Produced Speech Rhythm Depends On Predictability of Stress Patterns

Hugo Quené[*] and **Robert F. Port**[†]

[*] Utrecht institute of Linguistics OTS, Utrecht University
Trans 10, 3512 JK Utrecht, The Netherlands
hugo.quene@let.uu.nl

[†] Department of Linguistics, Indiana University
port@indiana.edu

ABSTRACT

If speakers repeat a phrase, their speech tends to be highly rhythmical. In a similar task without repetition, no such rhythmicity was found. This study investigates whether stress predictability, varied between tasks, affects speech rhythm. Results show that a regular speech rhythm emerges in repeated phrases (with highly predictable stress pattern), but not in other tasks without repetition (medium and low predictability of stress). For all tasks, the presence or absence of stress shift in a phrase partly depends on its global rhythmical pattern. A clear speech rhythm emerges only if a speaker can predict the upcoming stress pattern, and use this information to optimize the temporal organization of his output speech.

1 INTRODUCTION

Let's assume that you have to read out a phone number that you know by heart. The number contains repeated digit sequences, e.g. *2525250*. Most speakers will produce a rhythmical reading of such a well-rehearsed string of numbers. Prominent syllables are placed at periodically spaced temporal locations, yielding an alternating pattern of stressed and unstressed syllables.

Such a rhythmical pattern is also observed if speakers have to *repeat* a single short phrase with more than one stress, like *bake the bread* [1, 2, 3]. In this so-called "speech cycling" task, speakers tend to place stressed syllables at rhythmic time points, i.e. harmonic and equidistant fractions of the whole-phrase repetition period T, e.g. at $T/2$ or $T/3$ or $2T/3$.

By contrast, if speakers have to speak a similar phrase, but without repeating this phrase, no such rhythmical pattern was observed [4, 5]. Instead, speakers tended to align not the stressed vowel, but the initial vowel, to a target time point marked by a metronome pulse. This a-rhythmicity was presumably caused by the low predictability of the stress pattern. Stimulus phrases with varying stress patterns were listed in random order, and speakers had relatively little time to read and plan ahead for the next stimulus phrase of this list. This resulted in an emergency strategy, with minimal temporal organization within a spoken phrase.

The contrast between these two outcomes is further investigated here, by varying both the stress patterns of the stimulus phrases, and the predictability of the stress pattern among phrases. Both factors are varied within items and within speakers. Stress predictability is varied by having speakers produce the same materials in 3 different tasks, summarized in Table 1. Stimulus materials are similar to those used previously [4, 5]. Each phrase consists of a stress-shiftable number word like *thir.teen*, preceded by a content word with a varying number of unstressed syllables following the stressed syllable; examples are *cement* (wS), *pirate* (Sw), and *cinema* (Sww).

Our main hypothesis is that observed rhythmicity is modulated by the predictability of the stress pattern.

	predictability	stress patterns in list	repeating
1	low [4, 5]	randomized	no
2	medium	blocked	no
3	high [1, 2, 3]	blocked	yes

Table 1: Summary of tasks leading to varying degree of stress predictability for the current stimulus phrase.

2 METHOD

2.1 Materials

Stimuli consisted of two-word phrases like *galaxy thirteen*. The first word is a regular English noun or verb. The stress pattern was varied between phrases, as either wS, Sw, or Sww. For each pattern, 5 monomorphemic content words were selected. The second word

ISBN 1-876346-48-5 © 2003 UAB

in a phrase is an English number word, always with metrical pattern SS (*thirteen, fourteen, fifteen, sixteen.* (Other phrases, with number words having a SwS pattern, e.g. *twentythree*, were also included in the study, but left out of the analyses presented here). There were 5 stimulus phrases (word pairs) in each condition, yielding a total of 15 stimulus phrases.

2.2 Metronome pattern

In tasks 1 and 2, speakers heard a 2-beat metronome pattern consisting of a 4-beat cycle, with a low note on the first beat (400 Hz, 40 ms) and a high note on the second beat (800 Hz, 40 ms) [4, 5]. The third and fourth beats were unmarked. The metronome frequency was fixed at 43 cycles per minute.

In task 3, as in [1, 2, 3], speakers heard a 1-beat metronome with a mid-tone note on the first beat (600 Hz, 40 ms), at 76 cycles per minute (see below).

