THE INFLUENCE OF SILENCE ON PERCEIVING THE PRECEDING TONAL CONTOUR

David House
Dept. of Linguistics and Phonetics, Lund University, Helgonabacken 12, S-223 62, Lund, Sweden.

ABSTRACT

Interactive adjustment tests were carried out to test if a silent interval influences the perception of the preceding tonal contour. Results from 16 subjects show a strong influence of silence on tonal perception indicating that silence increases sensitivity for the preceding tonal endpoint with subjects showing greatest response consistency for the stimuli with the longest pause where adjustment is based on endpoint frequency before the pause.

INTRODUCTION

In both read and spontaneous speech, a prosodic phrase boundary is often accompanied by a silent pause which is preceded by a tonal contour marking the boundary. Considerable attention has been directed to the respective roles of silent pauses and boundary tones as markers of prosodic phrase and syntactic boundaries, see e.g. [1], [2], [3], [4], [5], [6], [7], [8] and [9]. The central question approached by this investigation is whether a silent interval influences the perception of the preceding tonal contour.

A number of general questions concerning boundary tones relate to the central issue of this investigation. Boundary tones may have several functions in addition to boundary signalling, e.g. signalling feedback seeking or turn regulation in spontaneous dialogue [10]. Are the tones and functions perceived categorically and, if so, does a silent interval facilitate perception?

In a previous study [11] it was shown that in synthesized VCVCV sequences where V= [a] and C= [m], the tonal configuration in vowels is perceptually more salient than in consonants. It can be conjectured, however, that if a silent interval is inserted before a vowel, tonal perception in the preceding consonant may be sharpened. This could give the final tonal level in the consonant greater perceptual significance than when immediately followed by a vowel. Thus, the following specific questions are addressed by this investigation: 1. Does a silent interval influence tonal perception? 2. Are final sonorant consonants important tone carriers? 3. Is perception of the tonal endpoint before a pause sharpened by the pause, and if so, does this sharpening increase with increased pause duration?

METHOD

Stimuli and task

To answer these questions, a set of adjustment tests was designed. Stimuli consisted of synthesized [amamam] sequences in three temporal conditions: 1) no pause between segments, 2) a 100 msec pause between the fourth and fifth segment [amam.am] and 3) a 1000 msec pause between the fourth and fifth segment [amam..........am]. Formant synthesis was used to generate the stimuli [12].

The subjects' task was to match different tonal configurations within each temporal condition. Matching was done interactively using a mouse pointer on a computer screen (Sun workstation, ESPS-Waves+ environment). The tonal configurations were 1) a falling F0 contour where the fall occurred through both the second vowel and second consonant and 2) a falling F0 contour through the second vowel only with a constant F0 on the second consonant. 10 Hz steps between 140 and 60 Hz were used to create 9 different stimuli in each tonal configuration making a total of 18 different stimuli for each temporal condition, i.e. a total of 54 stimuli for all three temporal conditions and both tonal configurations. See Figure 1 for stylized samples of stimuli.

Where endpoint frequency is most salient, subjects would be expected to match tonal configurations having the same endpoint frequency regardless of whether the contour falls in the vowel only or in both the vowel and consonant.

