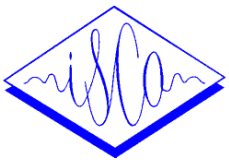

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

TIMING OF PITCH MOVEMENTS AND PERCEIVED VOWEL DURATION

H.H. Rump

Institute for Perception Research (IPO)
P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

ABSTRACT

The hypothesis was tested that the timing of accent-lending pitch movements influences the perceived duration of a vowel. Dutch subjects were asked to adjust the physical duration of a vowel so as to fit into the temporal structure of a sentence. The vowel occurred in a monosyllabic word embedded in a carrier sentence. Three pitch movements on the vowel were used, a rise, a rise-fall, and a fall.

Two opposite trends were found: the earlier the fall, the longer the duration of the target vowel was adjusted, the earlier the rise or rise-fall, the shorter its duration was adjusted. Control experiments indicated that the results should be interpreted in terms of a trade-off between the effects on prominence of timing of pitch movements and physical segment duration. It is concluded that late timing of pitch movements enhances the perceived vowel duration, but that this effect depends on the kind of pitch movement: the effect is cancelled in the case of late rises and rise-falls, whereas it is enhanced in the case of late falls by virtue of the enhancing effect on prominence of the accented syllable.

I. INTRODUCTION

The naturalness and acceptability of speech are to a large extent determined by the correct perceived duration (henceforth: "length") of the speech sounds. Several studies have shown that the length of vowels may not depend uniquely on their physical duration (henceforth: "duration"). One factor which appears to affect vowel length is pitch: In general, the presence of a pitch movement results in greater length [1,2,3]. In addition, there are reasons to assume that also the position of the pitch movement in the accented syllable may be of some importance for the perception of correct vowel length. From comparative studies between Dutch and German speech [4,5], there are indications that Dutch listeners when confronted with German, perceive a greater vowel length because the position of the pitch movement is later in the syllable than in Dutch. The relationship between timing of pitch movements and vowel length will be treated in greater depth.

In the present paper we thus focus on the timing of pitch movements as a factor which may influence the perceived vowel duration. We want to test whether Dutch listeners perceive differences in length between vowels containing differently timed pitch movements. In order to investigate the possible trade-off between timing of pitch movements and duration of a vowel on its length, an experiment is performed in which listeners adjusted the duration of a target vowel in a carrier phrase so that it fitted into the sentence, according to an internal acceptability criterion.

This specific kind of matching was chosen since a) this internal criterion was reported to be a very reliable reference, and b) two other methods had shown some major drawbacks. Firstly, when the method of matching with comparison stimuli was used, it had turned out that differences between adjusted durations disappeared in the course of the experiment [2]. It seemed that listeners learned to abstract from intonational phenomena; as a result, they directed all their attention towards the physical duration of the stimuli, which was constant. Secondly, when the method of paired comparisons was used, there was a large temporal order effect: The first stimulus within a pair always attracted relatively more "longer" judgments than the second one [1,3].

In the present experiment we found that differences in adjusted duration were small but consistent, and the results of additional control experiments indicate that variation in timing of accent-lending pitch movements has perceptually relevant consequences for vowel length in running speech. It appeared that vowels are judged to be longer under a late timing condition of pitch movements; in the case of falls this effect seems to be enhanced by virtue of extra prominence on the accented syllable due to the late timing of the fall.

II. EXPERIMENT 1: VOWEL DURATION ADJUSTMENT

Method

In experiment 1, we measured the adjusted duration of a target vowel under the condition of four different timing variants of three accent-lending pitch movements, rise, rise-fall, and fall. Subjects matched the length of the vowel to their internal criterion for optimal vowel length. The vowel (/i/) was embedded in the Dutch carrier sentence "Heb je LIEN laatst nog ontmoet?" ("Have you met LIEN recently?"). First, the sentence was recorded and stored in order to be manipulated. A male speaker was asked to read it aloud, realizing the sentence accent on "Lien" (a Dutch girl's name). The speech was digitalized (12 bits at 10 kHz sampling frequency) and stored in LPC-coded form (20 LPC coefficients; update rate: 5 ms). The duration of the vowel /i/ in the original utterance (110 ms) was manipulated by changing locally the update rate of the parameters. Nine versions of the utterance were prepared, only differing with respect to the duration of the target vowel. The duration values of /i/ ranged from 30 to 190 ms in steps of 20 ms.

