TWO KINDS OF STRESS PERCEPTION

Thomas Portele, Barbara Heuft
Institut für Kommunikationsforschung und Phonetik, Universität Bonn

ABSTRACT

This paper describes some preliminary results concerning the perception of syllable stress as either a binary feature or a nearly continuous parametric value. Two experiments were set up: a perception test where the subjects were forced to assign stress to one of two syllables, and a labelling experiment. Here, the subjects had to rate the degree of stress carried by a syllable using values between 0 and 32.

MOTIVATION

Word accent may serve as a distinctive cue in discriminating two words that are identical on the segmental level (in English, for instance, <pro'cess> and <'process> ; in German <'umlaufen> (to run over something) vs. <um'laufen> (to run around something)). An accented syllable may also serve as a first guess in segmenting the speech signal into words [1]. To perform these functions a syllable must be perceived as either stressed or unstressed. This implies that some kind of categorical perception takes place.

On the other hand, listeners are able to distinguish between syllables regarding the amount of stress they carry. The perception of focus accents is an indication for this ability. Fant and Kruckenberg [2] used a 30 point scale for subjective judgements of syllable stress and obtained reliable results.

To obtain an impression about how strong these abilities are developed in German listeners two experiments were set up. The first experiment was designed to assess the listeners' abilities in assigning stress to one of two syllables when one parameter, i.e. the position of the F_0 peak, is gradually changed. A similar experiment was carried out by Kohler [3] but he explored not lexical but semantic categories. In the second experiment three listeners judged more than 8500 syllables regarding their amount of stress. They used a scale from 0 to 31. The correlations between their ratings were evaluated as well as possible factors guiding their judgements.

EXPERIMENT 1

Method

The pairs <voll Milch> (full of milk) - <Vollmilch> (whole milk) and <zwei Räder> (two wheels) - <Zweiräder> (bicycles) were embedded in short texts and read by a male (<Vollmilch - voll Milch>) and a female speaker (<Zweiräder - zwei Räder>). A previous experiment [4] showed that the position of an F_0 peak serves as a cue to syllable stress for these stimuli, and that a movement of the peak to the other syllable often leads to a different stress assignment by listeners.

For each sentence containing one of the syllable pairs the F_0 contour was parametrized using a method by Portele et al. [5]. This method describes an accent by four numerical values, the position of the F_0 maximum being one of them. This parameter was modified and the position of the F_0 maximum was shifted towards the other syllable in 6 steps with a stepwidth of 50 ms (Figure 1). In the vicinity of the position where the stress shift was supposed to take place 10 steps with a stepwidth of 10 ms were used. These 15 intonation contours were imposed on the original sentence using the PSOLA method. Altogether, 60 stimuli were used. They were presented to the subjects (n=13) via headphones in a quiet room. The subjects' task was to decide which orthographic representation was correct for a given stimulus, e.g. whether <Vollmilch> or <voll Milch> was spoken. The linguistic content of the stimulus sentences was neutral to both interpretations. Each stimulus was presented

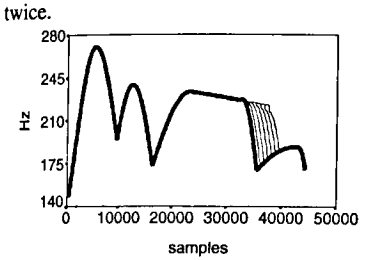

Figure 1. Parametrized intonation contour for the utterance "In diesem Keller wurden nachweislich Zweiräder montiert." (Provable, bicycles were assembled in this cellar). The thick line indicates the original contour, the thin lines show the shift of the F_0 peak in steps of 50 ms to the right.

Results

The results are displayed in Figure 2. Each picture displays in the upper half the individual scores of those subjects where the difference between the ratings to the left and the right from the vertical line (point of stress shift) is significant (t-test, $p < 0.075$). In the lower half (between 0 and 1) the pooled results from all subjects are shown. Rating 2 in the upper half and 1 in the lower half stands for perceiving the first syllable as stressed, rating 1 in the upper half and 0 in the lower half for preceiving the second syllable as stressed. The vertical line indicates the position of the F_0 peak where the change in perception occurs for most subjects (there are some individual differences). The subjects performed differently; some subjects obtained highly significant results for all four stimulus groups (<Vollmilch>: 10 subjects, <voll Milch>: 6 subjects, <Zweiräder>: 7 subjects, <zwei Räder>: 10 subjects) while others behaved completely erratic.

