aviv University, The Chaim Rosenberg School of Jewish Studies, 119-156.

[2] Brazil D., 1997. *The Communicative Value of Intonation in English*. Cambridge: Cambridge University Press.

[3] Campbell, N., 1993. Automatic detection of prosodic boundaries in speech. *Speech Communication* 13, 343-354.

[4] Chafe, W., 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.

[5] Cruttenden, A., 1997. *Intonation*. 2nd edition. (Cambridge Textbook in Linguistics.) Cambridge: Cambridge University Press.

[6] Du Bois, J. W.; Cumming S.; Schuetze-Coburn S.; Paolino D., 1992. *Discourse Transcription*. (Santa Barbara Papers in Linguistics, 4.) Santa Barbara, CA: Department of Linguistics, University of California, Santa Barbara.

[7] Du Bois, J. W.; Cumming S.; Schuetze-Coburn S.; Paolino D., 1993. Outline of Discourse Transcription. In *Talking Data: Transcription and Coding in Discourse Research*, J. A. Edwards and M. D. Lampert (eds.). Hillsdale, New Jersey: Lawrence Erlbaum Associates. 45-89.

[8] Enoch, P.; Kaplan G., 1969. The Physical Nature of Syllabic Stress in Israeli Hebrew. *Leshonenu* 33: 208-222. (Hebrew)

[9] Halliday, M.A.K., 1989. *Spoken and Wrtitten Language*. Second edition. Oxford: Oxford University Press.

[10] Halliday M.A.K., 1994. *An Introduction to Functional Grammar*. Second edition. London: Arnold.

[11] Hirst, D.J.; Di Cristo, A., 1998. *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.

[12] Izre'el, Sh., In press. From Speech to Syntax — from Theory to Transcription. In *Aaron Dotan AV*, Ch. Cohen and M. Bar-Asher (eds.). (Hebrew)

[13] Izre'el, Sh.; Hary, B.; Rahav, G., 2001. Designing CoSIH: The Corpus of Spoken Israeli Hebrew. *International Journal of Corpus Linguistics* 6, 171-197.

[14] Ladd, D. R., 1986. Intonational Phrasing: The Case for Recursive Prosodic Structure. *Phonology Yearbook* 3, 311-340.

[15] Laufer, A., 1987. *Hangana ('Intonation')*. Jerusalem: Institute for Judaic Studies, Hebrew University of Jerusalem. (Hebrew)

[16] Laufer, A., 1996. Pauses in Fluent Speech and Punctuation. In *Studies in Hebrew and Jewish Languages Presented to Shelomo Morag*, M. Bar-Asher (ed.). Jerusalem: The Center for Jewish Languages and Literatures, The Hebrew University of Jerusalem and The Bialik Institute, 277-294. (Hebrew)

[17] Miller, J.; Weinert, R., 1998. *Spontaneous Spoken Language: Syntax and Discourse*. Oxford: Oxford University Press.

[18] Mixdorff, H.; Amir, N., 2002. The Prosody of Modern Hebrew - A Quantitative Study. *Proceedings of Speech Prosody 2002*, 511-514.