Rhythmical Factors in Stress Shift

Hugo Quené Robert F. Port
Indiana University

Abstract
In phrases such as *THIRteen MEN*, stress in the first word is shifted forward from its canonical word-final position. Our general hypothesis is that this stress shift is controlled in part by the *global* rhythmical pattern of the utterance, informally defined as a Rhythmic Constraint on onset times of stressed vowels, in real time. This hypothesis yields two specific predictions about the occurrence of stress shift. First, stress shift should depend on preceding contexts as well as following contexts. Second, stress shift should be more frequent at faster speaking rates. These predictions were tested using 'shiftable' number words, preceded by words that are metrically wS, Sw or Sww. These two-word phrases were read in a list, at various speaking rates controlled by a metronome. The resulting proportions of stress shift verify both predictions. First, stress-shifted realizations are most frequent after Sww words, and less frequent after Sw or wS words. Second, for SS number words, the proportion of stress-shifted realizations increases with faster speaking rate. In addition, for Sww number words, speakers restructure the whole phrase, yielding low incidence of stress shift. These results support the existence of a Rhythmic Constraint that explains why and when stress is shifted.

1 Introduction

In languages such as Dutch and English, one syllable in each lexical item is regarded as the most prominent, carrying 'main lexical stress'. For each word, the location of this main stress is uniquely specified in its canonical sound form, which can be observed in the short answers in (1) (capitals indicate stress). Under some circumstances, however, stress may shift to a different syllable in the word, usually to an earlier syllable, as exemplified in (2).

(1) a. Q: Where does he come from? A: TennesSEE.
b. Q: When does he come? A: In the afterNOON.

(2) a. TENnessee VALley
b. AFternoon TEA

In English, words that are susceptible to stress shift consist of 2 or more syllables, with the main stress located on the final syllable, and with an unreduced preceding or initial syllable. Examples of such 'shiftable' words are *sixteen*, *twentyfive*, *chinese*, *afternoon*, *neolithic*, and *ideal*. Stress shift in such shiftable words is not limited to English and Dutch: similar shifts have also been described for Polish (Hayes and Puppel, 1985), Italian, Greek, Catalan (Nespor and Vogel, 1989), Biblical Hebrew (Churchyard, 1999), and Bedouin Hijazi Arabic (Al-Mozainy, Bley-Vroman and McCarthy, 1985). The same shift has also been observed diachronically in English: in loan words with final stress in the original, stress is shifted to initial position during adaptation of these words to English stress patterns, as in *hurricane, program, address, tyrant, caravan,* and other loan words (Bolinger, 1965a). The same has happened in loan names like *Delaware, Leroy*. The diachronic stress shift is still in progress in loans such as *cigarette, prestige, settee,* where both shifted and unshifted forms are now acceptable.

In the phonological literature, the stress shift in (2) is also known as “Iambic Reversal” or (the outcome of) the “Rhythm Rule”. Various phonological analyses have been proposed, using representations such as the metrical tree (Liberman and Prince, 1977; Kiparsky, 1979), the metrical grid (Selkirk, 1984), or a combination of both (Hayes, 1984; Visch, 1990). In these metrical accounts, stress shift is typically described as a shift of the main lexical prominence, when this prominence is closely followed by a subsequent one. Hence, stress shift is seen primarily as the result of a tendency towards alternation of metrically strong (S) and weak (w) syllables, although this tendency itself is postulated rather than argued.

A second class of phonological analyses describes stress shift in terms of phrase-level accent placement (Bolinger, 1965b; Gussenhoven, 1991; Shattuck-Hufnagel, 1994; Shattuck-Hufnagel, Ostendorf and Ross, 1994; Beckman, 1986). According to these authors, what seems to be ‘stress shift’ actually results from a speaker placing an ‘early’ pitch accent, on a syllable before the canonical stress location in the word. This occurs because speakers have a tendency to place accents at the two extremes, the beginning and the end, of a phrase (Bolinger, 1965b). Thus this analysis predicts that stress shift occurs only when a shiftable word occurs in the beginning of a phrase. The tendency for early accent placement is postulated for rhythmical reasons.

Both families of analyses acknowledge that stress shift is optional. Quantitative evidence for the variability is found in a corpus of American-English radio speech (Shattuck-Hufnagel et al., 1994). This corpus contained 405 relevant word tokens (i.e., polysyllabic words with an unreduced initial syllable, spoken with a pitch accent on this word token). In a clashing context, i.e. when immediately followed by an accented initial syllable in the next word, accent and hence stress location was shifted in 68% (n=93 tokens). In a non-clashing context, such a shift occurred only in 16% (n=312). Secondly, Grabe and Warren (1995) observed in a controlled production study that stress was shifted in 84% of the tokens in a clashing context, and in only 3% of the tokens in a non-clashing context. Both studies demonstrate that stress shift is neither obligatory in clashing contexts, nor forbidden in non-clashing contexts.

