PHONETIC REALIZATION OF FOCUS WITH NO ON-FOCUS PITCH RANGE EXPANSION IN TURKISH

Canan Ipek

University of Southern California, USA
ipek@usc.edu

ABSTRACT

This study investigates what acoustic parameters are relevant in terms of production and perception of focus in Turkish. Results from acoustic analyses shows that speakers do not expand on-focus pitch range but duration and intensity changed as a function of focus. Post-focus pitch is lowered in initial focus, but medial focus did not differ from neutral focus in any of the acoustic parameters. Listeners identified initial focus correctly with the highest rate and this shows the importance of post focus compression (PFC) in correct identification of focus.

Keywords: Turkish, focus, post-focus compression, focus perception

1. INTRODUCTION

A recurrent finding in studies on the phonetic realization of focus is that F_0 variations are the main correlate of focus while certain amount of adjustments in duration and amplitude is also involved [1, 5, 10, 14]. A focused item has higher pitch in comparison to its nonfocused counterpart. However, research on the temporal domain of focus has shown that the effect of focus on F_0 variations is not just limited to the on-focus word. In particular, a pitch range expansion on the focused word is usually followed by pitch range compression in the post-focus domain, if any [3, 8, 12, 14].

Encoding focus by F_0 variations both on the on-focus and post-focus domain seems to be a relevant cue for correct focus identification. Chen, et al. [2] compared focus perception rates of speakers from languages with and without PFC. They found that correct identification of focus for a language with PFC was over 90% whereas the percentage was below 75% for those that lack PFC. The study by Rump and Collier [10] investigated the relative height of two F_0 peaks in an utterance as a function of focus structure, and listeners' sensitivity to those F_0 variations. They found that listeners identified an early single focus only when the second F_0 peak was virtually absent.

This paper presents data from a production and a perception experiment on Turkish to further our understanding of the relevant acoustic components for encoding and decoding focus. No similar acoustic and perception analyses have been done on Turkish before. The aim is to investigate the phonetic realization of focus in Turkish on one hand and to analyze listeners' judgment of accent location, on the other in a systematic way.

2. METHOD

2.1. Production experiment

2.1.1. Stimuli

Four target sentences with an SOV order, each containing 3 words were used. There were four focus conditions for each target sentence. Those conditions were neutral, initial, medial, and final. To elicit different types of focus on a specific word, each target sentence was preceded by a prompt question.

Table 1: Target Sentences and their translations.

	Target Sentences	English translation
1	Tuna babamı dövmüş	'Tuna beat my dad'
2	Lale duvarı boyamış	'Lale painted the wall'
3	Döne dedemi kovmuş	'Döne sent away mygrandpa'
4	Mine burnunu yıkamış	'Mine washed her nose'

Target sentences and their prompt questions were grouped into four blocks and randomized. Each sentence was repeated five times. Thus there were 4 sentences × 4 foci × 5 repetitions × 6 speakers = 480 utterances.

2.1.2. Subjects

Six native speakers of Turkish, three males and three females, participated in the study. The mean age of the subjects was 25.2 years. All of the subjects were recruited from the University of Southern California and were paid for their participation. None of the subjects reported any speech or hearing problems.

2.1.3. Recording procedure

The recording sessions took place in a quiet room. The subjects were seated in front of a computer screen. A unidirectional, usb microphone was placed to the left of the computer, approximately 6 inches from the speaker's lips. The target sentences were given to the speaker before the recording session for the readings to be as natural as possible. During the recording, each prompt question was read aloud by the experimenter and the subjects read the target sentence aloud as an answer to the question.

2.1.4. F_0 extraction

F_0 extraction from the target sentences was done using a general-purpose Praat script [13]. The syllable boundaries were marked by hand. The script then removed local spikes and sharp edges from the raw F_0 contours, and computed mean F_0, intensity and duration of each syllable.

2.1.5. Analyses and results

The comparisons are made between the focused word and its neutral counterpart for F_0, duration, and intensity at the on-focus, pre-focus and post-focus regions. Figure 1 displays time-normalized F_0 curves averaged across all tokens and subjects.

