

CATEGORICAL PERCEPTION OF LEXICAL STRESS: THE EFFECT OF MANIPULATED DURATION

Noam Amir¹, Chen Ben-Chemo¹, Vered Silber-Varod²

¹Dept. of Communication Disorders, Sackler Faculty of Medicine, Tel Aviv University

²The Research Center for Innovation in Learning Technologies, The Open University of Israel

noama@post.tau.ac.il, chenb136@gmail.com, vereds@openu.ac.il

ABSTRACT

A recent production study found that duration is the main cue in realization of lexical stress in Hebrew. In order to examine this perceptually, two minimal pairs of bisyllabic words differing only in their stress patterns were uttered within a carrier sentence. Target words were then extracted and their vowel durations were manipulated in eight steps, either downward or upward, to form smooth transitions from original penultimate duration patterns to ultimate, and vice versa. These stimuli were then presented to 15 listeners. Results show that for most listeners, the changes in duration were sufficient to cause a categorical change in perception of the stress pattern.

Keywords: lexical stress, categorical perception, duration, manipulated stimuli, Hebrew.

1. INTRODUCTION

Lexical stress is produced by prominence of one syllable in a polysyllabic word. This prominence can be realized by prosodic means - Duration, intensity, F0, as well as spectral cues such as vowel formant frequencies. The order of importance of these acoustic cues is language-dependant and has been studied extensively in the past (inter alia, [8] for English; [9] for Swedish; A summary of lexical stress research is presented in [5])

The theoretical question that underlies the current study, and perceptual studies in general, is how does the listener decode the stream of speech sound to reconstruct the speaker's message? It is of interest to ask which aspects of the sound patterns are the "important" ones, the ones used by listeners to identify lexical stress patterns, especially in languages, such as Hebrew, in which "stress is contrastive and may distinguish between the meanings of otherwise identical words" [4, p. 625].

A recent study [1] on the *production* of lexical stress was carried with over 2000 utterances of minimal pairs in Hebrew. Results showed that speakers typically used duration, intensity and F0 to some extent, however duration was the strongest and most regular cue. This raises the question if duration in itself is a sufficiently strong cue to affect

perception of lexical stress in Hebrew. This should be viewed in light of the fact that Hebrew is not a quantitative language, i.e. duration is not a distinctive phonemic feature.

In a study on the effect of manipulated duration in English, Fry [7] showed that when the two parameters amplitude and duration were manipulated in 5 steps and F0 in 16 steps, judgments stress showed more effect of the durational than of the amplitude manipulation; change in F0 had an all-or-none effect. Moreover, when vowel quality judgments were compared to the above mentioned tests of durational and amplitude manipulations, it was found that the latter parameters were "more closely related to stress judgments than this type of vowel change." [5, p. 265].

1.2 Categorical perception

Categorical perception is a psycho-physical phenomenon which occurs when substantial variations lead to a shift of perception in category, however intra-category variations are ignored [10]. An acoustic cue is defined as categorical when: 1) A boundary point exists. This point is referred as "the 50% point", and it represents a relatively narrow ambiguity zone; and 2) When there are high identification rates for each of the categories at the two extremes of the scale.

Categorical perception phenomenon was explored in dozens of studies with respect to consonant discrimination (e.g., place, voicing, manner, etc.), but also in a variety of vowel and syllable contexts. Albeit, in [10] it is reported that the pattern of discrimination of vowel duration is *continuous* and not categorical. Among the first who reported on categorical perception of durational cues in speech was Denes [6], who discovered that due to voice onset time (VOT) effect, there is a categorical perception of the vowel duration between the noun *use* [jus] and the verb use [juz].

2. RATIONALE AND OBJECTIVES

This research was motivated by the findings of [1], which was a production study of the acoustic characteristics of lexical stress. It found that duration

is the most dominant and stable parameter in production of stressed syllables. However as yet no research has examined the importance of duration to the *perception* of lexical stress. In fact, no study on the perception of IH (Israeli Hebrew) stress-unstressed alternation in minimal pairs has been published to date.

Therefore, the objective of the current study was to determine if vowel duration is sufficient to cause a categorical shift in perception of stress in IH.

3. METHODS

3.1 Stimuli

Within a previous study [1], sentences embedding the target words – lexical stress-based minimal pairs – were given to the subjects in written form, and they were instructed to read them out loud, pausing for at least two seconds between sentences. All recordings were carried out in a quiet room, using an Audio Technica AT892 head mounted microphone in order to maintain a relatively constant mouth to microphone distance. Digitization was performed with a Centrance Micport Pro audio interface to a personal computer, maintaining constant gain levels throughout all the recordings. Sampling frequency was 44,100 Hz throughout.

In the present research only recordings of two minimal pairs were selected: the Hebrew words /'oto/ 'car' and /o'to/ 'him' uttered by a male speaker, and the words /'tofes/ 'formulaire' and /to'fes/ 'catches' uttered by a female speaker. The words were chosen to represent two different syllabic patterns: CV.CVC and V.CV. No voiced consonants are involved, and each word in each pair is from a different lexical category (two nouns, one verb, and one inflectional preposition). Vowel durations and duration ratios for these recordings are shown in Table 1.