Since stress shift is clearly optional, there have been several attempts to identify the factors governing its occurrence. As summarized above, phonologists have postulated rhythmical tendencies such as strong-weak alternation, “eurhythmy”, or “early accent placement”. In our opinion, however, these postulated concepts are poorly motivated and understood. Instead, we propose an *Equal Spacing Constraint* formulated as (3):

(3) **Equal Spacing Constraint**:
Prominent vowel onsets are attracted to periodically spaced temporal locations.

The attractors proposed in (3) have greater influence in some situations than in others. Speakers can produce non-rhythmical speech, with uneven distributions of stresses, accents, or pauses. In contrast to descriptions in terms of eurhythmy or early-accent, the Equal Spacing Constraint is concerned with timing of actual speech events, in real time. Moreover, it is firmly based on models of extra-linguistic, physical dynamical systems such as swings and pendulums. In recent decades, these models have also been applied successfully to bodily movements such as employed in speech (Haken, Kelso and Bunz, 1985; Saltzman and Kelso, 1987; Saltzman and Munhall, 1989; Kelso, 1995; Port and van Gelder, 1995), and to rhythm in speech produced in "speech cycling" experiments (Cummins and Port, 1998; Port, Tajima and Cummins, 1999). We hope to show that these models also provide a partial explanation for stress shift.

According to the Equal Spacing Constraint in (3), the production of stress shift is determined primarily by rhythmical factors during actual speech production[1]. Speakers should have a tendency to place stresses at periodically spaced temporal locations. Let's take a phrase like *their ideal partners*, containing the shiftable word *ideal* (adapted from Grabe and Warren, 1995, diacritics indicate pitch accent). If the preferred temporal location for a stressed vowel coincides with the unstressed initial syllable, as in (4.a), then stress is predicted to shift to that initial syllable. However, if the preferred stress location in *ideal* coincides with the canonical (final) stress location, as in (4.b) where the preceding word *their* is emphasized, then stress shift is unlikely to occur.

(4) a. [John and Ben have been searching for an acceptable partner, but]
they will NÉVer FIND their ÍDeal partners.
b. [Other people have found their perfect partners, but]
John and Ben will NÉVer find THÉIR idEAL partners.

This global rhythmical pattern which partially underlies the Equal Spacing Constraint can be described as resulting from the oscillatory behavior of a dynamical system (Abraham and Shaw, 1982; Cummins and Port, 1998; Port et al., 2002; Large and Jones, 1999). In this case, the system would consist minimally of an oscillator that creates an attractor at each pulse, or phase zero. This dynamical system tends to be stable enough that speakers tend to adhere to the resulting periodic pattern. In this view, stress shift is predicted to arise whenever the stress-shifted variant "fits better" in the global rhythmical pattern. In other words, stress shift reflects the speaker's tendency to maintain the rhythm. Hence, one crucial factor for produced stress shift may be not the stress position in the next word, nor the accentuation of the target word itself, but instead a global rhythmical pattern that underlies the speech behavior.

One prediction that can be derived from this view is that the occurrence of stress shift should be affected equally by the preceding and the following rhythmical pattern. Hence, stress shift could occur even if there is no right-hand context at all, contrary to the metrical descriptions discussed above. Recall that the oscillator model creates attractors for actual speech events in real time. If timing is enforced so that the stress-attracting rhythmical beat falls on the initial syllable of a shiftable word like *fifteen*, then stress shift is more likely to occur, as in (4.a). Similarly, if speakers are entrained to a rhythmical pattern with beats marked as in (5.a), then the shifted realization in (5.b) is more natural and acceptable than its unshifted counterpart in (5.c), because (5.b) matches the rhythmical pattern (5.a) whereas (5.c) does not. However, if the beat happens to fall on the *final* syllable of the shiftable word *fifteen*, as in (6), then stress shift is *not* likely to occur. The shifted realization in (6.b) is *less* natural and acceptable than its unshifted counterpart in (6.c).

