

PERCEPTION OF TONAL RISES AND FALLS FOR ACCENTUATION AND PHRASING IN SWEDISH

David House^{*†}, Dik Hermes^{**} and Frédéric Beaugendre⁺

^{*}Department of Languages, University of Skövde, Sweden

^{**}IPO, Center for Research on User-System Interaction, Eindhoven, The Netherlands

⁺Lernout and Hauspie Speech Products, Wemmel, Belgium

[†]Current affiliation: Department of Speech, Music and Hearing, KTH, Stockholm, Sweden

ABSTRACT

In previous experiments with Dutch, French and Swedish listeners, it was shown that the location in the syllable of the onset of a rising or falling pitch movement is critical for the perception of accentuation. As the onset of the pitch movement is shifted through the syllable, there is a point at which the percept of accentuation shifts from one syllable to the next. This point is termed the accentuation boundary. It has also been proposed that in certain positions, the percept of accentuation conflicts with the percept of phrasing. An experiment with Swedish listeners was carried out using the same stimuli as used for the accentuation study, but now the task was to determine the phrasing of the syllables. The results indicate that perceptual phrase boundaries can be determined in the same way as accentuation boundaries. Differences in the locations of the boundaries can be interpreted in terms of strengths of tonal cues for accentuation and phrasing.

1. INTRODUCTION

In recent work on the timing of pitch movements and perception of accentuation in Dutch, Hermes [8] established that the location of the onset of an accent-lending pitch movement determines which syllable is perceived as accented. By systematically shifting the timing of 80-ms, 120-ms and 160-ms accent-lending rises and falls in /mamamama/ and /a.a.a.a.a/ utterances from left to right, he determined the point for onsets of pitch movements where the percept of accentuation shifts from one syllable (S_n) to the next (S_{n+1}) and referred to it as the *accentuation boundary* (AB). For Dutch, the accentuation boundaries were located near the end of the /a/ vowel near the following syllable onset. In the same way, Hermes [8] could determine the high-low boundary of a stressed syllable. For the rise this is the location in the syllable before which a rise lends a high accent and after which it lends a low accent. For the fall this is the other way round. Similar results were obtained for two additional sets of /mamamama/ stimuli in which the /m/ duration was doubled and halved, and two additional sets of /a.a.a.a.a/ stimuli in which the duration of the intervocalic silent interval was doubled and halved.

These results led to the question of language specific accentuation boundaries. Are the boundaries universal, or do they differ in relationship to language dependent accentuation structure? To address this question, a set of similar experiments was carried out using French and Swedish (southern dialect) subjects [2]. The results for the French and Swedish listeners

were consistent with the finding for Dutch that the cue which induces the percept of accentuation is located at the onset of the pitch movement. However, the results for the French listeners differed in two ways from those of the Dutch and Swedish listeners. Firstly, the location of the accentuation boundaries for the accent-lending rises differed. For the Dutch and Swedish listeners, the boundary for the rises was located near the end of the vowel of S_n while for the French listeners, the boundary was located earlier, near the middle of the vowel of S_n . Secondly, for the accent-lending falls, the AB could be determined quite accurately for the Dutch and Swedish listeners. The result was very vague for the French subjects, however, indicating that they had difficulty perceiving the falls as accent-lending.

The accentuation boundaries for the three languages are plotted in Figure 1 for the /mamamama/ stimuli with the /m/ of normal duration and in Figure 2 for the /a.a.a.a.a/ stimuli with the intervocalic silent interval of normal duration. Similar results were obtained for stimuli with doubled and halved consonant and silence durations [2].

The language differences found in the above study were interpreted in terms of temporal-alignment categories of accent-lending rises and falls [12]. A model of perceived accentuation was proposed in which several different categories of accentuation are represented as the falling or rising movement is advanced through the test syllables. The perceived category would depend upon whether a pitch jump or the onset of the pitch movement (rise or fall) is perceived and whether or not the particular category is represented in the language in question. The percept of phrasing can also influence the perceived category. An attempt has been made to describe correspondences between these perceptually-based categories and phonological categories presented in theories of intonation for the three languages.