Several differences in vowel duration can be found between the two minimal pairs: In the case of the [oto] pair, the stressed vowel is about twice as long as the unstressed vowel. On the other hand, in the [tofes] pair, the second vowel (V2) is almost the same length in both stress patterns (93ms in the penultimate pattern, and 98ms in the ultimate pattern). The difference between the stress patterns in this case is mainly in the first vowel (V1). Nonetheless, above 95% of the listeners identified the words in the [tofes] pair correctly; In the [oto] pair the identification rates varied between 84% for /'oto/, and 97% for /o'to/.

Table 1: Durations (milliseconds) of vowels in the original recorded words and of vowels post-manipulation morphing.

Word		V1(ms.)	V2(ms.)	V _s /V _u
/'oto/	Original:	89	37	2.4
	±morphing:	53	165	3.1
/o'to/	Original:	64	138	2.1
	±morphing:	107	29	3.7
/'tofes/	Original:	77	93	0.82
	±morphing:	29	117	4
/to'fes/	Original:	36	98	2.7
	±morphing:	92	77	1.2

Vowel durations were then manipulated using Praat software [3] in order to shift from the duration values of the penultimate stress words (/ 'oto/, / 'tofes/) to those of the ultimate-stress words (/o'to/, /to'fes/) and vice versa. Ideally, the vowel durations of one form of the word (/ 'oto/ or / 'tofes/), should be replaced with the duration of the paired forms (/o'to/ or /to'fes/).

However, since the resultant effect was initially unknown, the target durations were exaggerated, in shortening and lengthening by approximately 20% beyond the original values. Morphing was performed in eight equal steps, for each word, giving overall nine different stimuli for each of the four recorded words: The original recording and eight morphed versions. Table 1 presents the original and morphed duration values of each word.

3.3. Procedure

The experiment was conducted in four blocks, one for each original recording and its morphed versions. In each block, the nine stimuli were presented to the listeners in eight repetitions, for a total of 72 presentations per block, all in random order.

The listeners were asked to mark which word of the minimal pair they perceived: The one with penultimate-stress or with ultimate-stress. The choices were presented both as a picture and in writing (Figure 1). It is important to note that in both minimal pairs used here, the Hebrew orthography for each of the words is different, thus the written versions are unambiguous. Pictures were added nonetheless, to make the task as clear as possible.

3.2 Participants

Fifteen women aged 24-31 years (M=26, STD=1.77) participated in the study. All participants were native Hebrew speakers, with no known hearing or attention deficiency problems. All of them were either graduates or students of a higher education program.

Figure 1: Illustration of the two screen displays.

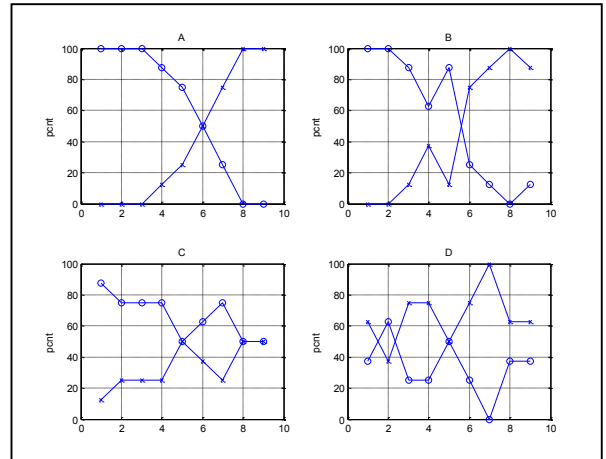


4. RESULTS

For each of the nine stimuli in each block the number of times (out of 8 presentations) it was identified as penultimate/ultimate was tabulated and calculated as a percentage. This can be plotted in the classical form for inspecting categorical perception, where the x-axis denotes the degree of morphing (nine points in the present research) and the y-axis denotes the recognition percentage of ultimate and penultimate. This resulted in 60 such graphs overall: Four graphs for each of the 15 participants. The graphs were then sorted into four prototypes:

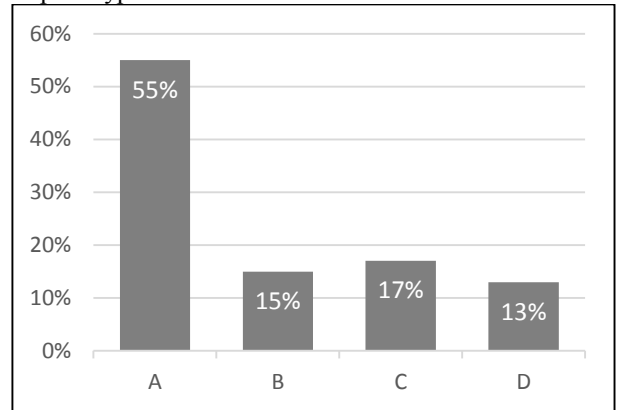
- A "classic" categorical perception with a clear crosspoint and at least 80% identification for each category (category A in Figure 2).
- When categorical perception occurred, but with a region of uncertainty in the centre.
- When one category was perceived, but not the other.
- Uncertainty throughout, or did not switch at all (category D).