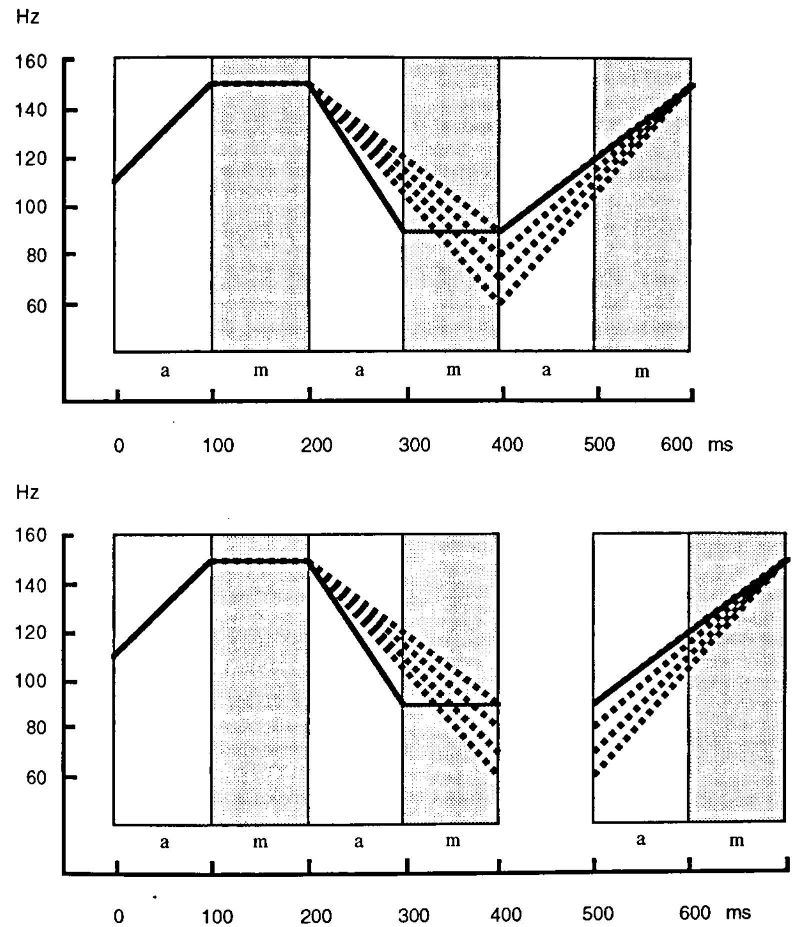

Figure 1. Stylized contours of some example stimuli. The upper panel represents the temporal condition "no pause" while the lower panel represents "short pause". The dotted lines represent tonal contours falling in both vowel and consonant (VC-fall) while the solid lines represent tonal contours falling in the vowel only (V-fall).

If endpoint frequency is of less perceptual importance, subjects would be expected to match contour shapes. This would result in subjects matching a lower endpoint for a vowel-consonant fall with a higher endpoint for a vowel only fall.

Test configuration

The middle five stimuli in the continuum of nine stimuli in each tonal configuration were presented as original stimuli. This resulted in six blocks of five stimuli each for a total of 30 presented stimuli. Stimuli were randomized in each block and block order was randomized between listeners. Subjects were asked to match each original stimulus to one of the nine stimuli having the same temporal conditions but the different tonal configuration. The test was presented as an adjustment procedure as in [13] with the nine choices presented in frequency order on the computer screen. All screen input was logged to a file.

Each subject began the test with a practice/calibration block in which the original stimulus was identical to one of the nine choices. The entire test took an average of 33 minutes with a minimum individual time of 13 minutes and a maximum of 55 minutes.

Subjects

16 subjects participated in the experiment. Subjects were mostly students and staff at the Dept. of Linguistics and Phonetics, Lund University, and all but two were native speakers of Swedish. Subjects were not paid, but were rewarded with chocolate and coffee after their participation in the test.

RESULTS

Results were very consistent between subjects: one factor ANOVA df=15, F=0.69, p>0.05, and within subjects df=2, F=43.17, p<0.0001. Figure 2 shows the percentage of same endpoint responses for the three temporal conditions and for the two tonal configurations within each condition.

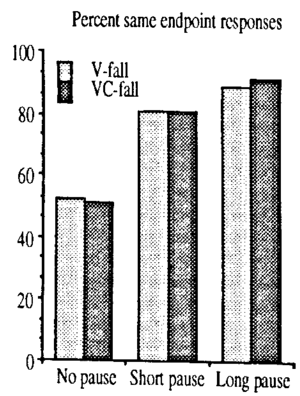

Figure 2. Graph showing percentage same endpoint responses as original stimulus when original is V-fall (falling contour in vowel only) and VC-fall (falling contour in vowel and consonant).

A strong effect of pause on endpoint perceptual salience can be seen. In the no-pause condition, only half the responses were same endpoint, while the other half were in the direction of a lower endpoint for vowel-consonant fall (VC-fall) being matched with a higher endpoint for the vowel fall (V-fall). Table 1 shows the response distribution where the direction is from the vowel fall. Endpoint salience also seems to increase somewhat with pause duration.

Table 1. Endpoint response distribution for the three temporal conditions. Direction is frequency of endpoint related to endpoint of tonal contour falling in vowel only.

	Lower	Same	Higher
No pause	76	83	1
Short pause	28	130	2
Long pause	12	144	4

A chi square test of independence on the above distribution results in $\chi 2=76.54$, df=4, p<0.001. One way ANOVA shows a significant difference comparing no pause with short pause F(2,45)=21.13, p<0.001 and comparing no pause with a long pause F(2,45)=40.5, p<0.001, but not when comparing a short pause with a long pause F(2,45)=3.13, p>0.05.

