THE ROLE OF STRESS AND ACCENT IN THE PERCEPTION OF RHYTHM

Cynthia Grover and Jacques Terken
Institute for Perception Research (IPO), Eindhoven, The Netherlands

ABSTRACT

We ask whether stress or accent alone, or in combination, induce the perception of rhythm. According to Grønnum [1], perceiving speech rhythm depends on the distribution of accented syllables, and durational variation is unimportant. We report an experiment on perceiving speech rhythm whose outcome confirms the primary role of accent in determining rhythmicity.

INTRODUCTION

Phoneme durations are highly variable [2], and so it seems unlikely that their duration is tightly controlled. Perhaps their durational variation is constrained by higher level influences, as our own recent work on German suggests [3]. One such influence could be speech rhythm.

The definition of speech rhythm is problematic. The usual implication of the words is regular alternation of two sorts of units, one more prominent and (at least) one less so. Often, this contrast is labelled in terms such as strong versus weak, eg.[4], stressed versus unstressed, eg.[5], or accented versus unaccented, eg.[1]. We use 'accent' to mean 'pitch accent', and stress to mean 'without pitch accent and bearing lexical stress'. Our purpose is to define rhythmicity, and so to provide a better basis for discussing rhythmic constraints in speech.

We also wish to gather information about the sources of the perception of rhythm. Specifically, to what extent do stress and accent induce rhythmicity, or the perception that speech is rhythmic?

We suppose that listeners might consider stimuli that can be analysed into regular metrical linguistic units (ie. feet) to be rhythmic. We ask whether this foot structure is defined more clearly by accents or stresses.

Our hypotheses are as follows. Stresses and accents are assumed to serve as the heads of feet.

1. A uniform foot structure (always the same number of unstressed syllables per foot) and uniform head type (always accent or always stress) should give rise to a clear perception of rhythmicity.
2. Variation in the foot structure or head type within the phrase should induce less rhythmicity than when these are regular. This variation could take two forms: A. The number of unstressed syllables per foot varies, or B. the type of head varies (mixed accent and stress).
3. Variation in both foot structure and head type should induce the least rhythmicity.
4. We assume that subjects initially hypothesize foot structure based on the initial part of a stimulus. We therefore propose that inconsistent foot structure early in a stimulus is more salient than variation late in a stimulus, so early variation should be less rhythmic.

METHOD

We wanted to know how rhythmicity related to perceived foot structure and head type, so we had subjects perform 3 tasks on reiterant "mamama.." stimuli: A. rating their rhythmicity, B. picking prominences, and C: placing boundaries.

Subjects

Eighteen Dutch adults participated in two half-hour experimental sessions.

Materials

A male native Dutch speaker produced series of the syllable "ma" with pitch accents and stresses occurring at regular intervals (ie. every third or fourth syllable). We selected a typical exemplar each of accented, stressed, and unstressed syllables, and concatenated them to produce 8 and 9 syllable strings of reiterant "mamama.."'s. Table 1 shows the syllable types' specifications.

Table 1. Syllable (ma) specifications. D: duration (ms); F_0: Hz; A: amplitude (N/m^2).

	D	F_0max	F_0min	Amax	Amin
Accent	269	85	126	12830	-12910
+Stress	216	85	89	4991	-4402
-Stress	197	85	88	2935	-3565

Nine full "mamama.." strings were concatenated, and from these an additional 9 truncated strings were made by removing the last syllable of the string. The patterns are given in Table 2.

Procedure

Dutch subjects performed 3 tasks: rhythmicity rating, prominence picking, and grouping in each of two sessions. They heard the stimuli through headphones. Each task began with 3 practice stimuli. Then for each of the 3 tasks, the 18 strings were called up by a control program. Subjects had to complete the indicated task for each string before hearing the next string.

Subjects rated rhythmicity on a scale of 1 to 10 (10: extremely rhythmic). They could pick as many prominences or boundaries as desired by clicking a mouse to select syllables or boundaries (represented graphically on screen).

Analysis

We performed repeated measures multiple regressions. For every syllable the expected response ("prominent" for stresses or accents, and "not prominent" for unstressed syllables) was registered in the datafile for matching against the actual subject response.