Figure 2: The four perceptual prototypes.



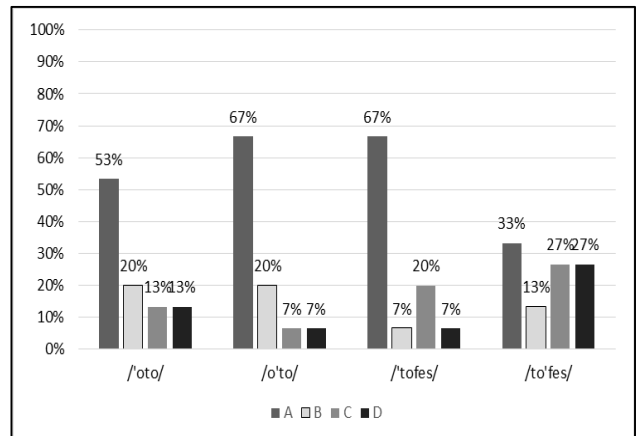
The percentage of graphs in each category, is presented in Figure 3. Summing the results for category A and B, a clear shift in stress occurred in 70% of these graphs.

Figure 3: Occurrences (%) of each perception prototype.



A further breakdown for each of the four blocks is presented in Figure 4, where it is demonstrated that categorical perception (types A and B) is above 70% in three out of the four words. However, for the word /to'fes/ only 46% of subjects perceived the changes in duration categorically.

Figure 4: Percentage of each perception prototype for each word.



5. DISCUSSION

The present study presents the effect of manipulating the duration of vowels on the perception of lexical stress in two IH minimal pairs. Subjects were asked to identify the word they heard, i.e., to perceive whether the stress pattern is penultimate or ultimate, for nine stimuli – the original recorded word and eight other stimuli that were manipulated in eight equal steps.

The findings suggest that there are in fact four perceptual prototypes that can be sorted according to the following: Uncertain judgment or no change in judgement at all; Certainty in the perception of one word form, but not the other; categorical perception, but with a region of uncertainty in the centre; and categorical perception with a clear crosspoint and at least 80% identification for each word.

"Clear" categorical perception was found to be the most common type (55%) in general. When looking at the perceptual behaviour of each word, the "clear" categorical perception type is again dominant (53%-67%). In fact, perceptual prototypes A ("with a clear crosspoint") and B ("with a region of uncertainty in the centre") can both be considered as categorical perception, thus making 73%-87% of the judgments as categorical in three out of the four words. The difference in the perceptual behaviour of one word form – /to'fes/ is evident and is may be ascribed to the morphing process, which produced low manipulated ratio (1.2, see Table 1) between the two vowels of the word.

It will be of interest to look at the crosspoint locations in the two directions of manipulation – lengthening and shortening – in order to see if there is any perceptual tendency towards one of the forms. As reported in [2] for monosyllabic minimal pairs in Thai, the zone of ambiguity was found later for the original short vowel and earlier for the original long vowel.

To conclude, duration is a major cue for the perception of stress pattern in Hebrew. Moreover, this prosodic cue is perceived categorically, in over 70% of the cases.

Further tests on the perception of lexical stress should be done with other minimal pairs, and with manipulation of other prosodic cues (F0 and intensity) in order to evaluate their relative perceptual importance in comparison to that of duration.

7. REFERENCES

[1] Silber-Varod, V., Sagi, H., Amir, N. The acoustic correlates of lexical stress in Israeli Hebrew. Submitted.

- [2] Abramson, A. S., Ren, N. (1990). Distinctive vowel length: duration vs. spectrum in Thai. *Journal of Phonetics*, 18(1), 79-92.
- [3] Boersma, P., Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4.04, retrieved 28 December 2014 from <http://www.praat.org/>
- [4] Cohen, E. G., Ussishkin, A. 2013. Stress: Modern Hebrew. In G. Kahn (ed.), *Encyclopaedia of Hebrew Language and Linguistics (EHL)*, Vol. 3, Leiden/Boston: Brill, 625-628.
- [5] Cutler, A. (2005) Lexical stress. In D. B. Pisoni and R. E. Remez (eds), *The Handbook of Speech Perception* (pp. 264-289), Blackwell.
- [6] Denes, P. (1955). Effect of duration on the perception of voicing. *The Journal of the Acoustical Society of America*, 27(4), 761-764.
- [7] Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, 27(4), 765-768.
- [8] Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, 32(4), 451-454.
- [9] Sluijter, A. M. C, Van Heuven, V. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100, 2471-2485.
- [10] Strange, W. (1999). Perception of consonants: From variance to invariance. In J. M. Pickett, *The acoustics of speech communication. Fundamentals, speech perception theory and technology*. Allyn and Bacon, 166-182.