Figure 3 shows the results from one subject. Here, the position of the F_0 peak where the stress shift takes place, is clearly recognizable.

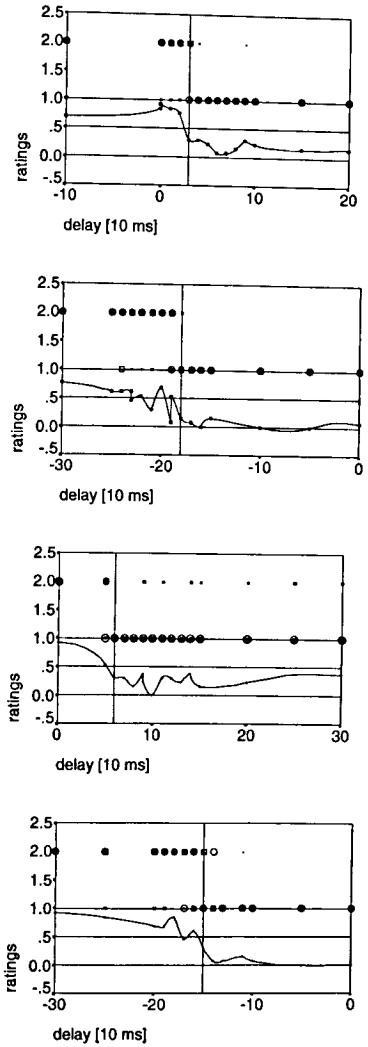

Figure 2. Results of the stress assignment experiment. The original stimulus was <Vollmilch>,<voll Milch>,<Zweiräder>, <zwei Räder> (from above). Further explanation is given in the text.

Discussion

The results show that the identification of the stressed syllable in suitable syllable pairs depends on the position of the F_0 peak. When this peak is moved towards the originally unstressed syllable there is a

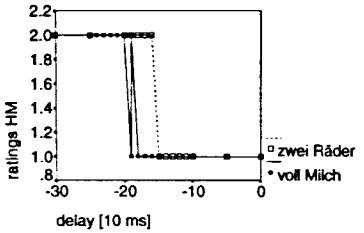

Figure 3. Judgements by one subject for the stimuli <zwei Räder> and <voll Milch>. Rating 2 stands for "first syllable stressed", rating 1 for "second syllable stressed".

certain point where the perception of the stressed syllable switches from one syllable to the other. However, not all subjects were consistent in their judgements. As shown in [4], for the material used in this experiment syllable duration is a stress cue as strong as F_0. The apparent mismatch between these two cues might be responsible for the erratic behaviour shown by some subjects. For other subjects (Figure 3) the position of stress switch can be located within 10 ms. This interval is a little longer than one pitch period.

These results are ony preliminary. A lot remains to be done, i.e. identifying the causes for the different behaviour of the subjects, looking for a connection between the position of the switch point and the properties of the speech signal, checking the discriminative ability of the subjects as the second requirement for categorical perception etc. However, the results show that, in German, the exact placement of the F_0 peak is crucial for correct perception, and, therefore, for intonation modelling in speech synthesis [3]. Our description of F_0 contours [5] was designed specifically to meet these requirements.

EXPERIMENT 2

Method

The corpus used in this experiment is part of a prosodic database [6]. More than 300 sentences were read by one male and two female speakers. Altogether, 8646 syllables were used. Their amount of stress was rated by three listeners on a scale between 0 and 31 [2]. The speech signal and the transcription were presented. They used a graphical scale to indicate the prominence level. They were allowed to listen to an utterance as often as they liked.

