A duration-based solution to the problem of stress realization in Turkish

Anne Pycha
University of California, Berkeley
Department of Linguistics
pycha@berkeley.edu

I. Introduction

In this paper, I argue that the realization of Turkish stress makes crucial reference not to syllable types (such as heavy versus light), but to syllable durations. In agreement with previous work, I show that fundamental frequency (F0) acts as a primary acoustic correlate of stress in the language. Contrary to what we might expect, however, F0 contours realize themselves fully on syllables closed by a sonorant consonant (abbreviated here as CVR), but are "clipped" on syllables with long vowel nuclei (CVV). This presents a puzzle for any theory of phonological weight, whereby CVV syllables should certainly be heavier, and therefore more capable of hosting prosodic information, than CVR syllables. The puzzle resolves itself when we examine phonetic duration: the F0-carrying portion of CVV syllables is in fact consistently shorter than that of CVR syllables. Following Gordon (1999), these findings suggest that exceptions to typological tendencies in syllable weight can find a principled explanation within phonetic implementation.

2. Background

Turkish has at least three ways of assigning stress (Sezer 1981, Inkelas & Orgun 1998, Inkelas 1999, Kabak & Vogel 2001, Inkelas & Orgun 2003, and references cited therein). Final stress is the default. Certain morphemes, however, condition the presence of non-final stress; an example is the negative morpheme *-mA,* which conditions stress on the syllable preceding itself. Place names undergo a totally different stress placement algorithm, dubbed "Sezer stress" after Sezer (1981): here, stress falls on the antepenultimate syllable if it is heavy and the penultimate syllable is light, otherwise it falls on the penultimate syllable.

(1a) git-'ti-m 'I went'
go-PAST-1SG

(1b) 'git-me-di-m 'I didn't go'
go-NEG-PAST-1SG

(1c) 'ankara, is'tanbul 'Ankara', 'Istanbul'

In this paper, I will not consider Sezer stress, but will be concerned only with final and non-final stress.

Previous researchers who have investigated the acoustic correlates of Turkish stress have concurred that F0 acts as a primary correlate (Konrot 1981, Konrot 1987, Levi 2005), but have also noted that final and non-final stress seem to get realized very differently. Konrot (1981) identified both F0 and intensity as correlates, but only in words with non-final stress. In words with final stress, he found no robust correlates at all. Levi (2005) identified F0, intensity, and duration as correlates, when averaging over words with final and non-final stress. But she too encountered the non-robustness of correlates for final stress, as revealed by her algorithm for determining the stressed syllable on the basis of acoustics: "In focused position, the target word will either have the accent on the final syllable of the word or on the syllable preceding the drop in pitch" (2005).

The lack of clear acoustic correlates might suggest that final stress is nothing more than a percept for Turkish listeners. But Konrot's (1987) follow-up study showed no bias toward final syllables in synthetic, monotonous stimuli. So it seems that *some* correlate must be present for final stress in Turkish, but that the exact conditions under which it occurs, or fails to occur, warrant a closer investigation.

I take up this investigation here, in three parts. First I show that final and non-final stress are indeed realized differently. I then show that F0 is the primary acoustic correlate for final position, even though its presence there is diminished compared to its presence in non-final positoin. Finally, I evaluate the hypothesis that certain syllables are more capable of hosting F0 differences in final position, and show that it is phonetically long syllables (and not phonologically heavy ones) which meet this criteria.

3. The current study

The first part of the current study shows that F0 is a primary correlate of stress in Turkish, even in final syllables. While other researchers have reported this finding, my work is the first to make a direct comparison between stressed syllables that differ *only* in the source of stress and not in segmental composition or location of the stressed syllable, a crucial control for the question that concerns us.

3.1 Correlates of stress in final and non-final position

To measure potential correlates of stress, I designed a list of paired words. Each member of the pair has stress on the second syllable, but the source of stress differs.