(5)	a.		*		*	
	b.		PRO...	blem	FIF...	teen
	c.		PRO...	blem	fif...	TEEN
(6)	a.		*		*	
	b.	pre...	FER	FIF...	teen	
	c.	pre...	FER	fif...	TEEN	

The actual alignment of syllables with rhythmical beats, as in (5) and (6), can be manipulated by varying the metrical structure of the word preceding the critical target word. In addition, it is helpful to impose strict timing constraints with the aid of a metronome which marks the beats indicated in (5.a, 6.a). Without the metronome, speakers might find other solutions to reconcile the Equal Spacing Constraint with the metrical pattern of the words in (5) and (6), e.g. by lengthening vowels, inserting pauses, etc. Although this behavior would be interesting, such speech would not be relevant to our specific predictions about stress shift.

A second prediction of the Equal Spacing Constraint follows from the properties of this kind of dynamical system. The Equal Spacing Constraint should become stronger as speaking rate increases. Stress should be shifted more often at faster speaking rates. This prediction follows from a general property of such a dynamical system: at higher frequencies of the oscillator, the attractor at phase zero becomes stronger (Haken et al., 1985; Kelso, 1995). In musical terms, this explains why up-beats and other notes off the main beat are more difficult to produce at faster rates of playing or singing.

In summary, then, we propose that stress shift should be affected by the global rhythmical pattern of the phrase, which provides attractors in real time for vowel onsets (Allen, 1972; Cummins and Port, 1998). This proposal will be tested in a speech production study, in which the global rhythmical pattern is varied by altering the preceding context, and in which different speaking rates are investigated.

2 Experimental method

This speech production study investigates the incidence of stress shift in two-word test phrases like *cinema fifteen*. The phrases consist of a critical number word with easily shiftable stress, e.g. *fifteen* (Shattuck-Hufnagel et al., 1994), preceded by a 2- or 3-syllable word. Table 1 below shows that 2 factors were varied paradigmatically among these phrases:
(a) the metrical pattern of the number word: either SS (*fifteen*) or wwS (*fiftythree*);
(b) the metrical pattern of the preceding word, with a varying number of unstressed syllables following the stressed syllable; wS (*announce*), Sw (*menu*), Sww (*cinema*). In addition, each phase was spoken at multiple speaking rates enforced by a metronome, beginning at the slowest rate, with stepwise increments in speaking rate between blocks of the experiment.

Table 1: Summary and examples of test phrases, broken down by metrical patterns of shiftable word and preceding word.

preceding word		shiftable number word	
		SS (*fifteen*)	wwS (*fiftythree*)
wS	(*announce*)	*announce fifteen*	*announce fiftythree*
Sw	(*menu*)	*menu fifteen*	*menu fiftythree*
Sww	(*cinema*)	*cinema fifteen*	*cinema fiftythree*

The dependent variable in this study is the incidence of stress shift in the shiftable number word (*fifteen, fiftythree*), as determined from acoustic measurements. This incidence is expressed as the proportion of initially-stressed realizations of the shiftable number word (relative to all realizations).

2.1 Materials

Stimulus materials for this experiment consisted of word pairs, in which the first word is a regular English word, and the second word an English number word. For critical number words with metrical pattern SS, the 4 words *thirteen, fourteen, fifteen, sixteen* were used. For number words with metrical pattern wwS, the 20 words *twentyone, twentythree, twentyfour, twentyfive, twentysix, thirtyone, thirtytwo, ..., sixtythree, sixtyfour* were used. Numbers with “double digits” like *twentytwo* were excluded, because these might be spoken in non-standard rhythmical patterns. In addition, distractors were made using the 5 words *twenty, thirty, forty, fifty, sixty*, with metrical pattern Sw where stress shift is not possible. For each metrical condition of the first word, preceding the critical number word, 20 monomorphemic content words were selected with the help of a dictionary.

The 3 categories of first words (wS, Sw, Sww) were crossed with the 2 categories of critical number words (SS, wwS), yielding 3×2 = 6 cells. There were

20 stimulus phrases (word pairs) in each cell, yielding a total of 120 stimulus phrases. In addition there were 60 distractor phrases.

2.2 Metronome pattern and rates

A metronome was presented to the speakers, to stabilize their rhythmical reading of the list of stimulus pairs. Unlike previous experiments using a similar task (Cummins and Port, 1998; Port et al., 2002), the present experiment does not involve "cycling" or repetition of a single, and hence familiar test phrase. Instead, speakers read a new two-word test phrase on each repetition period of the metronome pattern, with an unpredictable stress pattern for each new test phrase. The metronome pattern was long enough to allow some resting time for the speaker. This pause served three purposes: it gives speakers time to read and plan ahead, and to breathe in if necessary, and it should discourage carry-over of the rhythmical pattern to the next trial.