After the duration manipulation the utterances were provided with the appropriate intonation contours, of which examples are shown in Fig.1. The contours consisted of linear interpolations in the log-frequency domain between turning points. The question-marking rise at the end of the utterance had a size of 6 semitones (st) and started 120 ms before the end of the last voiced segment in the utterance. The size of the pitch movements was always 7 st.

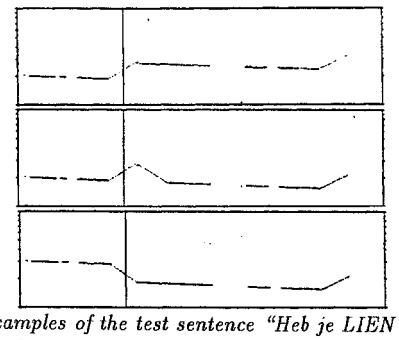

Figure 1: *Examples of the test sentence "Heb je LIEN laatst nog ontmoet?". Three pitch contours, containing a rise, a rise-fall, or a fall on the vowel /i/, are represented. The vertically drawn line indicates the onset of the target vowel.*

For each of the three accent-lending pitch movements four timing variants were prepared: "very early", "early", "late", and "very late". They differed with respect to both their starting point relative to the vowel onset and to their duration (see table I). Differences in rate of change between the pitch movements were too small to be noticed in running speech: They have to differ by at least a factor of two to be distinguishable.

Table I: *Timing of the pitch movements in ms. Onsets are relative to the onset of the target vowel /i/.*

	rise		fall in rise-fall		fall	
	onset	duration	onset	duration	onset	duration
very early	−120	120	30	120	−120	120
early	−70	120	80	120	−70	120
late	−60	180	150	180	−60	180
very late	0	180	210	180	0	180

The task for the subjects was an adjustment task, and the instruction was to adjust the duration of the target vowel making it fit into the temporal structure of the sentence. There were three sessions in which always the timing variants of only one pitch contour were tested. Adjustments alternately started at one of the extreme ends of the duration continuum ("downward" or "upward"). During a session each of the timing variants was presented twelve times (six times upward, six times downward), so that a session contained 48 trials (rise-fall: 36 trials)[1]. The replication data served the purpose of testing the consistency of the subjects. The total number of adjustments made by each subject was 132. No feedback was given.

Nineteen subjects participated in the duration adjustment experiment. All had normal hearing. They were tested individually. The stimuli were presented over headphones in a sound-treated room.

Results and discussion

In order to determine the internal consistency of the subjects' scores we defined the following criterion for subject selection. The dispersion of the observations within the same timing condition (replications) should be relatively small: as a measure for the dispersion we took the standard deviation (sd), which we related to the DL for vowel length. We assumed that this DL was not larger than about 20% of the vowel duration, i.e., < 15 ms (e.g., [3]). According to the criterion, the data of five subjects were discarded. The data of three other subjects were also discarded because of a so-called "floor effect". They adjusted the vowel duration to 30 or 50 ms in more than 70% of the cases, namely 73, 82 and 95%.

The pooled means and the standard error of the means for the remaining subjects are displayed in Fig.2. An overall analysis of variance of the data showed that the main effect of Timing Variants was not significant. However, the interaction effect of Pitch Contours x Timing Variants was highly significant ($F_{(4,20)} = 27.15$, $p < .001$). Because of this large interaction effect ANOVAs were performed for each pitch contour separately. They showed that the main effects of Timing Variants were significant: With the rise and the rise-fall, only the very early timing conditions caused the vowel to be adjusted to a significantly shorter duration than did the later ones (rise: $F_{(1,15)} = 17.73$, $p < .01$, rise-fall: $F_{(1,10)} = 24.73$, $p < .01$), but with the fall, the adjusted duration was significantly different for all timing conditions: The later the fall, the shorter the duration was adjusted (cf. Fig.2).

[1] At very short durations of the target vowel, the very late rise-fall mainly took place in the syllable following the word "Lien". As a consequence, there was an accent-shift. Therefore, the very late rise-fall was not included in experiment 1.