2)

Location: final Source: default		Location: non-final Source: pre-stressing morpheme *mi*	
be.'be	'baby'	be.'be.mi	'a/the baby?'
tʃe.'tʃe	'type of insect'	tʃe.'tʃe.mi	'a/the kind of fly?'
de.'de	'grandfather'	de.'de.mi	'a/the grandfather'
pe.'pe	'stammerer'	pe.'pe.mi	'a/the stammerer'

The words in the first column are simplex nouns, from Konrot (1981). Their first and second syllables are segmentally identical. These bare nouns have stress on the second syllable, which the grammar assigns by default because no other morphological source of stress is present. The words in the second column are suffixed nouns. They too have stress on the second syllable, but here a morphological source of stress is present, namely the pre-stressing suffix *-mi* (INTERROGATIVE).

The word list thus allows for a direct comparison between stress types. By examining, for example, the second syllable of *be.'be* and the second syllable of *be.'be.mi* we can compare F0, intensity, and duration measurements for syllables that share both segmental composition and stressedness, but differ in source of stress. This controlled approach ensures that any differences in phonetic measurements stem primarily from differences between final and non-final stress. This approach differs from that used in Konrot (1981) and Levi (2005), both of whom compared syllables which differed along more than one dimension. Konrot (1981) compared words such as *yaz.'ma* ‘writing’ and *'yaz.ma* ‘don’t write!’ which, although they illustrate that Turkish possesses minimal stress pairs, still entail the comparison of syllables with totally different segmental compositions. Levi (2005) compared words such as *de.niz.'de* ‘in the sea’ and *de.'niz.le* ‘with the sea’ which pose the same problem.

The pre-stressing suffix that I employ in the word list, *-mi* (INTERROGATIVE), presents significant advantages over other pre-stressing suffixes in Turkish. It can be suffixed to almost any word, and its sonorant onset allows for continous tracking of F0, which will be relevant when we look more closely at F0 contours in the next section. Most importantly, *-mi* does not trigger glide epenthesis when it attaches to vowel-final roots. Almost all other pre-stressing morphemes do: *be.'bej.le* ‘with a/the baby’, *be.'bej.di* ‘it was a/the baby’, *be.'bej.se* ‘if a/the baby...’. This epenthesized glide is quite difficult to segment in a principled way from the neighboring vowel, and so could easily introduce unwanted variation into the acoustic measurements. By contrast, the morpheme *-mA* (NEGATIVE) does not trigger glide epenthesis, but it attaches only to verbs. Most verbs in Turkish are monosyllabic, which makes combinations of a verb plus *-mA* less than ideal for our purposes because the resulting stressed and unstressed syllables lie in different morphemes (*'yap-ma* ‘don’t do (it)!’), again confounding the analysis.

The only potential drawback of the suffix *-mi* is that, as an interrogative, it may introduce a question intonation. But a follow-up analysis that I conducted with the pre-stressing suffixes *-le* (COMITATIVE) and *-di* (PAST) produced results which suggest that we can generalize on the basis of *-mi*.

The target words, along with a set of filler words, were embedded in focused position within carrier sentences. The order of presentation was randomized. Five native speakers of Turkish (two female, three male) recorded the word list in Istanbul, Turkey, using a head-mounted microphone and speaking at a natural rate.

3)

Final stress	Non-final stress	Repetitions	Speakers	Total
4 tokens	+ 4 tokens	x 2	x 5	= 80 tokens

3.1.2 Measurements

The target words were segmented using waveform displays and spectrograms produced by Praat acoustic analysis software (Boersma & Weenik 2003). Aspiration, when present, was segmented as part of the preceding consonant. For each vowel in each token, I measured the F0, intensity, duration, first formant (F1), and second formant (F2).

The F0 was measured in two ways. In the first method, replicating Konrot’s (1981) approach, I calculated F0 by averaging over the F0 found in the initial sonorant portion of the vowel and in the final portion. In the second method, I measured only the peak F0 after discarding the first quarter of the vowel duration, which is typically prone to segmental disturbances from the preceding consonant.

The intensity was measured by taking the peak intensity after discarding the first quarter of vowel duration. F1 and F2 were measured at the vowel midpoint, with nothing discarded. Although no previous research has indicated that unstressed vowel reduction occurs in Turkish, this does not exclude the possiblity that F1 and F2 have a qualitative effect, even if subtle, in cueing stress.