The optimal metronome pattern (chosen after several pilot trials) consisted 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). The third and fourth beats were unmarked. Speakers were instructed to align the precursor word and the number word to the first and second beat, respectively (see below).

The repetition frequency of the metronome constituted a third factor in this experiment, with values of 28, 33, 38, 43, 48 and 53 cycles per minute (or 0.467, 0.550, 0.633, 0.717, 0.800 and 0.883 Hz) corresponding with periods of 2.143, 1.818, 1.579, 1.395, 1.250 and 1.132 seconds. Note that the actual speaking time is approximately the first half of the metronome period.

2.3 Speakers

In total, 10 speakers participated in this experiment. All speakers were undergraduate or graduate students in the Department of Linguistics at Indiana University, Bloomington. They received a payment of $5 for their voluntary participation.

2.4 Procedure

Speakers wore headsets with a mounted microphone (Shure SM2), with the microphone positioned about 2 cm from the left corner of the mouth. The list of two-word test phrases was presented on paper. For each two-word phrase, the first word was presented in regular lowercase orthography; the second word or number was presented in its regular number form in arabic digits[2].

The metronome signal was generated in real time at 44.1 kHz sampling frequency (using Praat 4.01 software and a laptop-based ESS Solo1 sound card) and fed into an analog audio mixer (Mackie 1202-VLZ). Signals from the metronome and from the speaker's microphone were recorded on two separate tracks of a DAT recorder (Tascam DA-20 Mk.II) at 48 kHz. Both signals were also presented to the speaker's headphones (microphone left, metronome right) at a comfortable hearing level.

Speakers were instructed to align the first word with the low tone of the metronome (at the first beat of the metronome cycle), and to align the second word or number with the high tone (at the second beat of the metronome cycle). It is known that English speakers will align the vowel onset in the stressed syllable with the beat (Allen, 1972; Cummins and Port, 1998), as readers may confirm by repeating the word *append* in synchrony with tapping a finger. The experimenter demonstrated the intended alignment with a filler word pair not used in the experiment (*observe thirtytwo*, always realized with final stress in the second word), repeatedly speaking this single item through several metronome cycles. Without explicit instruction, speakers might have learned from this example that stressed syllables should be aligned with the metronome tones, that unstressed syllables may be spoken on the up-beat preceding the main beat, and that stressed syllables are easiest aligned at their vowel onset.

Speakers listened to the metronome signal for a few cycles, and then fell in with their realizations of the stimulus word pairs. All speakers started with the longest metronome period (or lowest frequency). After reading the whole list of stimulus phrases, speaking each phrase once, there was a short break of 1–2 minutes. Then the metronome frequency was increased to its next higher value, and the speaker read the same list again. Hence, metronome frequency was varied both within speakers and within stimulus phrases. The stepwise increments of metronome frequency continued until a speaker was unable to align stimulus phrases with the initial tones of the metronome pattern.

2.5 Analysis

Data from 1 speaker were partly lost due to equipment failure, so the remaining data from this speaker were excluded from further analysis. In addition, two other speakers were quite unable to align their phrases with the metronome beats. All data from these speakers were also excluded. Thus, data from 7 speakers (2 male, 5 female) were analyzed further.

The two channels of the DAT recordings were re-digitized into stereo sound files at 11025 Hz sampling frequency. In order to determine which syllable in the shiftable number word was stressed, measurements were made from the intensity contour of the speech channel. First, the signal was de-emphasized with –6 dB/octave, to enhance the low-frequency (vowel) region of the speech spectrum. Then the intensity contour of the filtered signal was computed (window length 40 ms, window shift 10 ms). For each token, the stressed syllable in the critical word was defined as the syllable with the highest peak intensity in its vowel.

Our stress classification was based on intensity only, disregarding other acoustic correlates of stress, such as syllable duration, F0 pattern, and spectral slope (Sluijter, 1995; Sluijter and Van Heuven, 1996; Beckman, 1986). These other correlates would have been unreliable for the materials collected here. This is because first, speakers had only limited control over syllable durations: intermediate syllables had to be compressed between two metronome beats, while the final syllable was typically lengthened into the following silent part of the metronome cycle. This makes the produced durations rather artefactual, and difficult to use for stress detection. In fact, informal inspection of the recordings showed that even if stress was unambiguously shifted to the initial syllable, final syllables were often considerably longer than initial ones). Secondly, F0 was difficult to use because speakers typically produced two pitch accents, on the stressed syllables of each word in a phrase. This means that the stressed syllable bore pitch accent. However, F0 was still an unreliable correlate of stress, because accentuation tones were often confounded with boundary tones — thus thwarting automatic extraction of the relevant properties of the F0 contour. Finally, the spectral slope was not reliable for stress detection because there were different vowels in stressed and unstressed syllables.

