The Acoustic Realization of Lexical Stress in Hebrew Speech

Noam Amir and Hagit Sagi
Dept. of Communication Disorders
Speech, Language and Hearing
Sackler Faculty of Medicine
Tel Aviv University
noama@post.tau.ac.il.com

Abstract—Lexical stress is an omnipresent phenomenon in spoken language, serving in many cases to disambiguate lexically identical words. Lexical stress in Modern Spoken Hebrew (MSH) is usually either penultimate or ultimate, however its prosodic markers have not been studied extensively. In the present study the acoustics of lexical stress were characterized in detail. A list of 34 two-syllable words was employed, where each word had different meanings when pronounced with penultimate or ultimate stress, e.g. "TOfes" (formulaire) vs. "toFES" (catches). 136 carrier sentences were devised, in which each word appeared four times: twice in the middle and twice at the end, with penultimate stress and ultimate stress in each position. Thirty speakers (15 female and 15 male) uttered each sentence, giving a total of 4,080 utterances. Pitch contours were calculated with Praat software and corrected manually, and all syllable boundaries of the target words were annotated manually also. Statistical analysis revealed that syllable duration was the dominant marker of Hebrew lexical stress, regardless of word position. Pitch contours were strongly affected by word position and played a very minor role in indicating stress. These results have important implications in Text to Speech applications, where generating the correct prosody has a large impact on naturalness. The study will be extended in the future to longer words that are not necessarily disambiguated by stress, in order to determine if the same mechanisms are present in such cases.

Index Terms—lexical stress, prosodic markers, Modern Spoken Hebrew

I. INTRODUCTION

Speech sounds are of two levels: the segmental level and the suprasegmental level. While the segmental level refers to phonemes, and their incorporation into words and utterances in a given language; the suprasegmental sound patterns are "a vocal effect which extends over more than one sound segment in an utterance, such as a pitch, stress or juncture pattern" (Crystal 2008, 446). According to McQueen and Cutler (2010), listeners make use of suprasegmental patterns first and foremost to parse speech into words: "the central knowledge store for speech perception is the mental lexicon, that is, our stored representations of words" (ibid., 489).

Syllables in words are not identical with respect to their perceptual prominence. At the word-level, some syllables are more prominent than others. In stress languages, like Hebrew, Arabic and English, each word has one syllable which has the maximal prominence. This is the syllable is called the stressed syllable. It should be mentioned that stress is relative, and that its relativeness is measured with comparison to the other syllables in the same word. For example, the stressed syllable in the word boker 'morning' is the first syllable (marked with boldface characters), while in yalda 'girl', it is the second syllable that is stressed. In Hebrew, mainly in nouns, the stressed syllable can be the last (called ultimate stress), the one before the last (called penultimate stress), or the third syllable from the last (antepenultimate stress). Moreover, in Hebrew there are identical words, with respect to their phoneme sequence, that differ in meaning only because of the stress location. For example: BOker 'morning' vs. boKER 'cowboy'. This type of minimal pair will be examined in the present research. Such pairs consist of two bisyllabic words, such that one syllable is stressed and the other is unstressed.

The present research focuses on the prosodic organizational level of words, and the effect of the word location in a sentence, meaning the effect of the intonation level contour on the stress characteristics. During speech, speakers articulate a sequence of speech segments (vowels and consonants), but their organization (regrouping and delimitation) is accomplished via suprasegmental features, such as duration (and rhythm, in consequence), fundamental frequency (F0), and intensity.

This research will be presented as follows: In Section I.1 the theoretical framework of prosodic hierarchy is presented, and in I.2 an overview of Hebrew stress systems, which often employ either ultimate, final, stress or penultimate stress, is described. Section I.3 then describes the acoustic characteristics of stress. Section II is dedicated to the methodology, regarding the relation between the two examined units: the word and the utterance. Hence, the sentence formulation and the acoustic measurements that were performed are described. In section III the results are presented. Summary and conclusions are in section IV.
I.1 The prosodic hierarchy

Prosodic features can be seen primarily as the result of laryngeal or sub-glottal activity. Tone and intonation are controlled by laryngeal muscles and accentual features are attributed to the respiratory muscles, while segmental features are primarily associated with the supralaryngeal component (except the voiced-unvoiced feature). Thus, prosody organizes structures that measure chunks of speech into countable units of various sizes (Fletcher 2009, 523). Lexical stress, for example, is realized in MSH on the word level, while the prominent syllable is realized on a higher level of the prosodic hierarchy – it is a prosodic cue which signals the important word in an utterance. An abstract model of the prosodic hierarchy, as is relevant for MSH phonology, is in Figure 1.

