LOCAL PROMINENCE OF ACOUSTIC AND PSYCHOACOUSTIC FUNCTIONS AND PERCEIVED STRESS IN FRENCH

P. Mertens

K.U.Leuven, Linguistics Department, Leuven, Belgium.

Abstract

Syllable duration, pitch, loudness, pause length, pitch change, and local difference values for the first 3 parameters, were studied for their ability to predict perceived stress as measured in a listening task. The best cues were duration increase relative to preceding and following syllable(s), followed by nucleus duration.

1. Introduction

Syllabic stress is a linguistic attribute realized in various ways, with or without prominence. It can not be observed directly. A measure of perceived prominence has to be established in order to classify the syllables. A brief review of terminology will clarify this point.

(1) A syllable is prominent when it stands out from its context due to a local difference for some prosodic parameter. Prominence is continuous (not categorical) and contributions of multiple parameters can interact.

(2) Stress is an abstract linguistic category, which can be realized by several types of prominence, in a way which is language-specific.

(a) In French, an intra-syllabic pitch glide of a given interval suffices to signal stress. Prominence by duration or loudness will be functionally redundant although very common.

(b) For static syllables prominence will results from an intersyllabic change of a parameter.

(c) Finally, stress can result from tone level itself, on the basis of tone distribution [3,6].

(3) Word stress (lexical stress) indicates the syllable in a word which can receive stress.

(4) French has two stress types: final (word stress position) and initial stress (emphatic), with a different distribution.

In a listening task, the stress judgment will be based on a mixture of heterogeneous factors: acoustic, structural, lexical. Subjects may focus on an isolated factor, or on many; they find it very difficult to separately rate prosodic parameters. The test can show how untrained subjects judge stress, and whether they agree. Given the continuous nature of prominence, a stress score, the number of listeners that perceived a syllable as stressed [1,2], allows for a classification in min. 3 categories: stressed, unstressed, ambiguous.

Because of space limitations, previous studies on stress perception and stress cues can not be reviewed here.

2. Method.

Six extracts (277 syll.) were selected from a corpus [3] in such a way that the test contained at least 2 occurences of each stressed tone. A male and a female speaker each provided 3 extracts. The mean length of 46 syll/test was suggested by [2] where it was found that the proportion of syllables judged stressed decreases as the length of the carrier sentence increases. For lengths above 40 syll. the ratings are simular to those for continuous speech. The passages were very different in

terms of prosodic complexity (table 1), which can be defined in terms of (1) rate of speech (without pauses), (2) proportion of stressed syllables, (3) of emphatic stresses, (4) of pauses, (5) of glides, and (6) rhythmic structure.

Table 1. Agreement between 20 raters, and prosodic complexity, (judged by phonetician) for tests 1 to 6.

TEST(S)	1	2	3	4	5	6	1-6
P(A)	.7667	.7705	.8333	.6939	.8590	.7756	.7867
P(E)	.6704	.6691	.7086	.6194	.6195	.6009	.6558
Kappa	.2922	.3064	.4281	.1958	.5429	.4379	.3804
Complexity	medium	medium	low	high	low	medium	

2.1. Perceptual experiment.

The 20 untrained subjects heard each passage once (with 25s silence) and 6 times (with 6s intervals) during which they had to indicate the stressed syllables on the test sheet.

Each syllable was judged either stressed or unstressed; so, it was assigned to 1 out of 2 categories. The nominal scale calls for a non-parametric test: the kappa statistic [9] was used. P(A), the proportion of times that the raters agree, and P(E), the proportion of estimated chance agreement, are determined. The kappa coefficient is the ratio of P(A) to the maximum proportion of times that raters could agree, both corrected for chance agreement. A kappa 1 indicates complete agreement, a 0 indicates no agreement other than chance. Since only 2 categories are used here, chance agreement is high, and kappa rather low (table 2). The pooled data (P(A)=.7867, kappa=.38) show a moderate agreement among the listeners, although significantly different from 0. The relation with prosodic complexity is obvious.