Thus, for each critical number word either the initial or the final vowel was determined to be stressed, on the basis of the intensity peaks. Trivial corrections were made by the first author during auditory validation of all measurements; such corrections were necessary in about 5% of all realizations.

3 Results

Figure 1 shows the proportions of realizations with initial stress, broken down by metrical patterns of the words constituting the pair. At first glance, these results underline the optional nature of stress shift that was also mentioned above. Average percentages of stress shift vary between 24% and 83%, with an overall average of 45%. Stress shift turns out to be a graded or optional phenomenon, since speakers in all conditions and at all speaking rates can choose whether or not to shift stress in the critical number word.

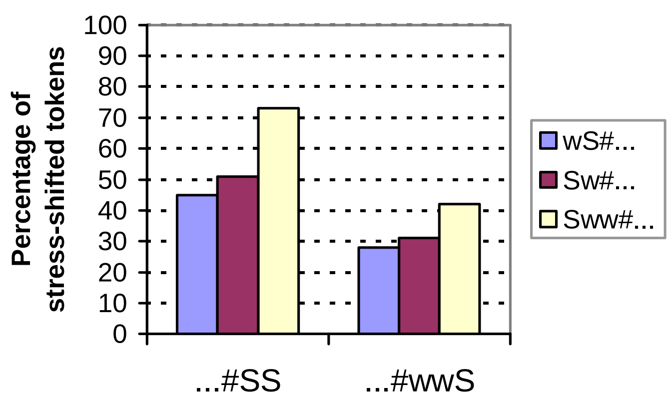

Figure 1: Percentages of stress-shifted tokens of the critical number word, broken down by metrical patterns of the stimulus phrase.

The effects on these percentages of metrical pattern of critical number word, metrical pattern of preceding word, as well as the effect of the metronome frequency, were investigated by means of logistic regression (Kleinbaum, 1992;

Hosmer and Lemeshow, 2000; Pampel, 2000). The dependent variable in this analysis is the *logit* of the percentages summarized in Figure 1. The logit of proportion P is defined as the logarithm of the odds of P, or logit(P) = log(P/(1-P)). This analysis was chosen, because logistic regression allows a combined analysis of discrete and continuous independent variables, as needed here, and because it is perfectly suitable for regression analysis with a dichotomous dependent variable, such as shifted/unshifted stress. The output of this analysis consists of estimated regression coefficients for each level of each factor, and for the linear effect of metronome frequency. Essentially, logistic regression attempts to *model* the logit data, given the independent variables, by estimating the best-fitting regression coefficients for their effects. This analysis therefore extracts more information from the data than classical hypothesis-testing methods like analysis of variance, which only determine whether the effects exceed a fixed significance level. Speakers varied considerably in their incidence of stress shift, with speakers' averages ranging from 18% to 72%. Hence logistic regression was done by means of a mixed-effects model, with speakers as an additional random factor (Goldstein, 1995; Snijders and Bosker, 1999; Pinheiro and Bates, 2000). This type of model is similar to a repeated-measures analysis of variance with the 3 predictors as within-subject factors. The optimal model for the percentage data is summarized[3] in Table 1.

Table 1: Logistic regression coefficients (with standard error), for "fixed" effects of metrical pattern of shiftable number word, metrical pattern of preceding word, and metronome frequency, expressed relative to the baseline condition <u>wS</u>#<u>SS</u>. Significant coefficients are printed in boldface.

condition or predictor	coefficient (s.e.)
baseline: <u>wS</u>#<u>SS</u>	–0.137 (0.285)
<u>Sw</u>#... conditions	+0.211 (0.081)
<u>Sww</u>#... conditions	**+0.893** (0.080)
...#<u>wwS</u> conditions	**–0.984** (0.067)
metronome | ...#<u>SS</u>	**+1.470** (0.365)
metronome | ...#<u>wwS</u>	–0.810 (1.887)

The resulting regression coefficients verify the main prediction of this study: the metrical pattern of the preceding word does indeed affect the incidence of stress shift. For both types of number words, <u>SS</u> *fifteen* and <u>wwS</u> *twentythree*, stress is shifted significantly more often if the number word is preceded by an <u>Sww</u> word (57%) than if preceded by an <u>Sw</u> (41%) or <u>wS</u> (37%) word, as indicated by the positive coefficient in the 3rd row of Table 1. Stress shift is indeed affected by the preceding rhythmical pattern. This is not accounted for by mainstream phonological descriptions of stress shift in English, which only inspect the following context but not the preceding context.