From Figure 1 we can see that focus does not cause on-focus F_0 increase for initial and medial focus conditions. On the other hand, final word has higher F_0 than its neutral counterpart when focused, however there is pitch range expansion on the syllable that immediately precedes the final word, so the interpretation of the on-focus F_0 difference for final word is not straightforward. Similar pre-focus pitch range expansion is also observed on the initial word when the medial word is focused. F_0 drop is clearly visible for initial focus condition but not for medial focus one.

For statistic tests, pre-focus domain for initial word is divided into two separate groups, namely pre-focus for medial focus condition and pre-focus for final focus condition. This is done particularly to be able to see whether the pre-focus pitch expansion relative to its neutral counterpart which is visible in Fig. 1 is significant. The same categorization is done for the final focus condition, too (i.e. two separate groups for post-focus condition). Thus, each word can be in four different conditions. For the initial word, these conditions are a) neutral, b) on-focus, c) pre-focus (medial), i.e., when medial word is on-focus, d) pre-focus (final); for the medial word a) neutral, b) on-focus, c) pre-focus, d) post-focus; and for the final word a) neutral, b) on-focus, c) post-focus (medial), d) post-focus (initial). Figure 2 displays the comparision of mean F_0 values with standard errors for each stressed syllable at different focus conditions a word can have.

In order to determine which acoustic features systematically vary as a function of focus, one-way repeated measures ANOVAs were conducted comparing mean F_0, intensity and duration for the same stressed syllable in four different focus conditions.

Figure 1: Time normalized F_0 contours averaged across 6 speakers, 4 sentences and 5 repetitions. The vertical lines indicate syllable boundaries. The thicker vertical lines indicate word boundary in an SOV order.

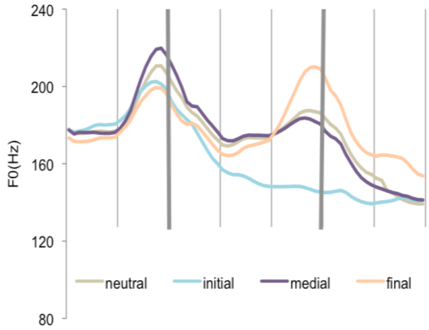
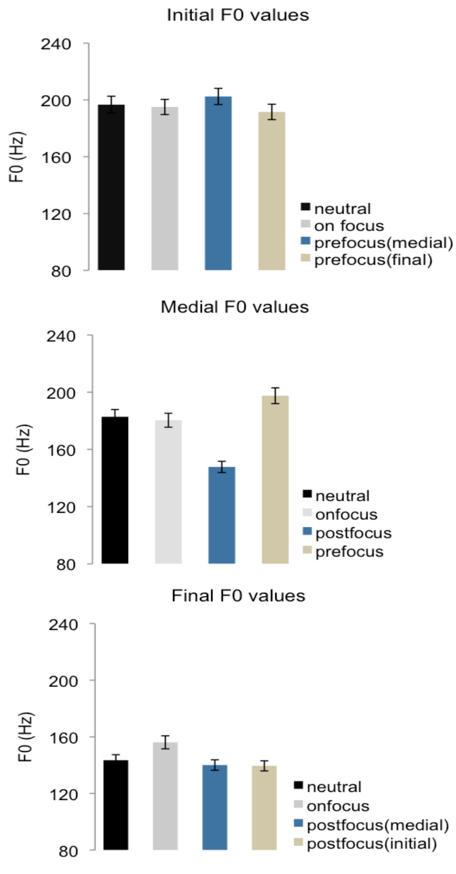


Figure 2: Means and standard errors of F_0 values on the stressed syllable for initial, medial, final and their corresponding neutral focus conditions.