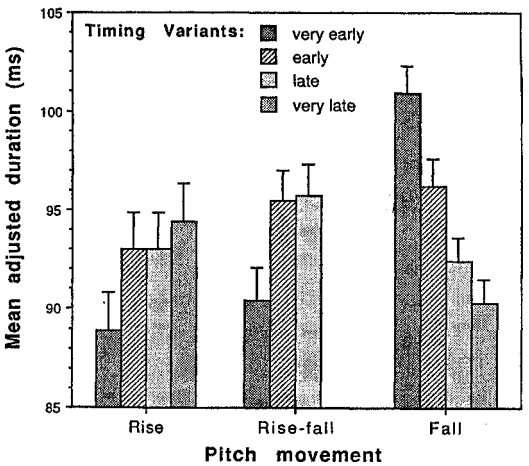

Figure 2: *Plotted means and standard errors of the means (error bars) of the adjusted vowel duration for each of the timing variants of rises, rise-falls, and falls (pooled data, 11 subjects, N=132).*

What surprised us most was the large interaction between timing of the pitch movements and the type of the pitch movements. In search for a possible explanation we found that timing of pitch movements may influence the perception of prominence of the accented syllable: From earlier research [6] it was concluded that a rise has to occur rather early in the syllable, whereas a fall has to occur relatively late in the syllable in order to make it sound prominent. Therefore, it seems plausible that the timing effects found in experiment were caused by differences in prominence-lending capacity of the pitch movements. Indeed, since a longer duration is one of the acoustical features of prominent syllables, we speculate that listeners have tried to keep the prominence relatively constant by adjusting the duration of the target vowel. This strategy may be regarded as some kind of compensatory lengthening or shortening in order to keep the amount of prominence constant. If this were true, we expect that an a somewhat exaggerated prominence, induced by some specific timing of a pitch movement, results in a short adjusted duration, whereas lack of prominence, induced by some specific timing of a pitch movement, results in a long adjusted duration.

In order to find out whether differences in adjusted duration found in experiment 1 reflect differences in the perception of prominence, we performed a control experiment. In a paired comparison experiment we tested whether listeners are able to indicate which timing variant of a certain pitch movement makes the syllable sound more prominent.

III. EXPERIMENT 2: CONTROL EXPERIMENT FOR PROMINENCE

Method

In experiment 2, listeners were presented with pairs of sentences from the adjustment experiment. The vowel /i/ had a fixed duration of 90 ms, but the members of a pair contained different timing conditions of the same pitch movement, i.e., the very early and very late timing variants. All different pitch movement types were tested. Each stimulus pair was presented 10 times (both orders five times). In addition, pairs containing identical timing variants of the same pitch movement were included. Sentence pairs containing different pitch movements were used as fillers.

The task of the subjects was to indicate on a score form in which of the sentences within a pair the syllable containing the target vowel sounded more prominent. Ten subjects participated in the experiment. Some of them had also participated in experiment 1.

Results and discussion

The often reported temporal order effect did not occur: the first stimulus in the pair containing identical stimuli is called "more prominent" in only 47% of the cases ($p > .05$ under the assumption of a binomial distribution). Therefore, the results are pooled over both orders of presentation.

The results of the paired comparisons are shown in Fig.3. For the rise and the rise-fall, the syllable containing the very early timing variant attracts significantly more "more prominent" judgments than the syllable containing the very late timing variant: 87% and 94%, respectively. For the fall, however, the target vowel attracts the majority of "more prominent" judgments under the condition of the very late timing variant: 69% ($p < .001$).

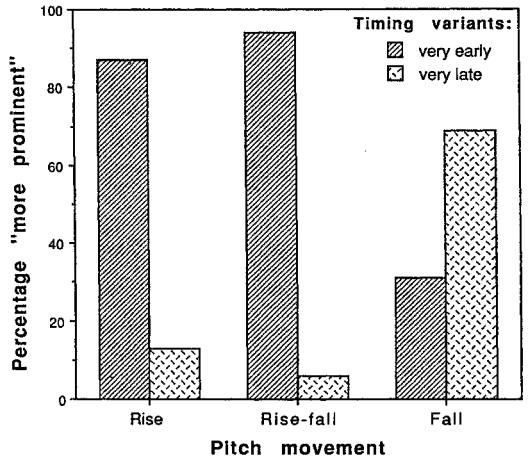

Figure 3: *Percentage "more prominent" judgments for the vowel under the very early and the very late timing condition , respectively, of each of the three pitch movements (N=100).*

The results show very clearly that prominence of the target vowel is strongly influenced by the timing of the pitch movements. When comparing the results of experiment 2 with the results of experiment 1 (cf. Fig.2), we notice that timing variants that enhance the prominence induce relatively short adjusted vowel durations, whereas timing variants that reduce the prominence induce longer adjusted durations. In conclusion, it appears that the timing of pitch movements has been a decisive factor in the adjustment experiment, but indirectly, namely by virtue of its effect on the prominence of the syllable containing the target vowel.