2.2. Acoustic measurements.

For each syllable, 5 primary attributes are obtained, using an interactive analysis program [3,5]: nucleus DURATION, PITCH peak, LOUDNESS peak, intra-syllabic GLIDE, PAUSE duration. The segmentation into syllabic nuclei [4] provides boundaries necessary for the parameter extraction and pitch contour stylization. PITCH is the peak and GLIDE the interval of the stylized contour, positive or negative according to slope. Pitch values are expressed in semitones (ST): the melodic (in mel) and harmonic (in ST) scales are almost identical in the F0-range of speech [10]. The results were hand-corrected where necessary.

The measurement of LOUDNESS [10,8] (in soneG) accounts for frequency dependence, critical bands, frequency masking, level, but ignores the effect of stimulus duration. Level values (dB SPL) for each critical band were obtained from the power spectrum (512pt FFT, 40ms), by summation of the components in the band range, and dB-conversion.

Prominence estimates were calculated for duration, pitch and loudness. Prominence is defined as the difference between the parameter value for a syllable and the parameter mean of the context, either left (L) or right (R), with length 1 and 2 syll., giving 4 relative values: resp. DL1, DL2, DR1 and DR2 for duration, PL1, PL2, PR1, PR2 for pitch, and LL1, LL2, LR1 and LR2 for loudness. This allows for a continuous scaling of prominence. A similar measure combining left and right contexts with length 1 was used in [7].

3. Results

Scatter diagrams were made for the 17 attributes, with stress SCORE as the dependent variable. Some results were predictable: PITCH varies randomly with stress score

Table 2. Correlation between stress SCORE and parameters (above line) prominence measure (below line), for the pooled data.

DURATION				PITCH				LOUDNESS				GLIDE	PAUSE
.477				.203				.299				.070	.296
DL1	DL2	DR1	DR2	PL1	PL2	PR1	PR2	LL1	LL2	LR1	LR2		
.49	.49	.41	.41	.47	.44	.45	.48	.31	.30	.43	.36		

Table 3. Mean values for 7 variables cross-tabulated with ranges for SCORE. N is the number of syllables in a group.

SCORE	N	DUR	DL1	DR1	PL1	PR1	LL1	LR1
0-20	277	88	0	0	0.1	0.1	-0.2	0.2
0- 3	194	72	-26	-19	-1.2	-0.9	-1.2	-1.0
4-11	50	109	45	25	1.6	1.7	2.1	1.6
12-20	33	156	80	77	3.6	4.8	2.5	5.4

(because of speaker's range, declination line, etc.) and so does LOUDNESS. There are too few cases of glides and pauses to find a relation with SCORE.

Although no clear linear relation was found, Pearson correlation coefficient r was used to estimate the amount of information that could be gained from each variable (table 2). r varies considerably from one passage to another: for DURATION, from .63 to .18. Test 4 (with high complexity) gives very poor correlation for all attributes and is to a large extent responsible for the low r in the pooled data.

DURATION is the only primary parameter with relatively high r: this can be explained by minimal syllabic duration, small variability for unstressed syllables and large for the stressed.

The best prominence estimates are DL1 and DL2, indicating that syllables with high SCOREs are generally longer than the preceding one(s). DL2 and DR2 give results close to DL1 and DR1. LOUDNESS, LL1 and especially LR1 score quite good (r=.5) in some tests, but not on the average.

Depending on the method used and the number of variables taken into account, multiple regression gives a correlation of .60 to .88 with the stress SCOREs.

Stress score can be used to classify the syllables in 3 groups: not prominent, ambiguous, and prominent (table 3), showing clear differences between groups. The choice of the ranges depended on the number of elements in each group.

Labeling according to the transcription by a phonetician gives a further classification (table 4). Group means show that the stressed are twice as long as unstressed; they are prominent by duration (DL1,DR1) and, in the case of low stressed, also by loudness (LL1,LR1). PL1, PR1 and GLIDE reflect the tones used (H,L,HL,LH,L-,H+). The values for emphatic stress are very close to those for unstressed syllables. The parameters do not reflect the evident phonatory effort of emphatic stress.