The results displayed in Table 2 show significant differences for F_0, intensity and duration for all focus conditions, except for duration for final focus. Post-hoc Bonferroni tests were carried out to identify which pairs differed significantly. Some significant differences obtained from the ANOVA test did not show significance in the post-hoc comparisons. This might be due to the relatively conservative post hoc test.

Table 2: Results of 1-way repeated measures ANOVA (df = 3, 15).

	Mean F_0	**Mean intensity**	**Mean duration**
Initial	F = 4.1194 *p* = 0.025	F = 6.0614 *p* = 0.006	F = 4.6657 *p* = 0.017
Medial	F = 21.221 *p* < .001	F = 30.742 *p* < .001	F = 4.1176 *p* = 0.025
Final	F = 5.3586 *p* = 0.01	F = 23.735 *p* < .001	F= 3.2205 *p* = 0.052

According to post-hoc tests comparing neutral and on-focus conditions, medial-focus words did not differ significantly from their neutral focus counterparts in any of the acoustic parameters. Initial words had significantly longer duration, and on-focus final words had greater magnitude for intensity. In order to determine whether there is significant post-focus lowering of F_0 and intensity, post-focus conditions are compared with neutral condition. The results indicate significant lowering for both F_0 and intensity for initial focus but not for medial focus. As for the pitch range expansions in the immediately pre-focus domain, post-hoc comparisons between the pre-focus and neutral condition show significant difference for final focus only.

2.2. Perception experiment

2.2.1. Stimuli

The stimuli used in the perception experiment came from the sentences recorded in Experiment 1. Two speakers, one male and one female from experiment 1, were selected. They were chosen because they had the minimum mean standard deviations of all F_0 point across four focus conditions. The third and fourth repetitions of each sentence for each focus type were used. Thus, there were 2 speakers × 2 repetitions × 4 sentences × 4 focus = 64 tokens.

2.2.2. Subjects

Ten native speakers of Turkish, five males and five females, ranging in age from 24-40 (mean = 28.1 years), participated in the study. Seven of them were recruited from the University of Southern California, and the rest were recruited from Turkish native speakers living in Southern California for less than five years. None of the subjects reported any speech or hearing problems.

2.2.3. Listening procedure

Subjects were asked to listen to the target sentences using headphones and choose the word emphasized, or none if they could not hear any emphasis. Subjects were given 10 practice trials before the real target items, but were not given any feedback on the correctness of their answers. The audio stimulus was played once, and there was no time limitation for the subject to make a judgment.

2.2.4. Results

The confusion matrix for focus perception is given in Table 3. One way repeated measures ANOVA showed significant effect of focus type (F[3, 27]=7.0474, p<.001). Recognition rates indicate that speakers identify initial focus correctly with the highest percentage. Neutral focus is the one recognized least correctly and it is mostly confused with medial focus. For final focus, subjects did slightly better than medial focus and it's mostly confused with neutral focus.

Table 3: Confusion matrix of focus perception (%). Correct identification is marked by bold face.

heard as / original	Neutral	Initial	Medial	Final
Neutral	**45**	11.25	31.87	11.87
Initial	16.25	**75.63**	5	3.12
Medial	21.87	18.75	**56.25**	3.12
Final	26.87	5.63	7.5	**60**

3. DISCUSSION AND CONCLUSION

The acoustic analyses in Experiment 1 showed that unlike many other languages, there is no on-focus pitch range expansion in Turkish. Rather, a focused word acoustically differs from its neutral counterpart in terms of greater magnitude in duration and intensity. However, this difference is not independent of position of focus. While the final word differed from its neutral counterpart in terms of intensity, focus on the initial word caused increase in duration, and medial word did not differ from its neutral counterpart in any of those acoustic parameters. The lack of any acoustic difference in the medial position might be argued to be due to that position being attested as the "syntactic" focus position in Turkish [4, 7]. Such "default" focus could be due to the fact that Turkish is SOV. Verbs are known to have the tendency to have lower F_0 than nouns [6, 11]. As shown in both Figs. 1 and 2, F_0 in the verb position is always low in the present data. As a result, the OV pair seems to always exhibit a high-low pitch pattern, which resembles a focus + post-focus construct. This may lead the object to be always heard as focused or at least more prominent. This seems to have rendered the pre-verbal position in Turkish somewhat like the sentence-final position in non-SOV languages, where a narrow focus is not very distinct from broad/neutral focus [1, 9]. If our interpretation is valid, it could be applicable to other verb-final languages as well.