<table>
<thead>
<tr>
<th>Highest Unit</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>Intonational Phrase</td>
</tr>
<tr>
<td>PhP</td>
<td>Phonological Phrase</td>
</tr>
<tr>
<td>PrWd</td>
<td>Prosodic Word</td>
</tr>
<tr>
<td>Ft</td>
<td>Foot</td>
</tr>
</tbody>
</table>

| Lowest Unit | Syllable |

Figure 1: The prosodic hierarchy (simplification of the illustration in Fletcher 2010, 529)

I.2 The Stress System of Hebrew

Hebrew word types can be distinguished by syllable structure and stress pattern. Certain syllable structures and stress pattern exist only in the verb system and in loan words (i.e. CVC(C)VC, as in katav 'you.F.SG wrote') while other structures exist only in (original Hebrew) nouns (CVCC, as in vered 'rose', and noam 'pleasantness'). Yet, in MSH certain noun patterns have special stress assignment, creating semantic differentiation from their "regular" counterparts. For example: rehovot 'Rehovot (the city)' vs. rehovot 'streets', respectively. Such minimal pairs are the focus of the present research. Apart from "special" vs. "regular" noun patterns, minimal pairs that rely on stress assignment only, exist also between nouns and verb in the present tense (Benoni), as in tofes 'form' and tofes '(He) catches', respectively.

Generally speaking, stress in Hebrew nouns may reside on any one of the last three syllables, ultimate (as in simla 'dress'), penultimate (as in tilboshet 'costume'), or antepenultimate (as in telefon 'telephone'). Antepenultimate stress is relatively rare, found mostly in borrowed nouns). Moreover, the position of stress is not phonologically conditioned, as it does not distinguish between syllable structures (unlike that of Arabic and Tiberian Hebrew). In other words, "Hebrew stress is quantity-insensitive, as expected in languages with no contrastive vowel length." (Adam and Bat-El, 2009: 257).

The dominant stress pattern of the Hebrew verb system is the ultimate stress, except for certain verb groups that have ultimate stress due to morpho-phonological processes (Schwarzwald 2002: 32-33). The dominant stress pattern of Hebrew nouns is also ultimate stress (84%. Schwarzwald 2002: 32-33).

I.3 The acoustics of lexical stress

As to the acoustics of stressed syllables in MSH, little quantitative research was done on this matter. Enoch and Kaplan (1969) measured stress in isolated word pronunciation and showed that the feature that distinguished between stressed and unstressed syllables was \( f_0 \), while syllable duration and intensity were found as a distinctive feature only in few cases.

Most (1999) examined the production and perception of syllable stress by children (10-13 years) with normal hearing compared with those of children with hearing impairment. The stimuli consisted of two recordings, the target word within a carrier sentence, and the target word in isolation. The production results showed that both groups of children produced stress similarly; they produced the stressed syllable with higher \( f_0 \), longer duration, and higher intensity, compared to the unstressed syllable.

In Becker (2003) it is reported that vowels in stressed syllables are phonetically twice as long as vowels in unstressed syllables (Becker 2003). In addition, it is said that the stressed syllable has a higher pitch when in phrase final position. Otherwise, the high pitch falls on the posttonic syllable (Becker 2003). Nonetheless, Becker does not give evidence to his claims. Further to Becker’s (2003) report, Adam and Bat-El (2009) claim that these phonetic characteristics of Hebrew stress are found in both adult and children’s speech (p. 257).

II. METHODOLOGY

II.1 Utterances

34 bisyllabic minimal pairs differing in their stress pattern served as stimuli within carrier sentences. The words consisted of familiar nouns and verbs. Both the ultimate stress word and the penultimate stress word are located in medial and in final positions of the sentence. An example of a four sentence set is as follows:

1. Ultimate stress in medial position: arey bira shonot mitkashatot lahag 
2. Ultimate stress in final position: yatsanu letiyul meurgan learey bira
3. Penultimate stress in medial position: shatiti bira hapab hasamuch

Overall each participant uttered 136 sentences.