RESULTS AND DISCUSSION

Before all analyses, we removed the variance due to control variables: individual differences and the order of presentation of the stimuli.

Rhythmicity

As predicted, there was a significant difference in the ratings for the different accent patterns, $F_{(8,621)}=26$, $p<.01$. Patterns with consistent head type or foot structure or both (strings 1 to 8 (S1-8)) were judged to be more rhythmic than those with variable foot structure and head type (S9), $F_{(8,621)}=23.5$, $p<.01$ (see Table 2). However, the clearest division between ratings fell between strings 1-6 versus S7-9, $F_{(1,621)}= 160$, $p<.01$. Thus an important factor in rhythmicity is not simply the presence or absence of consistent foot structure, but the particular type of inconsistency. S2 and 4 can be analyzed as alternating, except that the place of the third head is filled by a syllable which is not prominent; S7 and 8 cannot be analyzed as alternating.

Early inconsistency in foot structure (S8) induced less rhythmicity than the late inconsistency in foot structure (S7), but the difference failed to reach statistical significance, $F_{(1,621)}=3.5$, $p<.05$.

The strings with consistent foot structure and head type (S1 and 3) were rated as more rhythmic than the strings with just the consistent head type (S2 and 4), $F_{(1,621)}= 32.7$, $p<.01$, and also as significantly more rhythmic than the strings with just consistent foot structure (S5 and 6), $F_{(1,621)}= 25.3$, $p<.01$. Thus, our hypotheses that rhythmicity depends upon the consistency of foot structure and head type are supported.

Further, it is clear that a head may be defined by pitch accent or stress. While the ratings were higher for the strings

*Table 2. Rhythmicity, prominence and grouping per string. The numbers of boundaries and of prominent syllables are given. (#:string number. M:mean rhythm rating; sd: standard deviation; ma: unstressed; MA: stressed; **MA**: accented; P: number of prominences; B: number of boundaries. Parenthesis shows line of truncation.)*

#	Rhythm M	sd	Pattern
			2a4a6a8a
1	7.7	2.0	ma**MA**ma**MA**ma**MA**ma**MA**(ma
			P:4 66 4 66 4 66 6 63 1
			B:13 45 22 45 22 44 20 11
			2a4a6u8a
2	6.3	1.9	ma**MA**ma**MA**mamama**MA**(ma
			P:5 67 5 66 7 6 12 56 1
			B:12 40 16 47 15 6 22 7
			2s4s6s8s
3	7.2	2.6	maMAmaMAmaMAmaMA(ma
			P:8 28 8 25 8 25 8 24 5
			B:13 22 15 23 15 21 15 7
			2s4s6u8s
4	6.2	2.2	maMAmaMAmamamaMA(ma
			P:12 35 12 30 9 8 7 40 4
			B: 7 19 11 33 13 12 22 2
			2a4a6s8a
5	6.5	1.8	ma**MA**ma**MA**maMAma**MA**(ma
			P:4 64 7 61 9 8 12 49 2
			B:14 36 22 46 21 7 20 5
			2s4s6a8s
6	6.3	2.1	maMAmaMAma**MA**maMA(ma
			P:5 12 4 11 7 58 7 4 1
			B: 6 9 6 20 34 25 9 1
			2a4a7a8a
7	5.0	2.2	ma**MA**ma**MA**mama**MAMA**(ma
			P:4 67 4 63 7 11 60 55 3
			B:10 37 20 43 15 29 21 6
			2a3a6a8a
8	4.4	2.2	ma**MAMA**mama**MA**ma**MA**(ma
			P:4 67 68 8 12 54 20 41 2
			B:13 29 43 17 22 17 23 3
			2s3a5s8a
9	5.1	2.0	maMA**MA**maMAmama**MA**(ma
			P:4 6 62 7 5 7 6 55 3
			B: 1 16 41 19 7 8 19 3

with an accent (S1 and 2) rather than a stress head (S3 and 4), this difference was not statistically significant ($F_{(1,621)}$= 2.2, p<.05).

Prominence Picking

Accented syllables were perceived as prominent more often than were stressed syllables, p<.01. There was no statistically significant difference in the proportion of matches of response to predicted outcome for the accented and unstressed syllables (84% for accents, 90% for unstressed syllables, and 27% for stressed syllables).