For rises, a jump category was proposed and formalized as L+H* in autosegmental terminology [3]. This jump category corresponds to rise 1 in the Dutch description of intonation as presented by 't Hart, Collier and Cohen [15] and to a focal accent realization of an accent I word in Swedish as presented by Bruce [4]. The perceived tonal change is a jump up in tonal levels from low to high into the stressed vowel. Also for rises, a rise-onset category was proposed and formalized as L*+H corresponding to rise 3 in Dutch and to a focal accent realization of an accent II word in Swedish, especially in dialects of southern Sweden [5]. Here the perceived tonal change is a rise on the stressed vowel with the onset of the rise

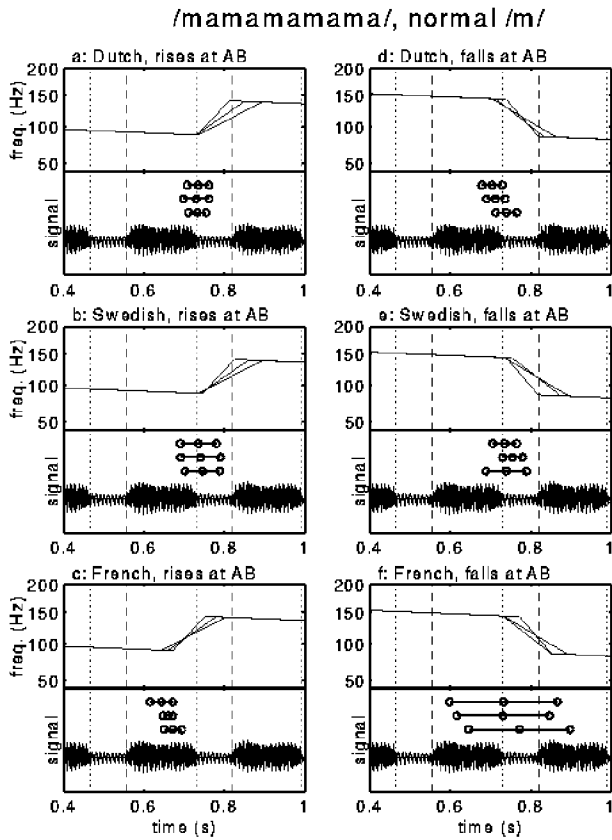


Figure 1: Accentuation boundaries (AB) obtained for /mamamama/ stimuli with the /m/ of normal duration for Dutch, Swedish and French subjects. Locations of the movements of three different durations (80, 120, and 160 ms) at the AB are plotted for the rises (left panels) and falls (right panels). The circles represent the quartiles of the error function fitted through the response distributions. These are plotted in triplets where the lower plot represents 80ms, the middle 120ms and the upper 160ms movement durations. The lower panel of each frame also presents the oscillogram monotonized at 100 Hz.

being the main cue for accentuation and occurring later in the vowel than the onset of the jump category which can even occur before the vowel. The language difference for the rises was explained in terms of the two categories where French listeners favored the jump category leading to an earlier AB while Dutch and Swedish listeners used both types of rise categories for accentuation.

For the falls, the jump category was formalized as H+L* corresponding to Swedish word accent I following Bruce [4] and the accent-lending fall in the IPO approach to intonation [15]. The fall-onset category was formalized as H*+L corresponding to Swedish word accent II. For Dutch, 't Hart et al. [15] present only one category of full-sized accent-lending falls, while Rietveld and Gussenhoven [14] distinguish between two phonological categories. The difficulty of the French