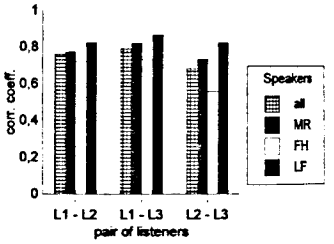

Figure 4. Correlation coefficients between the listeners' judgements.

Results

The correlation coefficients between the listeners' judgements are diplayed in Figure 4. The agreement between the subjects is slightly lower for speaker FH. An average correlation coefficient of 0.75 indicates the high similarity between the ratings. The agreement between the listeners was higher for prosodically marked utterances (orders or yes/no questions) than for simple statements.

The relation between acoustic properties of a syllable and its perceived prominence was also investigated. Figure 5 displays the connection between syllable duration and prominence rating.

There is a marked dependency between the existence of an F_0 peak asscociated with a syllable and the syllable's perceived prominence (U-test, $p < 0.001$). The average difference between syllables with and without F_0 peak is 12 prominence grades.

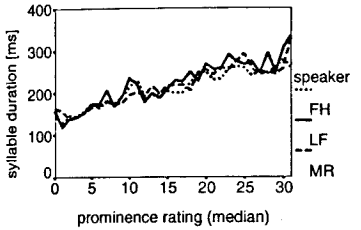

Figure 5. Relation between prominence rating and syllable duration displayed for each speaker.

The relation between the height of an F_0 peak and the perceived prominence is not very strong (but significant; Kruskal-Wallis-test, $p < 0.001$).

It was found that the offset between the start of a stressed vowel and the asscociated F_0 peak is significant for offset values greater than zero. Larger offsets (late peaks) induced higher prominence values [3].

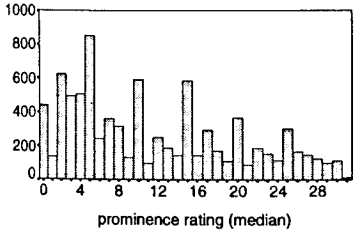

Figure 6. Histogram of the prominence ratings by all listeners.

Discussion

The results show that the listeners are able to differentiate between more than two or four levels of syllable prominence. How many prominence levels are sufficient, however, can not be deduced from the data, because the distribution is quite even (Figure 6). However, one can assume that judgements were made relative to other syllables in the utterance. It is unlikely that an isolated presentation of the syllables would have led to similar results. But this procedure would be far away from a real communication situation.

The acoustic properties of the speech signal have some measurable influence on the perceived prominence. However, linguistic factors that could be deduced from the transcription by the listeners seem to play a more important role [7].

CONCLUSION

The two experiments show that stress can be perceived as a binary feature and as a multi-level parameter of a syllable. Both kinds of perception are necessary for efficient communication.

ACKNOWLEDGEMENT

This research was partly funded by the *Deutsche Forschungsgemeinschaft* and by the *Verbmobil* project sponsored by the *Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie.* We thank Bernd Möbius for the texts used in the first experiment, Monika and Florian for painfully judging 8646 syllables, and all other subjects and speakers.

REFERENCES

[1] Cutler, A.; Norris, D. (1988), "Te role of strong syllables in segmentation for lexical access." J. Exp. Psych. Human Perception and Performance **14**, 113-121

[2] Fant, G.; Kruckenberg, A. (1989), "Preliminaries to the study of Swedish prose reading and reading style." STL-QPSR **2/1989**, 42-45

[3] Kohler, K.J. (1987), "Categorical pitch perception." Proc. XIth ICPhS, Tallin, 91.2

[4] Heuft, B.; Portele, T. (1994), "Zur akustischen Realisierung des Wortakzents." Proc. Elektron. Sprachsignalverarbeitung V, Berlin, 197-204

[5] Portele, T.; Krämer, J.; Heuft, B. (1995), "Automatische Parametrisierung von Grundfrequenzkonturen." (to appear in Proc. DAGA-95, Saarbrücken).

[6] Heuft, B., et al. (1995), "Parametric description of F_0 contours in a prosodic database." Proc. ICPhS'95, Stockholm

[7] Heuft, B., et al. (1995), "Betonungsstufen von Silben und ihre Beziehung zum Sprachsignal." (to appear in Proc. DAGA-95, Saarbrücken).