The stimulus type accounted for over .08 of the variance in the match of predicted to actual prominence, p<.001. Subjects picked prominences in the strings with either consistent head type or foot structure (S1-8) more in line with our predictions than in the string with variable head type and foot structure (S9), p<.01. There was virtually no difference in the proportion of matches to predicted prominences along the lines that were important in judging rhythmicity, namely between S1-6 and 7-9. Thus rhythmicity and prominence are distinct percepts.

Prominences were picked largely as predicted wherever accents consistently served as foot heads; the worst match of predicted to actual prominence occurred in S6 and 9, where both accent and stress were present. There were fewer matches toward the end of the truncated strings, p<.01.

Grouping

The number of groups perceived differed with stimulus type, $F_{(8,620)}$= 18.6, p<0.05. We had thought that subjects might attempt to analyze strings into feet to increase their rhythmicity. This probably does not occur. In the string with variable foot structure and head type (S9) subjects nominated fewer groups than for other strings, $F_{(8,620)}$= 24, p<.01. In the other 2 strings with low rhythmicity (S7 and 8), subjects nominated many groups, but the ratings were nonetheless low. Also, the contrast between strings which was important in judging rhythmicity (S1-6 vs S7-9) was not important in the grouping data. Thus grouping and rhythmicity judgment are distinct percepts.

The number of groups shows a fair correlation with the number of prominences, r =.51, $F_{(1,626)}$ = 186, p<.01. Subjects placed boundaries consistently next to accents and stresses, generally preferring to end a group with a stress or accent. The traditional view of starting feet with a stress is then not upheld.

Accent induces division into groups more reliably than does stress; strings of feet headed by stresses contain significantly fewer groups than those with feet headed by pitch accents (S3,4 and 6 vs S1,2 and 5: $F_{(1,620)}$= 91, p<.01).

Lastly, longer stimuli contained more boundaries. This could mean either that strings ending in a stress or accent are broken into groups differently when they end in an unstressed syllable, or that listeners tend to break strings of any length into units of a consistent size, in which case one would expect more groups in longer strings.

CONCLUSION

Our expectations about rhythmicity were in general upheld. Subjects distinguished degrees of rhythmicity clearly. High rhythmicity of strings arose in strings with consistent foot structure and head type. Inconsistency of foot structure does not necessarily reduce rhythmicity; strings in which feet contained 1 or 3 unstressed syllables (S2 and 4) were still perceived to be quite rhythmic. Indeed, the type of inconsistency is crucial: feet with 1, 2 or 3 unstressed syllables (S7 and 8) were perceived to be less rhythmic. This difference is nicely predicted by the clock formulation of Povel and Essen [6, 7]. Our results suggest strongly that rhythmicity in speech is a function of the regularity of a unit's recurrence, which is also the basis of their clock formulation. Here regularity means the consistency of foot structure, which is confounded with consistency of duration.

In any case, in judging rhythmicity subjects do not base their decisions wholly on the number of groups perceived or of prominences picked. Prominence picking can specify beats (in Povel and Essen's terms) and grouping can specify meter. Their regularity contributes to rhythmicity. Prominences and groups are picked on similar bases, that is, with respect to predicted accent position.

REFERENCES

[1] Grønnum, N. (1993). Rhythm in regional variants of standard Danish. *ESCA Working Papers, 41*, 20-23.

[2] Campbell, W.N. (1992). *Multi-level Speech Timing Control.* Ph.d. thesis. University of Sussex, Brighton, U.K.

[3] Grover, C. & Terken, J. (1994). Rhythmic constraints in durational control. *Proceedings of the International Conference on Spoken Language Processing (Yokohama, Japan), 1*, 363-366.

[4] Giegerich, H. (1985). *Metrical Phonology and Phonological Structure.* Cambridge U. Press: Cambridge, U.K.

[5] Bruce, G. (1987). On the phonology and phonetics of rhythm: Evidence from Swedish. *Phonologica (1984); Proceedings of the Fifth International Phonology Meeting (Eisenstadt 1984)*, 21-31.

[6] Povel, D-J. (1984). A theoretical framework for rhythm perception. *Psychological Research, 45*, 315-337.

[7] Povel, D-J. & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*, 411-440.