3.1.3 Results

Overall, the analysis of phonetic measurements indicates that stress has multiple correlates, in both positions. In accord with previous work, the force of F0 as a stress correlate is clearly diminished in final position, when compared to non-final position. Despite this diminished status, however, F0 plays the primary role in final position when compared to other correlates; intensity and intensity play only a secondary role, if any role at all.

Stressed versus unstressed syllable

To determine the acoustic correlates distinguishing stressed from unstressed position, a statistical comparison of each correlate on the stressed vowel (e.g., the second syllable of *be*.'*be*) with that same correlate on the unstressed vowel (e.g., the first syllable of *be*.'*be*) was conducted. Paired t-tests revealed significant differences (p<0.05) between stressed and unstressed syllables, for most correlates, in both final and non-final position.

4)

	F0	Intensity	Duration	F1	F2
Final	✓	✓	✓	×	✓
Non-final	✓	×	✓	✓	✓

✓ = significant difference between stressed and unstressed syllables
× = no significant difference

Analysis: paired t-test, with each measurement compared in stressed and unstressed syllables within the same word.

Final versus non-final position

Although multiple correlates for stress are present in both final and non-final syllables, the force of these correlates in each positions is not equivalent. We can see this by calculating measurement ratios of the stressed syllable to the unstressed syllable, and comparing the ratios across final and non-final position, as schematized below for F0.

Final stress:	*be*.'<u>*be*</u>	F0('<u>*be*</u>)/F0(*be*) = ratio for final position
Non-final stress:	*be*.'<u>*be*</u>.*mi*	F0('<u>*be*</u>)/F0(*be*) = ratio for non-final position

The ratios, averaged over all of the tokens, reveal that the realization of final and non-final stress does not differ significantly for duration and intensity, but it does differ for F0, and the formants F1 and F2.

5)

	F0	Intensity	Duration	F1	F2
Ratio '<u>CV</u>/CV	✓	×	×	✓	✓

✓ = significant difference between final and non-final positions (p<0.05)
× = no significant difference

Analysis: ANOVA with dependent variable = ratio, and factor = stress type (final or non-final)

The average F0 ratios for final and non-final position are shown below.

6) *Ratio 'CV.CV, for F0*

	Final	Non-final
Method 1: Averages $(F0_{start} + F0_{end})/2$	1.16	1.32
Method 2: Peaks Maximum in last ¾ of vowel	1.17	1.36

In other words, according to Method 2, the number of cycles per second (Hz) on stressed syllables is about 17% greater than that on unstressed syllables, for final position. For non-final position, the number of cycles per second on stressed syllables is 36% greater than on unstressed syllables.

These results thus confirm the findings previously reported in the literature. Because the current study carefully controlled for both segmental content and stressed syllable location, however, we can now be certain that the differences between stress realization in final and non-final position really are due to position alone, and not external factors.

A closer look at final position

Recall that our puzzle centers around the realization of stress in final postion, for which previous work found only non-robust correlates. Although the stastical analyses conducted here indicate that F0, intensity, duration, and F2 all distinguish final stressed syllables from unstressed syllables in the same word, a closer look reveals that only F0 can realistically function as a perceptual cue.

Turning to duration first, we see that the average difference in duration between stressed and unstressed syllables in words such as *be.'be* is just 7ms.

7) *Mean durations (SD) for vowels in stressed and unstress syllables, in words with final stress (be.'be), n=40*

Stressed	Unstressed	Difference
78 ms (11)	71 ms (13)	7 ms (14) ($t = -3.188$, $p<0.05$)

It is worth asking what an average difference of 7 ms could possibly mean for the listener. Within the range of speech sound durations (30 to 300 ms), the just-noticeable differences in duration lie between 10 and 40 ms (Lehiste 1970: 13). A difference of 7 ms is thus rather unlikely to function as a perceptual cue. Of course, the comparison of mean durations obscures the fact that vowels from individual words in the data set can and do exhibit duration differences greater than 10 ms. Thus duration may function idiosyncratically, but not systematically, as a cue to stress.