DISCUSSION

The results demonstrate a strong effect of silence on the perception of the tonal contour. They also demonstrate the importance of a sonorant consonant as a tone carrier particularly when followed by silence.

In the pause stimuli, matching seems to be based primarily on endpoint frequency before the pause, while in non-pause stimuli, listeners seem to be attending more to fall gradients or to average frequency through the fall. An interpretation concerning auditory memory may serve to help explain the results. If short-term auditory memory for frequency is sharpened by the presence of a pause, then subjects should find it more salient to match endpoint frequency even if the final segment is not a vowel. In the no-pause condition, auditory memory relies more on the tonal contour since endpoint frequency is rendered less salient by the following vowel.

This interpretation may be modified by the fact that, due to test construction constraints, there was also some

information after the pause which listeners could have used as well as the endpoint information before the pause. The fact remains, however, that the presence of the pause significantly influenced perception of the tonal contour.

This can have implications for perception of such tonal phenomena as boundary tones and discourse markers. The presence of a pause may therefore sharpen perception of a boundary tone or discourse marker. More precision may be called for in intonation modelling and automatic stylization of intonation, especially concerning tonal contours before pauses. This would be in line with data in [14] where automatic stylization for recognition tended to fail most often in prepausal positions.

ACKNOWLEDGMENT

Many thanks are due to Marcus Filipsson for writing the interactive program used for the perception test.

REFERENCES

[1] Beckman, M., and Pierrehumbert, J. (1986), "Intonation structure in Japanese and English", in J. Ohala (ed.), *Phonology Yearbook 3*, pp. 255-309.

[2] Gårding, E., and House, D. (1987), "Production and Perception of Phrases in some Nordic Dialects", In P. Lilius and M. Saari (eds.). *The Nordic Languages and Modern Linguistics 6*, pp. 163-175, Helsinki University Press.

[3] Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P.J. (1992), "Segmental durations in the vicinity of prosodic phrase boundaries", *Journal of the Acoustical Society of America*, vol. 91, pp. 1707-1717.

[4] Bruce, G., Granström, B., Gustafson, K. and House, D. (1993), "Phrasing strategies in prosodic parsing and speech synthesis", *Proceedings Eurospeech '93*, pp. 1205-1208, Berlin, Germany.

[5] Bruce, G., Granström, B., Gustafson, K. and House, D. (1993), "Interaction of F0 and duration in the perception of prosodic phrasing in Swedish", In B. Granström and L. Nord (eds.). *Nordic Prosody VI*, pp. 7-21. Stockholm: Almqvist & Wiksell International.

[6] Pijper, J.R. de, and Sanderman, A. (1993), "Prosodic cues to the perception of constituent boundaries", *Proceedings Eurospeech '93*, pp. 1211-1214, Berlin, Germany.

[7] Strangert, E. (1993), "Speaking style and pausing", *Reports from the Department of Phonetics, University of Umeå, PHONUM 2*, pp. 121-137.

[8] Strangert, E. and Strangert, B. (1993), "Prosody in the perception of syntactic boundaries", *Proceedings Eurospeech '93*, pp. 1209-1210, Berlin, Germany.

[9] Swerts, M. and Geluykens, R. (1994), "Prosody as a marker of information flow in spoken discourse", *Language and Speech*, vol. 37, pp. 21-43.

[10] Bruce, G., Granström, B., Gustafson, K., House, D. and Touati, P. (1994), "Modelling Swedish prosody in a dialogue framework", *Proceedings of the 1994 International Conference on Spoken Language Processing*, pp. 1099-1102, Yokohama.

[11] House, D. (1990), *Tonal Perception in Speech*, Lund: Lund University Press.

[12] Carlson, R., Granström, B. and Hunnicutt, S. (1991), "Multilingual text-to-speech development and applications", in W. Ainsworth (ed.), *Advances in speech, hearing and language processing*, pp. 269-296, London: JAI Press.

[13] d'Alessandro, C. and Castellengo, M. (1993), "The pitch of short-duration vibrato tones", *Journal of the Acoustical Society of America*, vol. 93, pp. 1617-1630.

[14] House, D. and Bruce, G. (1990), "Word and focal accents in Swedish from a recognition perspective", In K. Wiik and I. Raimo (eds.), *Nordic Prosody V*. pp. 156-173. Turku University.