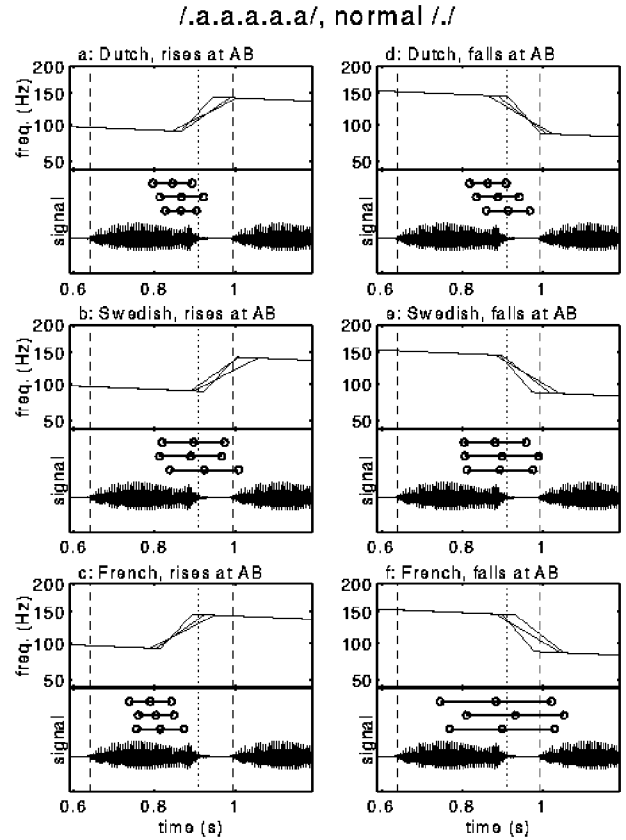


Figure 2: Accentuation boundaries (AB) obtained for /a.a.a.a.a/ stimuli with the intervocalic silent interval of normal duration for Dutch, Swedish and French subjects. For further explanations see caption of figure 1.

listeners to perceive falls as accentuation was explained by introducing a third category, a phrasing category (H%L). It was proposed that in some positions the falls seem to induce a strong percept of phrasing which may override the percept of accentuation (cf. [6, 7]). This interaction may also account for areas of ambiguity in the Dutch and Swedish responses. These areas of ambiguity have been further investigated for Dutch and were shown to be the result of individual response differences [9]. For French, it has been claimed that falls do not have a clear function in the accentual structure [1, 16, 17]. If the French listeners perceived phrasing categories (H%L) instead of accentuation categories, this would explain the ambiguity in the French responses for falls.

The purpose of the experiment reported here is to see if phrase boundaries (PB) can be elicited from Swedish listeners by using the same stimuli as in the accentuation-boundary experiments but with different instructions. If phrase boundaries can be elicited, we want to compare the cues involved and their locations with those for the accentuation boundaries and the high-low boundaries. This study will also further test the accentuation and phrasing categories proposed in [12].

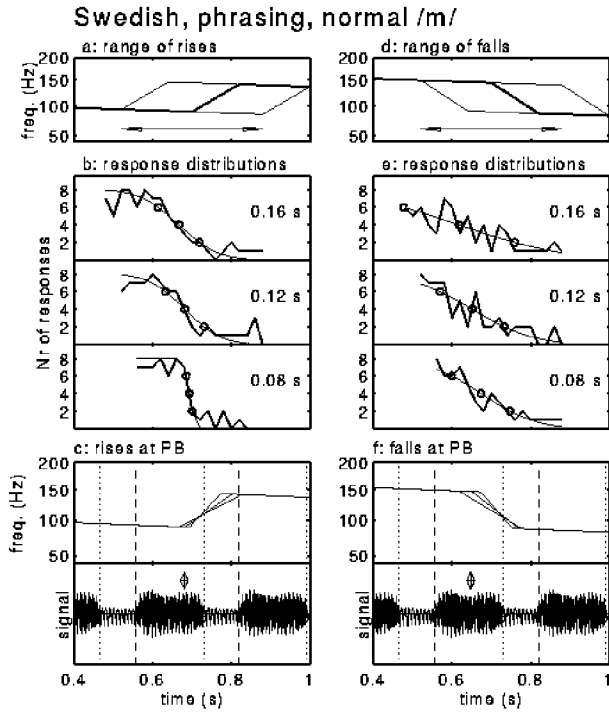


Figure 3: Collapsed results for eight subjects for the /ma/ stimuli. The range of the stimuli is plotted in (a) and (d). The response distributions (RDs) for the rises are plotted in (b) and those for the falls in (e). The three figures in the panels of (b) and (e) present the RDs for different durations of the pitch movement. The thin line indicates an error function fitted through these distributions. The circles indicate the quartiles of these error functions. Below, the rises (c) and the falls (f) are plotted at the phrase boundary (PB) as determined on the basis of the response distributions, i.e. the onset of these pitch movements coincide with the middle quartile of the error functions fitted through the distributions. The lower panel of (c) and (f) presents the oscillogram monotonized at 100 Hz. The dashed line represents the vowel onset, the dotted line the vowel offset, and the diamond indicates the location of the PB.