Turning next to intensity, we see that the average difference in intensity between stressed and unstressed syllables is 0.7 dB.

8) *Mean intensities (SD) for vowels in stressed and unstressed syllables, in words with final stress (be.'be), n=40*

Stressed	Unstressed	Difference
68.5 dB (4.8)	66.5 dB (4.6)	0.7 dB (2.1) ($t = 2.161$, $p<0.05$)

Again, it is worth asking what this difference in intensity could mean for the listener. The just-noticeable difference in loudness is about 1 dB (Lehiste 1970: 119), so a difference of 0.7 dB is again unlikely to function as a perceptual cue.

In my data set, a majority of words with final stress exhibit greater intensity on the unstressed syllable. This pattern, which is in line with Konrot's (1981) results but not with Levi's (2005), may be perhaps be the result of positional strengthening in initial syllables, since the unstressed syllable is also the first syllable in the word. As with duration, there are of course some words in which the intensity difference between stressed and unstressed syllables is greater than 0.7 dB, so intensity may act as an idiosyncratic perceptual cue in those words.

The second formant, F2, also distinguishes between stressed and unstressed syllables in final position although the first formant, F1, does not. This suggests that some amount of vowel reduction may occur in unstressed syllables. However, it is impossible to speculate on the role of F2 for listeners without conducting perceptual experiments that are beyond the scope of this paper. I leave this question for future work.

Turning finally to F0, we see that the average difference between stressed and unstressed syllables is either 23.7 Hz or 25.3 Hz, depending upon the F0 measurement used.

9) *Mean F0 (SD) for vowels in stressed and unstress syllables, in words with final stress (be.'be), n=40*

	Stressed	Unstressed	Difference
Method 1: Averages $(F0_{start} + F0_{end})/2$	212.6 Hz (56.2)	189.0 Hz (65.8)	23.7 Hz (15.2) ($t=-9.861$, $p<0.05$)
Method 2: Peaks Maximum in last ¾ of vowel	216.7 Hz (56.9)	191.4 Hz (68.4)	25.3 Hz ($t=-10.029$, $p<0.05$)

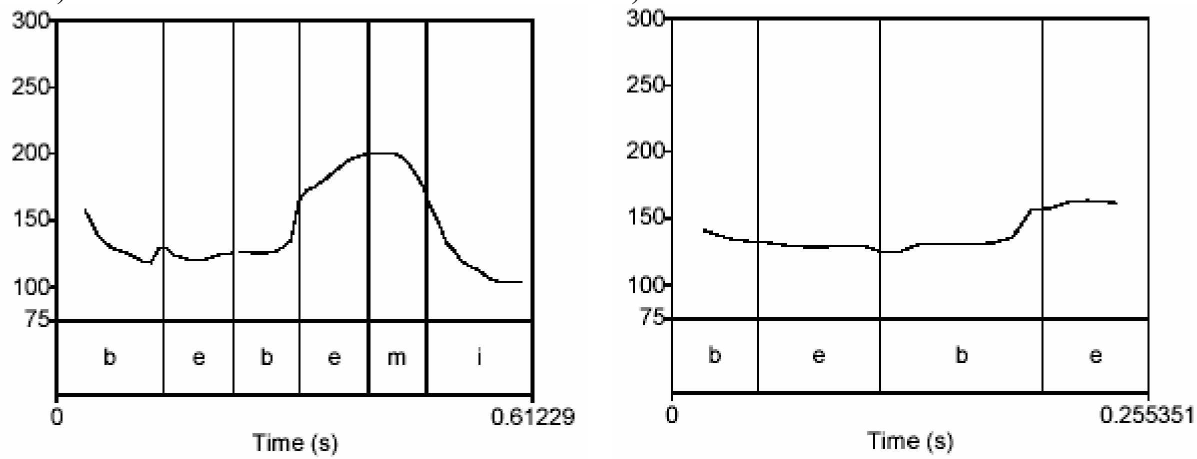
For the listener, it seems likely that such differences should be perceptible. Within fundamental frequencies in the range 80 to 160 Hz, the just-noticeable difference is about 1 Hz. Above these frequencies, the difference limen becomes progressively greater (Flanagan 1957: 534, via Laver 1994: 451).