Predictions by the intonation model are observed in the data: (1) syllables with extra-low tone (L-) can be short and weak because their stressed status is already indicated by tone level, (2) glides (HL,LH) lack loudness prominence because stress is already signaled by the glide.

4. Conclusion

A listening task provided ratings of perceptual prominence for 277 syllables. The relative agreement between the raters indicates the perceptual reality of prominence. The importance of acoustic parameters as well as of four prominence measures were studied. The stress scores by the listeners are best pred'cted by durational prominence relative to the preceding 1 or 2 syllables, and by syllabic duration itself. When the transcription of intonation by a phonetician is used for syllable classification, the same order of importance for the studied parameters is found. Predictions by the intonation model on the relative importance of individual prosodic parameters depending on the tone used, are confirmed by the data.

Table 4. Parameter means for syllables classified according to transcription by phonetician. The mean score by the untrained listeners is shown under SCORE. N is the number of elements in a category.

	N	DUR	DL1	DR1	PL1	PR1	LL1	LR1	GLIDE	SCORE
EMPHATIC	15	72	-19	-12	2.8	2.5	-1.3	1.4	0.0	6.3
STRESSED	74	137	70	65	2.6	2.9	1.6	2.7	0.0	10.6
H	28	135	64	53	4.3	4.7	1.4	3.5	0.3	11.3
HL	3	186	120	115	5.0	2.6	-0.3	-1.3	-4.3	11.0
L	25	139	74	77	1.8	2.2	3.7	4.3	-0.6	9.1
LH	4	228	168	133	4.2	3.7	2.0	0.7	7.2	13.2
H+	3	103	56	56	10.6	9.6	4.0	2.6	0.0	17.3
L-	11	100	32	33	-3.0	-1.9	-2.7	-0.6	-1.2	9.4
UNSTRESSED	188	71	-26	-24	-1.4	-1.1	-0.8	-0.9	-0.1	1.8
h	13	61	1	-46	-0.3	1.6	0.3	1.0	-1.0	4.6
l	168	71	-29	-23	-1.5	-1.4	-1.0	-1.0	0.0	1.6
l-	7	79	-20	-7	-1.1	0.1	-0.2	-1.2	0.4	2.4
POOLED	277	88	0	0	0	0.1	-0.2	0.2	0	4.4

5. References

[1] Allen, G.D. (1972a) The location of rhythmic stress beats in English: an experimental study (Part I.), Lang. & Speech 15(1), 72-100.

[2] McDowall, J.J. (1974) The reliability of ratings by linguistically untrained subjects in response to stress in speech, J. Psycholing. Res. 3, 247-259

[3] Mertens, P. (1987a) L'intonation du français. De la description linguistique à la reconnaissance automatique. Unpubl. Ph.D.

[4] Mertens, P. (1987b) Automatic segmentation of speech into syllables, Proc. Eur. Conf. on Speech Techn., II, 9-12.

[5] Mertens, P. (1989) Automatic recognition of intonation in French and Dutch, Eurospeech 89, I, 46-50

[6] Mertens, P. (1990) Chapitre IV. "Intonation", in Blanche-Benveniste, C. et al. (1990), Le français parlé, Paris: Ed. du CNRS, 157-176.

[7] Gaitenby, J.H. & Mermelstein, P. (1977) Acoustic correlates of perceived prominence in unknown utterances, SRSR-49, 201-216.

[8] Paulus, E. & Zwicker, E. (1972) Programme zur automatischen Bestimmung der Lautheit aus Terzpegeln oder Frequenzgruppenpegeln, Acustica 27(5), 253-266.

[9] Siegel, S. & Castellan, N.J. (1988) Nonparametric Statistics, NY: McGraw-Hill.

[10] Zwicker, E. & Feldtkeller, R. (1981) Psychoacoustique. L'oreille, recepteur d'information, Paris: Masson.