2. METHOD

Again, both /mamamama/ and /a.a.a.a.a/ stimuli were used, and will here be referred to as /ma/ and /a/ stimuli. The stimuli were derived from natural three-syllable utterances, spoken with an accent on the second syllable. Speech modifications were carried out with time-domain pitch-synchronous overlap/add (PSOLA) techniques [13]. The first step was to triplicate the middle syllable, resulting in five-syllable utterances of which the middle three syllables were identical as to duration and spectral envelope. Next, the original pitch contour was replaced by a rising or a falling pitch movement superimposed on a declination line. The range of timings of the pitch movements used in this experiment was such that the pitch movements

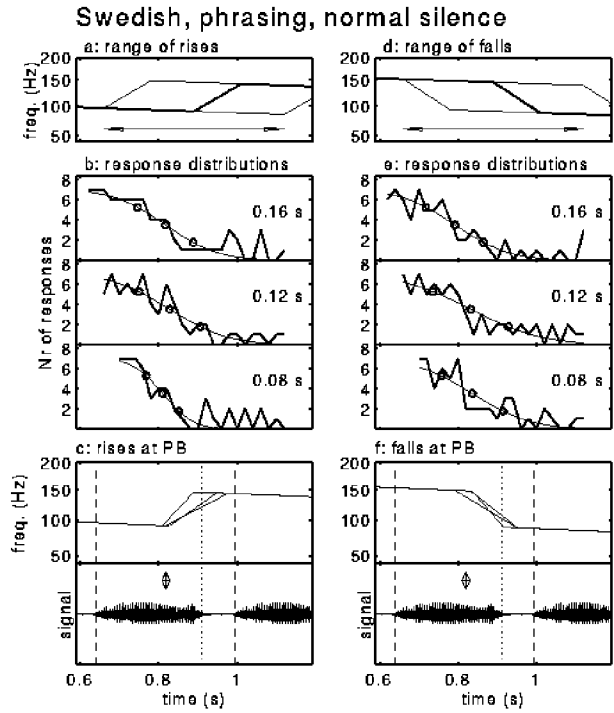


Figure 4: Collapsed results for seven subjects for the /a/ stimuli presented as in figure 3.

accented the third or the fourth syllable of the utterance. Only stimuli with normal durations of /m/ and intervocalic silent intervals taken from the natural utterances were used in this experiment.

Eight speakers of central Swedish participated in the experiment with /ma/ stimuli and seven in the experiment with /a/ stimuli. Five of the subjects participated in both experiments. Stimuli were presented twice over headphones at an interval of 0.5 seconds. There was then a silent interval of 2 seconds which provided time for subjects to indicate on a response sheet whether the series of syllables was grouped as 2+3 or as 3+2. The stimuli were presented in random order.

3. RESULTS

The results of eight subjects for the /ma/ and seven for the /a/ stimuli clearly indicate that phrase boundaries can be elicited using the stimuli from the earlier experiments on accentuation. Furthermore, the PB is located in the same general area as the AB, i.e. toward the end of the vowel. The overall results of the /ma/ stimuli are presented in Figure 3 and of the /a/ stimuli in Figure 4.

A comparison of the phrasing experiment results with the accentuation boundary results shown in figures 1 and 2 reveals a difference in the exact location of the respective boundaries for Swedish. The PBs for both the /ma/ and /a/ stimuli are located earlier in the vowel of S_n than are the ABs. The phrase boundary location for these Swedish listeners actually coincides

more closely with the accentuation boundary location for rises for French listeners as shown in figures 1 and 2, panels (c).

The sharpest PBs are elicited by the shortest pitch movement (80ms) as shown in the bottom panels of (b) and (e) in figures 3 and 4. For the rises, it is the onset of the pitch movement which defines the location of the PB. This is, however, not so clear for the falls. Furthermore, the PB for the falls are generally not as sharp as for the rises.

4. DISCUSSION

It is clear that there is an interrelationship between cues for accentuation and for phrasing. In Bruce et al. [6] different degrees of interdependency between tonal gestures for accentuation and phrasing in Swedish are exemplified. The results of this experiment indicate that the same tonal cues can quite effectively function as cues for both accentuation and phrasing, and in a similar manner.