A glance at the table reveals that the F0 measurements in this study go well above 160 Hz; this is primarily due to the higher F0s used by the two female speakers. When we examine the three male speakers only, the average F0s are almost all below 160 Hz, placing them squarely in the range quoted by Flanagan (1957). For words with final stress produced by male speakers, the average difference between stressed and unstressed syllables is 31.4 Hz (Method 1) or 34.3 Hz (Method 2). This suggests quite clearly that F0 is a robust perceptual cue for final stress in Turkish.

3.2 F0 contours on syllables of different shapes

The importance of F0 as a perceptual cue in final position would seem to be at odds with the fact that its presence is clearly diminished there, relative to non-final position. This paradox, apparent from the statistical analyses reported in the previous section, is made even sharper by an examination of two representative F0 tracings.

10 a) Non-final: *be.'be.mi* b) Final: *be.'be*



The F0 tracing in the non-final example (a) shows a sharp rise throughout the vowel of the stressed syllable, followed by a sharp fall that begins mid-way through the onset of the final syllable *mi*. By contrast, the F0 tracing of the analagous final example (b) remains mostly flat throughout the vowel of the stressed syllable, and its absolute value is only slightly higher than that found in the initial, unstressed syllable.

This visual comparison suggests that non-final stressed position represents a canonical realization of F0, i.e. a high-low contour, while final position represents a “clipped” instantation of F0, i.e. a contour that gets cut short. Under what conditions, exactly, do we get clipping? The F0 tracing in (a) suggests two straightforward, but very different answers. We could say that the high-low contour that realizes canonical stress in Turkish requires a two-syllable window. This requirement is met in words like *be.'be.mi,* but not in words like *be.'be.*

Alternatively, we could say that the high-low contour requires a minimum duration. Such a duration-based analysis makes very different predictions than a syllable based one. The minimim duration requirement, for example, could conceivably be satisifed in a number of different ways: by two syllables, as in (), or by a single syllable that happens to be long enough.

A similar question has been implicitly raised by Gordon’s (1999) analysis of contour tones. Using typological evidence drawn from a large survey of languages with contour tones, Gordon argues for an implicational hierarchy of tone-bearing syllables.

CVV < CVR < CVO < CV

CVV syllables are most likely to host contour tones, CVR syllables (closed by a sonorant) are next, CVO syllables (closed by an obstruent) are after that, and CV syllables are least likely. Furthermore, there is an implicational relationship among the syllable types: if a language has contour tones on CV syllables, then it also has contour tones on CVO, CVR, and CVV syllables. And so on.

Gordon (1999) offers a phonetic explanation for these typological facts, based upon the duration of the syllable’s sonorant portion. This is the portion of the rime that can actually convey F0 information: for CVV and CVR syllables, this includes the entire rime while for CVO syllables this includes the vowel only. It thus makes intuitive sense to think that CVV syllables should be typologically most likely to host contour tones, because the duration of their sonorant portion should be the longest. As Gordon shows, however, exceptions to this implicational hierarchy can and do occur, as in Cantonese. He accounts for these exceptions by claiming that raw duration sometimes subverts the implicational hierarchy.

The crucial question posed by Turkish, then, is similar to that confronted by Gordon: should we analyze the data in terms of syllable *types,* or in terms of syllable *durations?* To address this question, I compiled a list of words containing stressed CVV, CVR, CVO, and CV syllables in final position, using the Turkish Electronic Living Lexicon (Inkelas, Küntay, Orgun, & Sprouse 2000). The words were all disyllabic, and the vowel nuclei were all /a/. Because there are only so many real disyllabic words with identical /a/ vowels in both syllables, the onset consonant in each target syllable varied. Sample words are shown below.