Significant F_0 drop in the post-focus domain has been found for initial focus, but not for medial focus. An unexpected finding of the study is the immediate pre-focus pitch range expansion when the final word is on-focus. This pre-focus raising might be due to the fact that the verb in Turkish always has to have lower pitch than the object. So the raising of pre-verbal pitch might be due to speakers' attempt to guarantee this when the pitch of the verb is raised due to focus. It might as well result from a mechanism that tries to maximize the difference between the maximum F_0 of the object and the verb, and the magnitude of this difference might be relevant in terms of perception.

The results of the perception experiment showed that initial focus, encoded with PFC, has the highest recognition rate. The recognition rate for final focus was lower than that of initial focus which is in line with findings from other studies where final focus had the lowest recognition rate compared to medial and initial focus [1]. The relatively poor performance for medial focus perception might be attributed to the fact that acoustically medial focus and neutral focus did not differ from each other significantly.

The present study concentrated on the acoustic correlates of prosodic focus in Turkish. The results from the production experiment showed that narrow focus did not cause significant changes in F_0 on the word in focus, however increases in duration and intensity were observed for initial and final focus word, respectively. Focus, on the other hand, caused significant F_0 increase in the pre-focus domain when the final word was focused. The effect of focus in the post-focus domain was observed as a significant fall in F_0 values when the initial word was on-focus. This PFC in the case of initial focus seemed to have facilitatory effect with respect to correct focus identification such that results of the perception experiment showed that listeners identified focus correctly at the highest rate when the initial word was on focus. Whether pre-focus F_0 increase observed for the final focus condition has any effect on the final word on-focus F_0 and/or its correct identification as prominent requires further investigation.

4. REFERENCES

[1] Botinis, A., Fourakis, M., Gawronska, B. 1999. Focus identification in English, Greek and Swedish. *Proc. 14th International Congress of Phonetic Sciences* San Francisco, 1557-1560.

[2] Chen, S.-W., Wang, B., Xu, Y. 2009. Closely related languages, different ways of realizing focus. *Interspeech* Brighton, UK, 1007-1010.

[3] Cooper, W.E., Eady, S.J., Mueller, P.R. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of Acoustical Society of America* 77, 2142-2156.

[4] Erguvanlı, E. 1979. *The Function of Word Order in Turkish Grammar*. Ph.D. thesis, University of California.

[5] Gu, W., Lee, T. 2007. Effects of tonal context and focus on Cantonese f0. *16th ICPhS* Saarbrücken.

[6] Hoskins, S. 1996. A phonetic study of focus in intransitive verb sentences. *Proc.* 4th *ICSLP* Philadelphia, 1632-1635.

[7] İşsever, S. 2003. Information structure in Turkish: the word order-prosody interface. *Lingua* 113, 1025-1053.

[8] Lee, Y., Yi, X. 2010. Phonetic realization of focus in Korean, *Proc of Speech Prosody* Chicago.

[9] Liu, F., Xu, Y. 2005. Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 62, 70-87.

[10] Rump, H.H., Collier, R. 1996. Focus conditions and the prominence of pitch-accented syllables. *Language and Speech* 39, 1-17.

[11] Shih, C. 2000. A declination model of Mandarin Chinese. In Botinis, A. (ed.), *Intonation: Analysis, Modelling and Technology*. Dordrecht: Kluwer Academic Publishers**,** 243-268.

[12] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of F_0 contours. *Journal of Phonetics* 27, 55-105.

[13] Xu, Y. 2005-2011. ProsodyPro.praat. Online: *http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/*

[14] Xu, Y., Xu, C.X. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33, 159-197.