Although the results presented here are from only eight subjects and must therefore be considered preliminary, the differences in location between the accentuation boundaries and the PB can lead to some speculation. If the PB are in fact earlier, it may be that onset of movement cues are stronger cues for accentuation than for phrasing. This would mean that, for example, the perceived onset of a rise in the final half of the vowel of S_n would cue an accent on S_n but be perceived as a jump cue for phrasing and cue the beginning of the new phrase on S_{n+1} . This would in turn mean that jump cues are stronger for phrasing. The jump cues may be more effective in signaling discontinuity between groups and therefore be stronger cues for phrasing than for accentuation [10, 11]. The jump cues may also be more effective for rises than for falls as a falling contour may be more likely to signal a boundary than a rise (final lowering). This would also account for the greater ambiguity for the falls than for the rises and be consistent with the accentuation results for French interpreted in terms of perceptual jump cues.

Further experimentation is planned where the phrasing experiment will be run with Dutch and French listeners. Additional experiments are also planned using subjects with different dialects of Swedish to check the possible influence of dialect differences in word accent realization on perception results. The results will have implications for our understanding of language dependent interplay between accentuation and phrasing and also for the placement of tonal rises and falls in synthesis of accentuation and phrasing.

5. REFERENCES

1. Beaugendre F. *Une étude perceptive de l'intonation du français*. Thèse de doctorat de l'université Paris XI, Orsay, 1994.
2. Beaugendre, F., House, D., and Hermes, D.J. "Accentuation boundaries in Dutch, French and Swedish," *Proceedings of the ESCA Tutorial and Research Workshop on Intonation, Athens, Greece*, 43-46, 1997.
3. Beckman, M.E. and Pierrehumbert, J.B. "Intonational structure in Japanese and English," *Phonology Yearbook* 3, 255-309, 1986.
4. Bruce, G. *Swedish word accents in sentence perspective*, Gleerup, Lund, Sweden, 1977.
5. Bruce, G. and Gårding, E. "A prosodic typology for Swedish dialects," in Gårding, E., Bruce, G. and Bannert, R. (eds), *Nordic Prosody*, 219-228, Department of Linguistics, Lund University, 1978.
6. Bruce, G., Granström, B., Gustafson, K. and House, D. "Phrasing strategies in prosodic parsing and speech synthesis," *Proceedings Eurospeech '93. Berlin, Germany*, 1205-1208. 1993.
7. Bruce, G., Granström, B. and House, D. "Prosodic phrasing in Swedish speech synthesis," In Bailly, G., Benoît, C. and Sawallis, T.R. (eds) *Talking Machines Theories, Models and Designs*, 113-125. Amsterdam: North Holland, 1992.
8. Hermes, D.J. "Timing of pitch movements and accentuation of syllables in Dutch," *Journal of the Acoustical Society of America* 102, 2390-2402, 1997.
9. Hermes, D.J., Beaugendre, F. and House, D. "Individual differences in accentuation boundaries in Dutch," *IPO Annual Progress Report* 32, 131-138, 1998.
10. House, D. *Tonal Perception in Speech*, Lund University Press, Lund, Sweden, 1990.
11. House, D. and Gårding, E. "Phrasing in some Nordic Dialects," In Gregersen, K. and Basbøll, H. (eds) *Nordic Prosody IV*, 61-70, Odense University Press. 1987.
12. House, D., Hermes, D.J. & Beaugendre, F. "Temporal-alignment categories of accent-lending rises and falls," *Proceedings Eurospeech '97. Rhodes, Greece*, 879-882, 1997.
13. Moulines, E. and Laroche, J. "Non-parametric techniques for pitch-scale and time-scale modification of speech," *Speech Communication* 16, 175-205, 1995.
14. Rietveld T. and Gussenhoven C. "The influence of syllable composition on the alignment of pitch targets," *Journal of Phonetics* 23, 375-385, 1995.
15. 't Hart, J., Collier, R., and Cohen, A. *A perceptual study of intonation. An experimental-phonetic approach to speech melody*, Cambridge University Press, Cambridge, UK, 1990.
16. Touati, P. *Structures prosodiques du suédois et du français*, Lund University Press, Lund, Sweden, 1987.
17. Vaissière, J. "La structuration acoustique de la phrase française," *Annali della scuola normale superiore di Pisa*, 530-560, 1980.