Second, the metrical pattern of the critical number word itself was also expected to affect its susceptibility to stress shift. More stress shift was expected for wwS words (*twentythree*) than for SS words (*thirteen*). This expectation was based on the increased speaking time to the final stressed syllable, combined with an excellent landing place for the main stress, viz. the unreduced initial syllable. The results in Figure 1, however, clearly show that the two types of shiftable number words differ in the *opposite* direction: stress is shifted significantly *less* often in wwS words (34%) than in SS words (56%), as indicated by the negative coefficient in the 4th row of Table 1. This result will be further discussed below.

Third, a main effect was predicted for the metronome frequency, with more frequent stress shift as the metronome frequency, and hence the speaking rate, increases. Figure 2 shows that the effect of metronome frequency on the incidence of stress shift depends on the metrical pattern of the critical number word. This is also reflected in the regression coefficients in Table 1. There is a strong linear effect of metronome frequency for the SS number words (*fifteen*, 5th row, solid lines) but no such effect for the wwS number words (*twentythree*, 6th row, dashed lines).

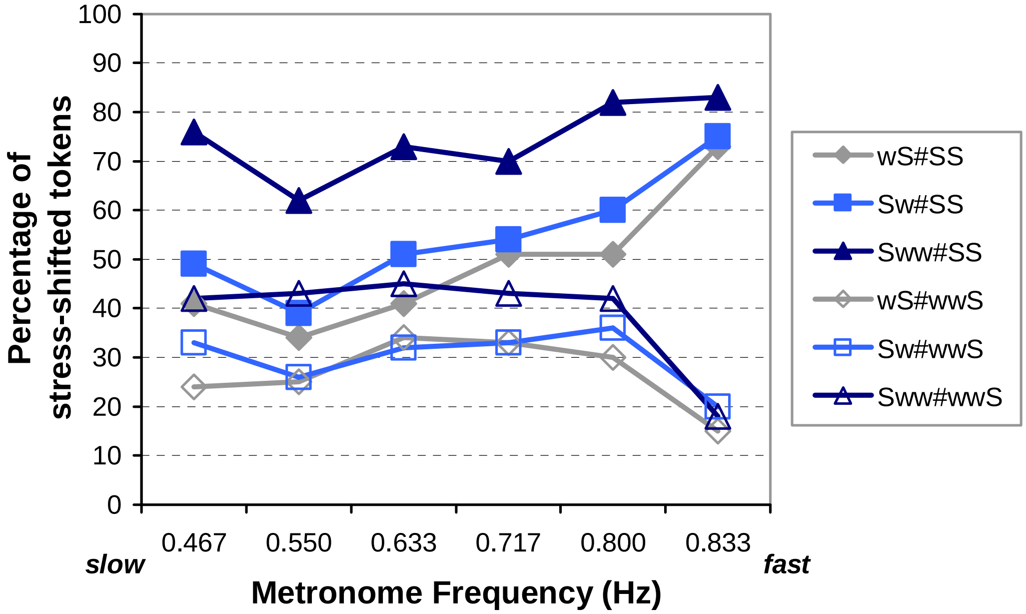

Figure 2: Percentage of stress-shifted tokens of the critical number word, broken down by metrical patterns of the stimulus phrase, and by metronome frequency.

Hence, the predicted effect of metronome frequency (and speaking rate) was confirmed for shiftable SS words like *fifteen*, but not for shiftable wwS words like *fiftythree*. It seems that when speaking phrases with wwS number words, speakers adhered to the Equal Spacing Constraint (3) by other means than stress shift.

4 Discussion

The pattern of results in this study suggests that in some conditions, speakers were indeed guided by an Equal Spacing Constraint as proposed in the Introduction. As predicted, this leads to a higher incidence of stress shift if there are more unstressed syllables immediately preceding the critical word, and if the speaking rate increases. This effect of speaking rate becomes apparent in the SS number words such as *fifteen*.

