# A RHYTHM-BASED METRIC FOR TURN-TAKING

**E. Couper-Kuhlen**

**University of Constance, Germany**
**University of Zürich, Switzerland**

## ABSTRACT

A rhythm-based metric for turn-taking is introduced here and its implications for the identification of marked and unmarked transitions in everyday verbal interaction are spelled out. Empirical evidence to support two specific predictions of a rhythm-based metric is presented, showing (i) that for transition times <0.2 sec. overlap, latching, and micro-pausing can be predicted in relation to the tempo of prior talk, and (ii) that for transition times >0.2 sec. a rhythmic analysis makes a better prediction of noticeability for pauses in the 0.2-0.6 sec. range. It will be argued that a metric based on speech rhythm is superior to one based on absolute time because it allows for a more reliable identification of communicatively significant timing.

## 1. INTRODUCTION

Early instrumental studies of spoken English utterances showed isochrony to be present only under ideal conditions [2]. However, the more recent discovery of P-centers [6] has reopened the debate [4]. In order to produce perceptual isochrony between a set of monosyllabic words, it has been shown that it is necessary to advance or retard the acoustic onsets of syllables according to their phonetic make-up. In one case the amount of offset shown to be necessary was 80 ms. or 16% of the duration of the interval involved [6]. This suggests that interstress intervals in connected speech could vary at the worst by as much as 16% with respect to the duration of a prior interval and still be considered isochronous. Presumably, however, the ratio for permissible variation within the bounds of perceptual isochrony increases when the intervals under consideration are polysyllabic and/or contain phrase boundaries. In an auditory/acoustic investigation of spontaneous English speech [1], I have found cases of perceptual isochrony in which the percent difference to a preceding interstress interval is as much as 30%. Perceptual isochrony is, however, rarely if ever found with % differences greater than this.

In everyday English conversation, perceptual isochrony is not constant; there are often syncopated beats and noticeable shifts in tempo. Nor is it continuously present. However, it does tend to become particularly pronounced and clear at the ends of turns in the conversations I have examined. This suggests that it may play a role, perhaps even a facilitative role, in the temporal coordination of turn-taking.

## 2. TRANSITION TIMING: THE UNMARKED CASE

My working hypothesis is that the norm for transition timing in everyday English conversation is temporal coordination between the first beat of the new turn and the last two beats of the prior turn such that an isochronous sequence results. Beats are typically created by syllables with relative prominence at some one level of a prosodic hierarchy constituted by syllables, feet, phonological phrases and intonation phrases. Isochronous patterns arise when every syllable, every unreduced syllable, every (pitch-) accented syllable and/or every

intonational onset or nucleus in a stretch of speech is timed at perceptually regular intervals of time. In contrast to earlier studies on isochrony then, the phenomenon in this approach is not restricted to an uninterrupted succession of stressed syllables only.

Once a rhythmic pulse is established in speech, it creates the expectation that it will continue. Therefore, temporary interruptions, provided they are not too long, can be tolerated without causing the rhythmic gestalt to break down. When a pre-established rhythmic pulse coincides with silence rather than with some prosodically prominent syllable, this is referred to as a silent beat.

Given the above hypothesis, prototypical unmarked timing for turn transition might be represented schematically as in Figure 1, where A and B are different speakers. As this diagram suggests, B must position the first prosodic prominence of the new turn such that it follows the last two prominences of A's turn at an approximately equal interval of time. Any intervening non-prominent syllables (whether post-tonic in A's turn or anacrustic in B's turn) will need to be incorporated into the transition interval in such a way that overall isochrony is not disturbed. If there are numerous non-prominent syllables intervening, this may entail latching or even minimal overlap, depending on the tempo of prior speech. If there are few or no non-prominent syllables to fill the interval, a small 'pause' may be called for, again depending on prior tempo.

## 3. TRANSITION TIMING: A MARKED CASE

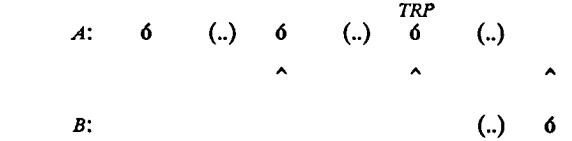
According to the present hypothesis, marked timing in turn transition occurs when the first prosodic prominence of a second turn comes either earlier or later than the next rhythmic pulse following a TRP in the prior turn. Only 'late' timing will be considered here.

A prototypical 'late' second turn can have two variants, represented schematically in Figures 2a/b. In both variants the first rhythmic pulse following the last two beats of A's turn is unfilled, i.e. the next turn is late. But in Figure 2a, when B's turn does begin, its first prosodic prominence coincides with a continuation of the pulse, which retrospectively converts the intervening silence into a silent beat and re-establishes the rhythm. In Figure 2b, by contrast, the first prosodic prominence of the new turn does not coincide with the pre-established pulse and it can be assumed that the gestalt-like rhythm rapidly breaks down. The status of silence following a TRP in the marked transition is thus quite different from its potential status in an unmarked transition. In the marked case it contains or coincides with a rhythmic pulse, whereas in the unmarked case any silence which occurs is incorporated into the rhythmic interval between two pulses.

## 4. EVIDENCE FOR A RHYTHM-BASED METRIC

The rhythm-based metric hypothesized here makes a number of predictions with respect to transition timing which can be investigated empirically. The following will deal with two of these, using two randomly

**Figure 1. Prototypical unmarked timing for turn transition**




[ó represents a prosodically prominent syllable, (..) optional non-prominent syllables, and ^ the rhythmic pulse established by a perceptually regular succession of ós.]

chosen fragments from natural, spontaneous English conversation (one telephone, one face-to-face) as a data base. In both cases prosodic prominences and isochronous rhythms were determined auditorily by two trained specialists, working first independently and later as a team. The intervals between perceptually regular prosodic prominences were then measured acoustically, from vowel onset to vowel onset, using wideband sonagrams as produced by a Kay Elemetrics Digital Sonagraph Model 5500. Only those patterns in which intervals deviated by less than 30% from the duration of a prior interval were retained. In addition to rhythmic intervals, all transition times were measured to the nearest one-hundredth of a second.