11)

CV.'CVV	ha.'taa	‘mistake, error’
	ka.'zaa	‘accident’
CV.'CVR	da.'var	‘sheep or goat’
	ba.'kan	‘minister’
	tʃa.'tal	‘fork’
CV.'CVO	ma.'kas	‘scissors’
	na.'maz	‘prayer’
	ya.'tak	‘bed’
CV.'CV	ba.'ba	‘father’
	fa.'va	‘mashed beans’

Each word (n = 217) was placed in focused position within carrier sentences, and produced at a natural rate by a female speaker of Istanbul Turkish (n=1).

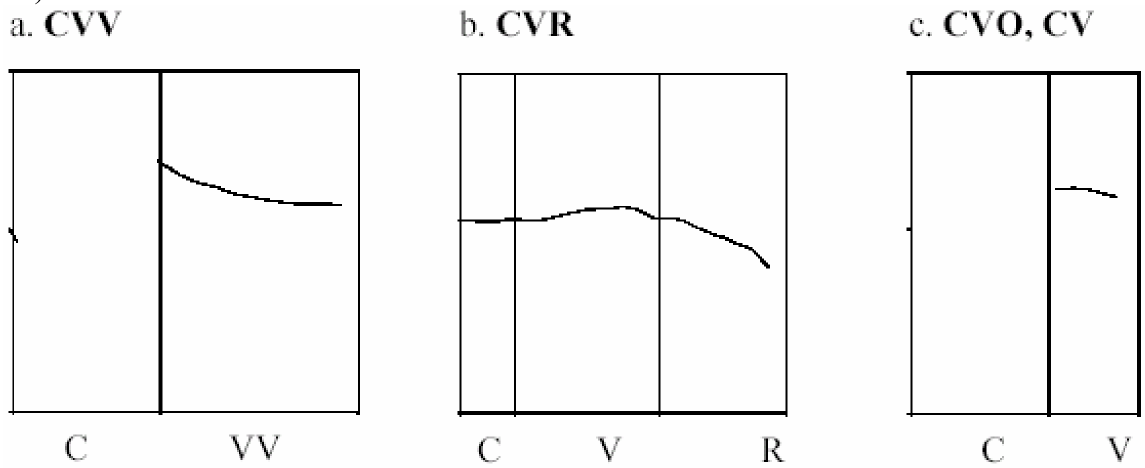
3.2.1 Measurements

For each word, I conducted a visual inspection of the realization of F0 on the stressed final syllable, in order to see how much the F0 contour resembled the high-low movement that is so typical of non-final stress. Such a visual inspection has obvious limitations, and would ideally be quantified by slope measurements, but nevertheless offers an efficient and reasonable means of analyzing contours. For each word, I also measured the duration of the sonorant portion of the stressed syllable’s rime. Thus for CVV and CVR syllables, I measured the entire rime. For CVO and CV syllables, I measured the vowel only.

3.2.3 Results

Syllables closed by a sonorant consonant (CVR) consistently realized a high-low contour. Surprisingly, however, open syllables with long vowel nuclei (CVV) did not. Representative examples are shown in (), which also shows that syllables with an obstruent coda (CVO) or with no coda (CV) fail to realize a contour.

12)



These results clearly cannot be analyzed in terms of syllable types. Any weight-based theory, and Gordon's (1999) theory in particular, would claim that CVV syllables must be heavier than CVR syllables and thus more capable of hosting an F0 contour. But the Turkish data run counter to this generalization.

The duration measurements, however, offer a straightforward explanation. Contrary to what we might expect, the mean duration of the sonorant portion of the rime is consistently *greater* in CVR syllables than in CVV syllables.

13) *Mean duration of sonorant portion of rime, for different syllable types*

CVV (n=7)	CVR (n=98)	CVO, CV (n=92)
191 ms	238 ms	117 ms

These findings suggest a duration-based analysis for the realization of Turkish stress: the longer the sonorant portion of a final syllable, the more of the high-low contour we see. It does not matter what "type" the syllable is, or how heavy we might expect it to be on the basis of typological criteria. What matters is its duration.