In the wwS number words, speakers are under stronger time pressure, because they have to produce 3 instead of 2 syllables, with more intrinsic duration for the additional phonetic material. Especially in the Sww#wwS phrases like *cinema twentyfive*, speakers are pushed to their limits. They have to align the two main stresses (wherever these fall) on the metronome pulses at the 1st and 2nd beat (marked with asterisk in 7.a). Meanwhile, they also need to read ahead for the next phrase in the list, and breathe in during the silent metronome interval (3rd and 4th beat), in order to produce that next phrase. Speakers have two options to reconcile these conflicting demands. First, they can pronounce each syllable clearly, respecting their inherent durations, as in unshifted (7.b) or shifted (7.c), where the latter realization is probably easier.

(7)	a.	*1	*2	3	4
	b.	CI . ne . ma	twen . ty . FIVE	(rest)	(rest)
	c.	CI . ne . ma	TWEN . ty . five		(rest)

Alternatively, speakers can solve their timing problems by compressing and reducing the unstressed syllables, thus violating the inherent duration of these syllables. This strategy yields the largest gain if the longest string of unstressed syllables is compressed, as in (8.b).

(8)	a.	*1	*2	3	4
	b.	CI n’ m’ twen . FIVE		(rest)	(rest)

Informal auditory analysis confirms that speakers often reduced and resyllabified stimulus phrases with 3-syllabic wwS number words, as in (8.b). This strategy leads to the observed lower incidence of stress shift for this type of words, as compared to 2-syllable number words. If these wwS words have to be spoken at the fastest speaking rate, then speakers often used this restructuring strategy to escape the strong time pressure, yielding a remarkably low incidence of stress shift (cf. Figure 2).

The present study has shown that stress shift may occur even if there is no “triggering” following context, and that its occurrence can depend on the preceding context. These findings suggest that *produced* stress shift is conditioned, at least in part, by the global rhythmic pattern of an utterance. Interestingly, the global rhythmic pattern seems to be relevant for *perceived* stress shift as well (Grabe and Warren, 1995). If a stress-shifted token of the critical word *ideal* is presented in its full sentence context (9.a), then three trained judges report perceived stress shift in 100% of their responses. If only the preceding context is provided (9.b), stress shift is still perceived in 67% of their judgments, falling to 38% if the critical word is presented alone without any context (9.c).

(9) a. As they will never find their ideal partners,
they must learn to compromise
b. As they will never find their ideal...
c. ...ideal...

Listeners notice stress shift better, if they hear more of the sentence context. In our interpretation, the global rhythmical pattern in the sentence context leads to a "rhythmic expectancy" that may contribute to the perception of stress shift (Grabe and Warren, 1995, p.104; Large and Jones, 1999).

4 Conclusion

Both predictions derived from the Rhythm Constraint were verified in the speech production study. The patterns of stress shift incidence are difficult to explain by means of metrical or intonational phonology. By contrast, the results demonstrate that stress shift is controlled in part by rhythmic constraints on the speech. The proposed rhythmic constraints that explain speakers' behavior are supported by independent evidence of similar rhythmic behavior, from physics and biology. In sum, actual speech rhythm as effected by stress shift cannot be explained as a purely linguistic phenomenon, dependent only on serially ordered phonological objects. Instead, speech rhythm is firmly grounded in our actual speaking behavior in real time.

Acknowledgements

Our thanks are due to Adam Leary, Allard Jongman and Joan Sereno, for fruitful discussions. This work was supported by Utrecht University, The Netherlands, and by the Fulbright Visiting Scholar Program.

Notes

1. The proposed Equal Spacing Constraint works similar to the Aerodynamic Voicing Constraint discussed by John Ohala (this volume), although it affects different properties of speech. Both constraints may be violated. Both constraints are based on extra-linguistic, physical models of aerodynamics and of dynamical systems, respectively.
2. Numbers were printed in digits (e.g. *15*) since pilot work had shown that speakers found this easier to read and to process than spelled-out forms (e.g. *fifteen*). Test phrases were printed in Arial font, boldface, font size 20 points, with double blanks between words in pair, double line spacing, horizontally centered on page.
3. Metronome frequencies (in Hz) were centralized to their mean before being used in the logistic regression. Significant coefficients have a BIC of 2 or more, where Z = coefficient/s.e. and BIC = $Z^2 - \ln(n)$ with n=4188 here (Hosmer and Lemeshow, 2000, p.48; Pampel, 2000, p.31 ff).

References

Abraham, R.A., and Shaw, C.D. 1982. *Dynamics: the Geometry of Behavior*.vol. 1: Periodic Behavior. Santa Cruz, CA: Aerial Press.

Allen, G.D. 1972. The location of rhythmic stress beats in English: An experimental study. Parts I and II. *Language and Speech* 15:72-100 and 179-195.