2.3 Speakers and Procedure

In total, 4 speakers participated in this experiment.

Speakers performed tasks 1, 2 and 3 in this fixed order. For task 1 (no repetition, no blocking) [4, 5], they listened to the metronome for a few cycles, and then fell in with their realizations of stimulus phrases, read from a list of all phrases in randomized order. Speakers were instructed to align the first word with the low metronome tone (1st beat), and the second word with the high tone (2nd beat).

Task 2 (no repetition, with blocking) was similar to task 1, except that the list of stimulus phrases was not in random order, but blocked by stress pattern. Hence, the stress pattern of a phrase was predictable from its preceding items in the list (limited to the 5 items within each block).

Task 3 (with repetition, with blocking) was a "speech cycling" task [1, 3, 2]. The same blocked list of stimulus phrases was used as in task 2. Speakers did not require reading time between tokens, so the metronome frequency was changed to about the double of its frequency in tasks 1 and 2. (Pilot experiments suggested that the intended 86 cycles per minute was uncomfortably fast; this was changed to 76 cycles per minute). Subjects repeated each phrase about 8–10 times, and then skipped several metronome cycles to breathe and to prepare for the next phrase on the list.

2.4 Analysis

Speakers' realizations were analyzed with the same software and procedures as before [5]. After deemphasis filtering with −6dB/octave (to enhance the vowel region of the spectrum), the intensity contour of the filtered speech was used to determine vowel onsets, viz. as the midpoint between the steepest rise in intensity (typically just before the vowel onset) and the peak intensity (typically at maximum vowel amplitude).

The stressed syllable in the number word was defined as the syllable with the highest **peak intensity**. Hence, other acoustic correlates of stress, such as syllable duration, F_0 pattern, and spectral slope [6], were disregarded here, because these parameters would have introduced artefacts. Because the temporal organization was highly constrained by the metronome pattern, speakers had only limited control over syllable *durations*. *Pitch* was disregarded because F_0 proved an unreliable correlate of word stress in this task. Finally, the *spectral slope* was not reliable for stress detection, because there were different vowels in stressed and unstressed syllables. Thus, stress location was determined from the peaks in the intensity contour. Trivial corrections were made by the first author during auditory validation of all measurements; such corrections were necessary in about 5% of all realizations.

The acoustic analysis of speakers' production yields three dependent variables. The first is the incidence of stress shift in the shiftable number word (*thirteen*), expressed as the proportion of initially-stressed realizations of the number word. The second and third are the phases of the initial and final vowels of the number word, relative to the metronome cycle. These are typically at phase angles between 0.2 and 0.8, because the preceding (first) word is aligned to the metronome pulse at phase zero.

3 RESULTS

Table 2 shows the **incidence** of stress shift, i.e. the proportions of realizations with stress shifted to the initial syllable, broken down by the two main factors. These results underline the optional nature of stress shift. Average percentages of stress shift vary between 44% and 95%, with an overall average of 65%. Speakers in all conditions and in all tasks can choose whether or not to shift stress in the critical number word.

task	predictability	preceding stress pattern		
		wS	Sw	Sww
1	low	70%	85%	85%
2	medium	75%	95%	95%
3	high	44%	63%	79%

Table 2: Percentages of stress-shifted tokens of the critical number word, broken down by preceding stress pattern, and by task or stress predictability.

The effects of the main factors (stress pattern, and predictability of stress pattern) on these percentages were investigated using logistic regression [7]. Speakers varied in their overall incidence of stress shift, with speakers averages ranging from 57% to 80%. Hence, logistic regression was done with a mixed-effects model, with speakers as an additional random factor [8]. The optimal model for the incidence data is summarized in

Table 3.

effect	coefficient (s.e.)
average	0.222 (0.176)
preceding wS#...	0.719 (0.417)
preceding Sw#...	**1.537** (0.426)
preceding Sww#...	**2.250** (0.437)
low predictability	0.000 (0.000)
medium predictability	0.670 (0.528)
high predictability	**-0.980** (0.345)

Table 3: Logistic regression coefficients (with standard error), for fixed effects of stress pattern and stress predictability. Significant coefficients are printed in boldface.

These results indicate that there is a significantly higher incidence of stress shift as the number of preceding unstressed syllables increases. Secondly, the difference between low and medium predictability is not significant, and the incidence of stress shift is significantly *lower* for the speech cycling task (3) with high predictability. This unexpected result will be discussed below. Finally, the interaction between the 2 main effects was not significant, and these components were therefore left out of the final model in Table 3.