The duration-based analysis also offers a straightforward solution to the two-syllable realization of F0 contours that we noted above, in reference to the F0 tracing in (a). If stress realization makes fundamental reference to duration, then F0 contours would be expected to spread over two syllables whenever circumstances permit. A syllable-type analysis, by contrast, has no principled way to unify the behavior of single- and multiple-syllable windows for F0 contours.[1]

These findings must be taken as preliminary, due to the small subject pool and the small number of CVV words, the latter of which results directly from constraints on the lexicon of Turkish (there are few words with stressed CVV syllables). The duration measurements reported here, however, are in agreement with those reported elsewhere. Kopkalli-Yavuz (2000) found that Turkish vowels are longer in closed syllables than in open ones, and showed that the

[1] A mora-based argument, of course, could unify single CVV syllables and multi-syllable CV.CV sequences, by claiming that two moras are required for the realization of the canonical F0 contour. But the fact that CVV syllables are not, in general, capable of hosting these contours in Turkish throws a wrench into this idea.

duration differences in all three vowel types are statistically significant. Barnes (2002) also found that vowel durations in Turkish are longer in closed syllables: in particular, vowel duration is longer in syllables closed by voiceless stops, longer still in syllables closed by nasals, and longest in syllables closed by /r/. Despite this, "overall rhyme duration in closed syllables remains remarkably constant" (2002: 5). In light of the arguments presented here, Barnes's results are particularly interesting because they suggest that the longer vowel duration in CVO syllables could make them more capable of hosting F0 contours than CV syllables, although my data shows that this is not the case. CVO and CV syllables behave similarly in their failure to host a contour, because they both fall well below the "sweet spot" -- or the minimum duration needed for canonical realization of stress.

4. Conclusion

I have shown that stress is indeed realized differently in final and non-final positions in Turkish, that F0 is a primary correlate even in final position, and that final syllable types differ markedly in their ability to host the canonical high-low F0 contour. I have argued that an appeal to phonological syllable types cannot account for the facts, but that a reference to phonetic duration can.

This research raises some interesting questions that warrant further investigation. The first has to do with the status of F0 in final position: if it is realized only on syllables of a certain duration (e.g., CVR syllables), how is it that listeners still "perceive" stress on those syllables where it is not realized (CVV, CVO, CV)? Recall from §3.1.3 that intensity, duration, and F2 also differentiated stressed from unstressed syllables. I argued that intensity and duration were nevertheless unlikely to play a real role in perception, but it may be that the differences in these two correlates are greatest *just* in those cases where overall phonetic duration prevents realization of canonical F0 contours. In other words, we might predict that short stressed syllables (CVV, CVO, CV) become longer and louder, while long stressed syllables (CVR) do not.

The second question has to do with the nature of phonological and phonetic categories. Turkish clearly subverts a typological hierarchy whereby CVV syllables should be heavier than CVR syllables, but this may not necessarily mean that the language makes direct reference to phonetic duration. It could simply mean that the hierarchy is wrong, or wrong for Turkish. To pursue this question and maintain a strictly phonological analysis of the question, it would be worthwhile to further investigate the grammar of Turkish for phonological processes that make abstract reference to CVR syllables specifically.

Acknowledgments

I gratefully acknowledge the contributions of Özlem Çetinoğlu, Orhan Bilgin, Sharon Inkelas, Susannah Levi, Ian Maddieson, Hasan Mesut Meral, Kemal Oflazer, and audience members at the annual meeting of the Linguistic Society of America, 2004. Flaws are mine.

References

Gordon, Matthew. 1999. The process-specific nature of weight: The case of contour tone restrictions. Kimary Sharon, Susan Blake, and Eun-Sook Kim (eds.) *WCCFL 17.* Stanford, CA: CSLI.

Konrot, Ahmet. 1981. Physical correlates of linguistic stress in Turkish. *University of Sussex Language Centre Occasional Papers 24:* 26-52.

Konrot, Ahmet. 1987. Hendrik E. Boeschoten and Ludo Th. Verhoeven (eds.) *Studies on modern Turkish: Proceedings of the third conference on Turkish linguistics.* Tilburg University Press.
Laver, John. 1994. *Principles of Phonetics.* Cambridge University Press.
Lehiste, Ilse. 1970. *Suprasegmentals.* Cambridge, MA: MIT Press.
Levi, Susannah V. 2005. Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association* 35(1):73 -97.