Al-Mozainy, H.Q., Bley-Vroman, R., and McCarthy, J.J. 1985. Stress shift and metrical structure. *Linguistic Inquiry* 16:135-144.

Beckman, M.E. 1986. *Stress and Non-Stress Accent*. Dordrecht: Foris.

Bolinger, D.L. 1965a. A theory of pitch accent in English. *Forms of English: Accent, Morpheme, Order*, ed. by I. Abe and T. Kanekiyo, 17-55. Cambridge, Mass.: Harvard University Press.

Bolinger, D.L. 1965b. Pitch accent and sentence rhythm. *Forms of English: Accent, Morpheme, Order*, ed. by I. Abe and T. Kanekiyo, 139-180. Cambridge, Mass.: Harvard University Press.

Churchyard, H. 1999. Topics in Tiberian Biblical Hebrew Metrical Phonology and Prosodics, University of Texas: Ph.D. thesis.

Cummins, F., and Port, R. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26:145-171.

Goldstein, H. 1995. *Multilevel Statistical Models*. London: Edward Arnold.

Grabe, E., and Warren, P. 1995. Stress shift: do speakers do it or do listeners hear it? *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, ed. by B. Connell and A. Arvaniti. Cambridge: Cambridge University Press.

Gussenhoven, C. 1991. The English Rhythm Rule as an accent deletion rule. *Phonology* 8:1-35.

Haken, H., Kelso, J.A.S., and Bunz, H. 1985. A theoretical model of phase transitions in human hand movements. *Biological Cybernetics* 51:347-356.

Hayes, B. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15:33-74.

Hayes, B., and Puppel, S. 1985. On the Rhythm Rule in Polish. *Advances in nonlinear phonology*, ed. by H. van der Hulst and N. Smith, 59-81. Dordrecht: Foris.

Hosmer, D.W., and Lemeshow, S. 2000. *Applied Logistic Regression*. New York: Wiley.

Kelso, J.A.S. 1995. *Dynamic Patterns: The Self-Organisation of Brain and Behavior*: MIT Press.

Kiparsky, P. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10:421-441.

Kleinbaum, D.G. 1992. *Logistic Regression: A Self-Learning Text*: Statistics in the Health Sciences. New York: Springer.

Large, E.W., and Jones, M.R. 1999. The dynamics of attending: How people track time-varying events. *Psychological Review* 106:119-159.

Liberman, M., and Prince, A. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8:249-336.

Nespor, M., and Vogel, I. 1989. On clashes and lapses. *Phonology* 6:69-119.

Pampel, F.C. 2000. *Logistic regression: A primer*: Quantitative applications in the social sciences; 07-132. Thousand Oaks, CA: Sage.

Pinheiro, J.C., and Bates, D.M. 2000. *Mixed-effects models in S and S-Plus*: Statistics and Computing. New York: Springer.

Port, R.F., and van Gelder, T. eds. 1995. *Mind as Motion: Explorations in the Dynamics of Cognition*: MIT Press.

Port, R.F., Tajima, K., and Cummins, F. 1999. Speech and rhythmic behavior. *Non-linear Developmental Processes*, ed. by G.J.P. Savelsburgh, H. van der Maas and P.C.L. van Geert. Amsterdam: Royal Dutch Academy of Arts and Sciences.

Port, R.F., De Jong, K., Kitahara, M., Collins, D., Leary, A., and Burleson, D. 2002. Temporal attractors in rhythmic speech. Ms.

Saltzman, E., and Munhall, K.G. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1:333-382.

Saltzman, E.L., and Kelso, J.A.S. 1987. Skilled actions: A task-dynamic approach. *Psychological Review* 94:87-106.

Selkirk, E. 1984. *Phonology and Syntax: The Relation between Sounds and Structure*. Cambridge, MA: MIT Press.

Shattuck-Hufnagel, S. 1994. "Stress shift" as early placement of pitch accent: a comment on Beckman and Edwards. *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, ed. by P.A. Keating, 34-42. Cambridge: Cambridge University Press.

Shattuck-Hufnagel, S., Ostendorf, M., and Ross, K. 1994. Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics* 22:357-388.

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

Sluijter, A.M.C., and Van Heuven, V.J. 1996. Acoustic correlates of linguistic stress and accent in Dutch and American English. Paper presented at *International Conference on Spoken Language Processing (ICSLP)*, Philadelphia.

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

Visch, E. 1990. *A Metrical Theory of Rhythmic Stress Phenomena*: Publications in Language; 34. Dordrecht: Foris.