**4.1. Prediction of Overlap, Latching or Micro-pause**

From the stipulation that prosodic prominences must be timed regularly in the unmarked case, it follows that the rhythmic interval containing a turn transition may be empty, or filled with a varying number of post-tonic or anacrustic syllables. A rhythm-based metric would predict that for the unmarked case overlap, latching or micro-pause between two turns is a function of how many (and possibly what kinds) of syllables the transition interval contains, in interaction with the rate of the rhythmic pulse (tempo). Where there are few or no non-prominent syllables to be incorporated, a micro-pause is likely, provided the tempo is not fast. Where there are several or many non-prominent syllables, latching or in the extreme case overlap is likely, provided the tempo is not slow. If tempo is held more or less constant, there should be less 'space' between turns the more syllables there are to be incorporated, and vice versa.

When all transition times of 0.2 sec. or less (including cases of minimal overlap between non-prominent syllables) in the conversational fragments examined are now grouped according to the tempo of the rhythmic pulse in surrounding talk, the pattern which emerges supports the prediction that transition times will be shorter where more syllables must be incorporated. In transition intervals of 0.71-0.72 sec., for instance, a 0.2 sec. pause occurs with no incorporated syllables, a micro-pause of 0.1 sec. with 2 incorporated syllables and latching with 4 incorporated syllables. In transition intervals of 0.78 sec. there is a micro-pause of 0.1 sec. with 2

**Figure 2a. Marked 'late' timing for turn transition (rhythm preserved)**

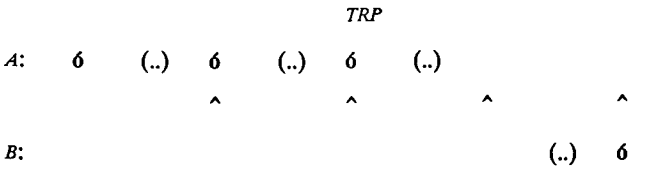
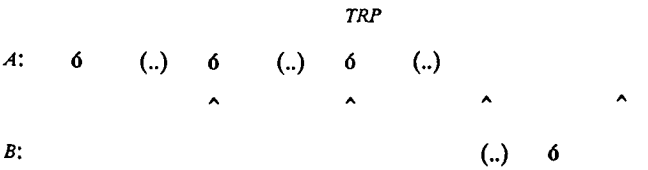



**Figure 2b. Marked 'late' timing for turn transition (rhythm disturbed)**



incorporated syllables but overlap with 6 incorporated syllables.

### 4.2. Identification of Noticeable or Significant Pausing

From a comparison of unmarked vs. marked timing, it also follows that silence occurring between two turns may either fall within a rhythmic interval, or instead coincide with a rhythmic pulse. In the latter case the silence can be expected to be more salient or noticeable, since interlocutors presumably monitor the rhythmic pulse in order to use it as an orientation in timing their entries to the floor. On the other hand, longer pauses are arguably more likely to contain a rhythmic pulse and for this reason may be more noticeable than shorter ones anyway [3].

When all transition times of 0.2 sec. or more in the conversational fragments examined are ranked according to absolute duration, the following pattern emerges. In general, longer pauses are more likely to coincide with rhythmic pulses; pauses longer than 0.7 sec. were never incorporated in the data examined. However, pauses shorter than this were not uniformly incorporated. There were 4 cases of incorporation to 7 of non-incorporation for pauses in the 0.2-0.6 sec. range. Thus, pauses of intermediate duration may or may not contain a rhythmic pulse, in function of the tempo and rhythm of surrounding talk. This suggests that conversationalists may be aware of and attribute signficance to some transitional silences in this range but not to others. An informal test of pause noticeability in the fragments under consideration offers initial support for the prediction that pauses which contain a rhythmic pulse will be more noticeable than those which are incorporated.

## 5. CONCLUSION

Although more evidence must still be brought to bear on these issues, transition timing in the conversational fragments examined here appears to support a rhythm-based metric for turn-taking in English conversation. If fully confirmed, the hypothesis would offer support not only for a prosodic hierarchy such as that advocated e.g. by Nespor/Vogel [7], but also for a non-concatenative, hierarchical model of speech processing [5]. Just as it can be argued that speech segments are not processed sequentially, so interactional timing may be 'processed' not in terms of cumulative duration but with respect to a rhythmic pulse.

## 6. REFERENCES


[1] COUPER-KUHLEN, E. (forthcoming), *"English speech rhythm: Form and function in everyday verbal interaction"*.

[2] CLASSE, A. (1939) *"The rhythm of English prose"*, Oxford: Blackwell.

[3] DUEZ, D. (1985) "Perception of silent pauses in continuous speech", *Language & Speech 28*, 377-389.

[4] FOWLER, C.A. (1979) "'Perceptual centers' in speech production and perception", *Perception and Psychophysics 25*, 375-386.

[5] MARTIN, J.G. (1972) "Rhythmic (hierarchical) versus serial structure in speech and other behavior", *Psychological Review 79*, 487-509.

[6] MORTON, J./MARCUS, S./FRANKISH, C. (1976) "Perceptual centers (P-centers)", *Psychological Review 83*, 405-408.

[7] NESPOR, M./VOGEL, I. (1986) *"Prosodic Phonology"*, Dordrecht: Foris.