In addition, the temporal alignment of the vowels in the shiftable number words were investigated, by calculating the **phase angle** of the vowels, relative to the metronome period. From previous research, it was expected that *stressed* vowels are always aligned to the metronome beat [9, 1]. Figures 1 and 2 show the densities of the phase angles of vowels in the number word, for initial and final vowel (upper and lower panel) and for shifted and unshifted stress (black and blue curves). Results for non-repetitive tasks 1 and 2 were highly similar, and these are pooled in Figure 1.

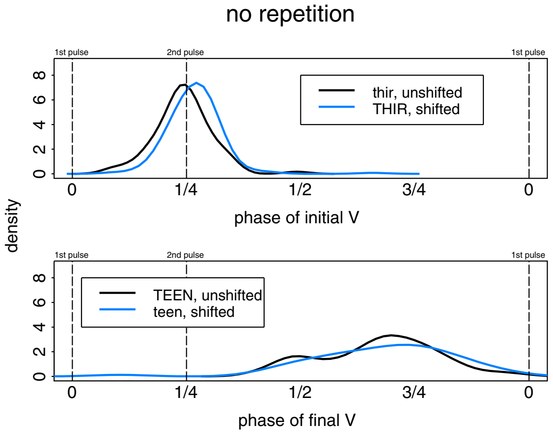

Figure 1: Densities of phase angles of vowel onsets in the shiftable number word, produced with tasks 1 and 2, for initial and final syllable (upper and lower panel) and for realizations without (black) and with (blue) stress shift.

If stresses are aligned to the metronome pulse (in tasks 1 and 2), then the modes (peaks) of these densities should coincide with this preferred location. The phase densities in Figure 1 for tasks 1 and 2 do not conform to this pattern. Instead, the initial vowel (in e.g. *thir*) is *always* aligned to the second metronome pulse, whether this initial syllable is unstressed (unshifted, black) or stressed (shifted, blue). Likewise, the final vowel (in *teen*) is *never* aligned to this metronome pulse, not even when it is stressed (unshifted, black).

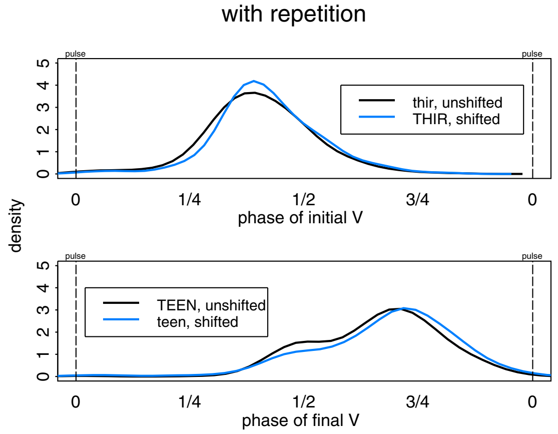

Figure 2: Densities of phase angles of vowel onsets in the shiftable number word, produced with task 3, for initial and final syllable (upper and lower panel) and for realizations without (black) and with (blue) stress shift.

The phase densities of the repetitive task 3 (Figure 2) reveal clearly rhythmical realizations of the shiftable number words. Initial and final vowels are aligned to silent beats at about 1/3 [shifted and unshifted: 0.35] and 2/3 [shifted: 0.67; unshifted: 0.71] of the metronome period, respectively, irrespective of whether these vowels are stressed or unstressed.

4 DISCUSSION

The purpose of this production study is to investigate the effects of stress predictability on the temporal organization of repetitive speech. Highest predictability is found in “speech cycling”. As in previous studies using this task [1, 2, 3], speakers tend to locate the strong syllables at rhythmic time points. Here, each phrase contains 3 metrically strong syllables, viz. 1 in the first word (e.g. *GALaxy*) and 2 in the shiftable number word (e.g. *THIR.TEEN*; both syllables in a word are metrically strong by definition). These 3 strong syllables are placed at equidistant points through the phrase repetition cycle, yielding the distributions in Figure 2.

Since both strong syllables of the number word are aligned at harmonic phase angles, stress assignment among these syllables becomes somewhat irrelevant.