II.2 Participants

30 subjects aged 28-35 (Mean: 32, ST: 2.53), 13 Men and 17 Women, participated in the study. All were native Hebrew speakers without any known speech or hearing problems.

II.3 Recordings

The subjects received the sentences in written form, and were instructed to read them out loud, pausing for at least two seconds between sentences. An experienced research assistant was present during the recordings, and requested the subjects to repeat the utterance whenever a mistake was made. All
recordings were carried out in a quiet room, using an Audio Technica AT892 head mounted microphone, through a Centrance Microport pro audio interface to a personal computer. Sampling frequency was 44,100 Hz throughout.

II.4 Data preparation
The entire corpus, comprised of 4,080 utterances, was annotated manually by an experienced research assistant. The target word in each utterance was segmented into syllables. F0 was calculated using Praat software, and corrected manually. Annotation and F0 data was saved in a text format, and custom written Matlab programs were written to collect this data over the entire corpus.

III. RESULTS
Three prosodic parameters were calculated for each syllable: duration, intensity and F0. Two general types of comparisons were then carried out. Intrinsic comparisons were comparisons between the stressed and unstressed syllables of a given word. Extrinsic comparisons were comparisons of the same syllable in the minimal pair words. For instance, we compared the duration of the first syllable in the medial penultimate-stressed word and medial ultimate-stressed utterances of the same word. Due to the large number of such comparisons, only a subset of the results will be presented here.

III.1 Duration
A methodological difficult confronting the issue of duration was that different syllables could be expected to have a-priori different durations. This is mainly due to differences in consonant durations, between plosives, fricatives, etc. In the future we intend to annotate the utterances on the phoneme level, in order to be able to compare vowel durations only. For the moment, however, this biased the intrinsic comparisons, thus the extrinsic comparisons are more useful. One way to overcome this obstacle is to compare the differences between first and second syllable duration over different stress patterns, instead of their absolute value.

Figure 2 shows an extrinsic comparison of duration of the first syllable in the medial position. The mean duration over all speakers is given, for each word (numbered on the x-axis) separately. Figure 3 shows the same comparison for the final position. A characteristic variation in duration can be seen for every one of the 34 words. However, it can be seen quite clearly that the first syllable is lengthened when it is stressed, i.e. when stress is penultimate. Individual t-tests were carried out for each of the words. The difference in mean duration was found to be significant for 32 out of the 34 words in the medial position, and 25 of the 34 words in the final position. From these figures it is clear that when the word has penultimate stress, the first syllable is lengthened in comparison to the same word with ultimate stress. Similar results were obtained in extrinsic comparisons of the second syllable durations.

Figure 4 shows an intrinsic comparison of durations, through the differences between durations of the first and second syllables, for words in the medial position.

The difference was found to be significant in 33 out of the 34 words. Clearly, the difference was larger in the ultimate-stressed words. Once again, this strengthens the conclusion that the stressed syllable is lengthened in comparison to unstressed. The same results were observed for words in the final position.

III.1 Intensity
To compare intensity, RMS values were calculated over the voiced parts of each syllable only. Figures 5 and 6 show intrinsic comparisons of intensity, through the ratio of S2/S1 intensities, in the medial and final positions.

The differences were found significant for 30 words in both cases.
In the medial position it can be seen that almost uniformly, the stressed syllable has a higher intensity, since the upper line is nearly always above 1, and the bottom line below it. In the final position, however, the situation is more complicated, due to the fact that there is an overall trend towards lower intensity towards the end of an utterance. Thus the upper line (ultimate) is higher than the bottom one (penultimate), however the two do not fall neatly above and below the ratio of 1.

III.3 F0

Results for F0 were less conclusive than those for duration and intensity, and will be presented only briefly. Extrinsic comparisons of mid-vowel F0 of the first syllable, for medial and final positions are presented in figures 7 and 8. In the medial position, F0 appears to be a clear extrinsic marker of stress, and it is significant for 32 words. However, this tendency is less pronounced in the final position.

I. SUMMARY AND CONCLUSIONS

The results presented here shed light on the debate over the prosodic markers of lexical stress in Hebrew. The importance of such markers can be examined in either extrinsic or intrinsic comparisons, and the results denote a certain hierarchy. It appears that duration and intensity are the strongest indications of stress, whereas F0 is mainly an extrinsic marker of stress but not an intrinsic one.

REFERENCES