However, we suppose that differences in adjusted duration only partly reflect differences in prominence. If prominence were the only factor that determines to what duration the vowel is adjusted, we would expect that differences in adjusted durations are largest in the case of rises and rise-falls. They turned out, however, to be largest in the case of the falls. Therefore, we suppose that they also reflect differences in vowel length.

In order to test whether the expectation holds that vowel length has contributed to the adjustment of different durations in experiment 1, we set up a second control experiment, in which we tested whether subjects are able to indicate under which timing condition of a certain pitch movement the vowel is perceived as "longer" given a certain physical duration.

IV. EXPERIMENT 3: CONTROL EXPERIMENT FOR DURATION

Method

The method and materials were the same as in the first control experiment. But now, the task of the listeners was to indicate in which of the sentences within a pair the target vowel sounded longer. Eight subjects participated in the second control experiment. Some of them had also participated in the adjustment experiment, but none of them participated in experiment 2.

Results and discussion

Like in the previous control experiment, the temporal order effect did not occur: the first stimulus in the pair containing identical stimuli received only 53% of all "longer" judgments ($p > .05$). Therefore, the results are again taken over both orders of presentation. They are shown in Fig.4. For the rise and the rise-fall, the differences in "longer" judgments turn out not to differ from chance level (53 and 59%, respectively). For the fall, the difference in "longer" judgments is significant: the vowel is judged to be longer under the very late timing condition than under the very early timing condition (63%, $p < .05$).

The general trend found in this experiment is very clear (see Fig.4): later timing of a pitch movement enhances the length of the vowel on which the pitch movement takes place.

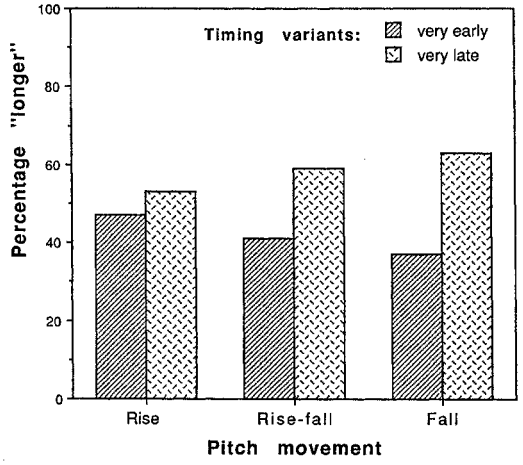

Figure 4: *Percentage "longer" judgments for the vowel under the very early and the very late timing condition, respectively, of each of the three pitch movements (N=80).*

It seems thus reasonable to assume that the effects of timing on length and on prominence are related in a certain way. When comparing the results of experiments 2 and 3 (Figs.3 and 4), one can see that in the case of rises and rise-falls, the effect of timing on prominence is in the opposite direction compared to the effect of timing on vowel length, whereas in the case of falls, it is in the same direction. The combined effect of timing on length and prominence may thus explain for the fall the clear trend in the results in experiment 1 and the significance of length differences found in experiment 3. It may also explain why the results in experiments 1 and 3 were less clear for the rise and rise-fall.

In conclusion, we suppose that the different influence of the nature of the pitch movements on timing can be explained by considering the combined effect of timing on length and prominence.

V. GENERAL DISCUSSION

In the experiments, in which we investigated the relationship between the timing of pitch movements and adjusted vowel durations, we found that Dutch listeners adjusted significantly different durations of the target vowel under different timing conditions. However, differences in adjusted durations were relatively small. In our view, there are two possible reasons: 1) the vowel /i/ is phonologically short in Dutch, so that the range of adjustable durations is limited: Only relatively short adjusted durations resulted in acceptable vowel length; 2) the target vowel had been embedded in a carrier sentence. It seems plausible that the range of possible adjusted durations is constrained by the rhythmical structure of the utterance: It should stay intact in order to make the utterance sound natural [7].

There were two major consequences originating from the fact that the range of adjustable durations was quite small. First, only experienced subjects were able to perform the adjustment task consistently; the data of the other subjects had to be discarded. Secondly, the step size between two adjustable durations was relatively large when compared to the effect of timing on length. As a result, some of the subjects complained that they could not adjust the right duration. Choosing a smaller step size, however, would have made the task even more difficult for the other subjects.