This corresponds with the lower incidence of stress shift observed in the speech cycling task. Within the same temporal organization, speakers could choose either the unshifted or the shifted stress pattern. Hence, speech produced with this task is always highly rhythmical, which renders the actual stress location irrelevant. Speakers sometimes switched back and forth between unshifted and shifted realizations, even between consecutive repetitions of the same phrase.

Nevertheless, stress shift tends to occur more often as the number of preceding unstressed syllables increases. This suggests a tendency for alternation of stressed and unstressed syllables [10].

The other 2 tasks used in this study yielded entirely different results (here discussed together): speakers always attempted to align the initial syllable to the second metronome pulse, even if this syllable was unstressed. Likewise, the final syllable was not aligned to the target pulse, even if this syllable wàs stressed. Only 18/154 (12%) of these stressed final syllables have a phase angle below .50, and none of these is close to the target phase 0.25.

In these 2 non-repetitive tasks, speakers did not tend to locate the stressed vowel at the metronome pulse. Contrary to our prediction, they did not shift the stressed syllable in time to let it coincide with the metronome pulse. This may have been due to the very constraining nature of these tasks, which gave speakers insufficient time and freedom for temporal reorganization. Remember that in these tasks, subjects always had to read the next phrase during the second (silent) half of the metronome cycle, and speak out this item on the next cycle. Blocking items by stress pattern (which was not pointed out explicitly to the speakers) apparently did not decrease the difficulty of this task. In less constrained circumstances, speakers might attempt to align stressed syllables to rhythmic time points, and they might shift stress locations to achieve this goal, but they were unable to do so in this study.

In natural spontaneous speech production, stress predictability is intermediate to the conditions investigated here. Speakers usually do not repeat a single phrase, but yet they do not have to read a new phrase from a paper list either. Because speakers can plan ahead to a moderate extent (compared to the conditions in this study), their spontaneous speech is predicted to be moderately rhythmical in nature.

5 CONCLUSION

First, stress predictability does affect speech rhythm, and hence it provides a plausible explanation for the contrasting outcomes of our previous studies (see Introduction). Regular speech rhythm emerges only if a speaker can predict the upcoming stress pattern, and use this information to optimize the temporal organization of his output speech.

Second, stress shift is controlled in part by rhythmic constraints on the temporal alignment of syllables in real time. These constraints refer to the global rhythmical context of the shiftable target word, and not only to its immediate following context.

ACKNOWLEDGMENTS

The first author gratefully acknowledges support from the Fulbright Visiting Scholar program. We thank Ken de Jong, Adam Leary, and Andrea Juhasz, for valuable comments and assistance.

REFERENCES

[1] F. Cummins and R.F. Port, "Rhythmic constraints on stress timing in English," *Journal of Phonetics*, vol. 26, no. 2, pp. 145–171, 1998.

[2] R.F. Port, K. De Jong, M. Kitahara, D. Collins, A. Leary, and D. Burleson, "Temporal attractors in rhythmic speech," March 2002.

[3] K. Tajima and R.F. Port, "Speech rhythm in English and Japanese," in *Phonetic Interpretation: Papers in Laboratory Phonology VI*, John Local, Richard Ogden, and Rosalind Temple, Eds. Cambridge University Press, Cambridge, to appear.

[4] H. Quené and R. Port, "Stress shift in rhythmical speech," *J. Acoustical Society of America*, vol. 111, no. 5, pp. 2477 (abstract #5aSC12), 2002.

[5] H. Quené and R.F. Port, "Rhythmical factors in stress shift," in *CLS 38: Papers from the 38th Meeting of the Chicago Linguistic Society.*, M. Andronis, E. Debenport, A. Pycha, and K. Yoshimura, Eds., vol. 1: Main Session. Chicago Linguistic Society, Chicago, 2003.

[6] A.M.C. Sluijter, *Phonetic correlates of stress and accent*, Foris, Dordrecht, 1995.

[7] D.W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, Wiley, New York, 2nd edition, 2000.

[8] T. Snijders and R. Bosker, *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*, Sage, London, 1999.

[9] G.D. Allen, "The location of rhythmic stress beats in English: An experimental study. Parts I and II," *Language and Speech*, vol. 15, pp. 72–100 and 179–195, 1972.

[10] B. Hayes, "The phonology of rhythm in English," *Linguistic Inquiry*, vol. 15, no. 1, pp. 33–74, 1984.