Differences in means of adjusted durations were found to range up to about 12 ms. However, for the individual subjects these differences were somewhat larger (up to about 20 ms). Differences of that magnitude are supposed to be larger than the DL for vowel duration, so that it is plausible that differences in timing of pitch movements are perceptually relevant for the perception of length of vocalic speech sounds.

In the present experiments, another important perceptual aspect of timing of pitch movements in running speech may have been revealed. It was found that difference in timing caused differences in prominence: One would expect that, in general, differences in length of the magnitude of the DL are perceptually too small to outweigh these differences in prominence. However, if we assume that listeners have tried to adjust equal prominence by lengthening or shortening of the target vowel, one may conclude that the relatively large differences in prominence, caused by the differences in timing of the pitch movements, were fully compensated for by the relatively small differences in adjusted (physical) duration of the vowel. This would imply that listeners (at least trained ones) are quite sensitive for differences in vowel length, although this is a consequence of their (much larger) sensitivity for differences in prominence. Because of its influence on prominence, the correct timing of pitch movements will thus be of great importance for synthesizing natural sounding intonation.

Further research may shed light on the above suggested relationship between perception of duration and prominence in running speech. Another important question to be answered is why timing differences resulted in differences in prominence. In future experiments, a larger array of timing variants of the various pitch movements compared to the present one, may be covered.

VI. CONCLUSIONS

In the present report we studied the relationship between timing of pitch movements and vowel length (i.e., its perceived duration). In the first experiment we found that timing of pitch movements had a significant effect on the duration to which the target vowel was adjusted. Timing differences resulted in a significantly different duration adjusted by Dutch listeners, although the influence was not the same for all timing variants and all individual subjects. In general, early falls resulted in long adjusted durations, whereas early rises and rise-falls resulted in short adjusted durations. From the results of the control experiments it turned out that differences in adjusted duration were mainly caused by differences in prominence of the syllable containing the target vowel. The target vowel is adjusted to a relatively short duration when the timing of a pitch movement makes the syllable sound too prominent in the given context, and vice versa.

Our main conclusion is that it is plausible that the length of a vowel in an accented syllable can be enhanced by the prominence induced by the specific timing of the pitch movement: A relatively late timing of the fall, which lends much prominence, results in a longer perceived vowel duration than a relatively early timing. However, such a simple relationship between timing of pitch movements, prominence, and perceived vowel duration does not hold for rises and rise-falls: The enhancing effect of late timing on length is more or less cancelled by the enhancing effect of early timing on prominence.

Acknowledgments

I would like to thank René Collier, Jacques Terken, and Marc Swerts for their helpfull comments.

References

[1] Lehiste, I. "Influence of fundamental frequency pattern on the perception of duration," *Journal of Phonetics*, vol. 4, pp. 113-117, 1976.

[2] Rosen, S. M. "Linear FM pitch sweeps and perceptual duration in speech and non-speech," Speech Transmission Lab., Stockholm, Sweden, *Quarterly Progress and Status Report*, vol. 4, pp. 1-12, 1976.

[3] Rosen, S. M. "The effect of fundamental frequency patterns on perceived duration," Speech Transmission Lab., Stockholm, Sweden, *Quarterly Progress and Status Report*, vol. 1, pp. 17-30, 1977.

[4] Van Dommelen, W. A. *"Temporale Faktoren bei ausländischem Akzent. Eine kontrastive deutsch-niederländische Untersuchung zur Produktion und Perzeption von Segmentdauerwerten,"* Ph.D. Thesis, University of Leyden. 1980.

[5] Adriaens, L. M. H. *"Ein Modell deutscher Intonation. Eine experimentell-phonetische Untersuchung nach den perzeptiv relevanten Grundfrequenzänderungen in vorgelesenem Text,"* Ph.D. Thesis, Technological University of Eindhoven. 1991.

[6] Van Katwijk, A. & Govaert, G. A. "Prominence as a function of the location of pitch movement," Inst. Perception Res., Eindhoven, The Netherlands, *Annual Progress Report*, vol. 2, pp. 115-117, 1967.

[7] Huggins, A. W. F. "On the perception of temporal phenomena in speech," *Journal of the Acoustical Society of America*, vol. 1, pp. 1279-1289, 1972.