

TIME SHRINKING EFFECTS ON SPEECH TEMPO PERCEPTION

Petra Wagner & Andreas Windmann

Faculty of Linguistics and Literature, Bielefeld University, Germany
 petra.wagner@uni-bielefeld.de; andreas.windmann@uni-bielefeld.de

ABSTRACT

Time shrinking denotes the psycho-acoustic phenomenon that an acoustic event is perceived as shorter if it follows an even shorter acoustic event. Previous work has shown that time shrinking can be traced in speech-like phrases and may lead to the impression of a higher speech rate and syllable isochrony.

This paper provides experimental evidence that time shrinking is effective on foot level as well as phrase level. Some examples from natural speech are given, where time shrinking effects may be deliberately employed in poetry and rap music.

Keywords: speech rate, perception, timing, rhythm, isochrony

1. INTRODUCTION

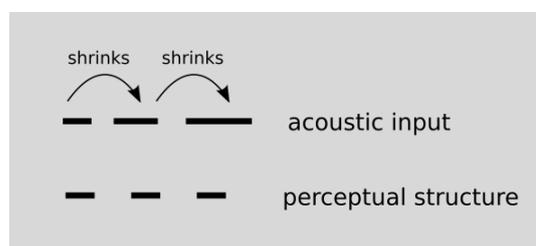
The link between psycho-acoustic findings of rhythmic pattern perception and speech has so far been explored only to a very limited extent, even though there seem to be recurrent ideas of some languages being perceived as more syllable-isochronous [1] or faster [10] than others. [9] suggests a model where speech tempo perception is largely explained by a multiple linear regression combining syllable and phone rate (or information density), a conjecture similarly made for tempo perception in music [2]. [5] showed the influence of phonotactic complexity on perceived isochrony, and [6] emphasized the impact of objective speech rate on the perception of isochrony. However, the interplay between stress patterns, e.g. a language-specific preference for trochaic or iambic feet, and perceived speech rate has so far not been experimentally tested. The time shrinking illusion does provide evidence for a possible influence of such a structural organization – its effect on speech rate perception is examined further in this paper.

2. THE TIME-SHRINKING EFFECT

Time shrinking describes the widely published psychoacoustic illusion that a comparatively long interval is perceived as substantially shorter when following a comparatively short interval [8]. [11]

showed that time shrinking does propagate across intervals, i.e. a “perceptually shrunk” interval may still have a shrinking effect on the next interval in a decelerating series of acoustic events (cf. Fig. 1).

Figure 1: A schematic illustration of the time shrinking effect. A short acoustic event has a shrinking effect on a subsequent longer one. Time shrinking can propagate across several events.



In [12] it was shown that time shrinking is likely to be responsible for the phenomenon that a series of speech-like events (syllable sequences consisting of /ba/) decreasing in tempo is perceived as faster than a series of objectively isochronous syllables or a series of syllables increasing in tempo – despite the series having an identical objective rate. Decelerating series were furthermore perceived as more isochronous. It was speculated that time shrinking may be at least partly responsible for a language like French being perceived as faster and more syllable isochronous than a language like English, with French showing a tendency for decelerating syllable sequences and English showing a tendency for accelerating sequences (across feet or foot-like groups).

Even though the study has shown the general effect of time shrinking on speech-like stimuli, several open questions remain. The syllable-series contained no rhythmic substructure as they consisted of monotonically decelerating or accelerating syllable series such as

- /baba:ba:/ (monotonically decelerating, ending long)
- /ba:baba/ (starting long, accelerating)

On a prosodic level, they can thus be regarded as phrases containing a single foot. Consequently, it could be argued that results presented in [12] apply to prosodic phrase, but not necessarily to

foot level. A likely explanation would then be that listeners simply ignore final lengthening when estimating speech rate. In that case, a sequence starting long would automatically be perceived as slower than one starting short. Given the probably universal nature of final lengthening, this would mean that time shrinking has in fact little influence on speech rate perception.

A simple way of testing the independence of phrase final lengthening and time shrinking is to use an increasing number of feet in the tested stimuli. With more feet, the influence of final lengthening on total utterance duration is automatically weakened.

Furthermore, time shrinking has been shown to propagate across several monotonically decelerating syllables (cf. Fig. 1). Thus, a higher number of syllables per foot ought to strengthen rather than weaken the phenomenon as well.

Another open question is the transferability of the shown phenomenon to real speech rather than nonsense-syllables. Where linguistic information needs to be comprehended by a listener, it is likely that subtle psycho-acoustic effects may be overridden by higher level linguistic processing, e.g. syntactic parsing or the decoding of semantics.

These open questions were addressed in the subsequently described experiment and discussion of rhythm in poetic speech and song. The tested hypotheses are the following ones:

1. Time shrinking is effective on foot level as well as phrase level. It cannot be explained by merely filtering out phrase final lengthening in speech tempo perception. (H1)

2. Time shrinking effects can be traced in natural speech as well. (H2)

3. EXPERIMENT: INVESTIGATING FOOT LEVEL TIME SHRINKING

In the experiment, H1 is tested based on the following reasoning: If the effect hitherto attributed to time shrinking were explicable as a final lengthening phenomenon, its effect ought to diminish given an increasing number of syllables per utterance/foot/phrase. On the contrary, if the described influence of duration order on tempo perception is a result of time shrinking, its influence ought to increase with the number of syllables per foot. We thus expect that in a pair-wise comparison task of patterns leading to time shrinking, the effect should be stable with an increasing number of feet/syllables per utterance.

Also, the effect ought to persist with an increasing number of feet per syllable. If an increase number of syllables per foot/phrase weakens the influence of time shrinking, the alternative hypothesis needs to be accepted. In this case, previous results cannot be attributed to time shrinking.

3.1. Experimental method

In a pair-wise comparison task, nonsense series of syllables consisting of 2-syllabic and 3-syllabic feet were tested. If time shrinking is effective on foot level as well as phrase level, it should be perceptible in phrases consisting of several feet as well as those consisting of one foot only.

In order to explore the effect of time shrinking in speech-like stimuli, various sequences of the syllable /ba/ were recorded, spoken by a female native speaker of German. The tempo of the recorded speech was slow, about 3 syllables per second. This tempo was kept in the subsequent experiment. In order to investigate the presence of time shrinking on various foot patterns, one 2-syllabic foot /baba/ and one 3-syllabic foot /bababa/ was recorded. The recordings were scaled to an average intensity of 60 dB and f_0 was flattened to 169Hz throughout the utterance, corresponding to the speaker's average. Each recording was then modified in duration according to one condition that does not license time shrinking (starting with a long, "stressed" syllable) and one condition that does (ending with a long, "stressed" syllable). The feet licensing time shrinking can be regarded as iambs or anapaests, the others as trochees or dactyls. Modifications were carried out using TD-PSOLA [4]. The 2- and 3-syllabic feet were then concatenated to "phrases" consisting of two or three feet. The binary feet were also combined to sequences of four feet, in order to better test a possible effect of utterance duration. In order to make the patterns sound more natural, the last syllable was added some final lengthening.

This procedure resulted in 10 different test stimuli in total (cf. Tab. 1,2). Duration manipulations were carried out in such a way that the total duration of each stimulus containing the same number of syllables per feet was identical. This results in an identical articulation rate (syllables/sec) for all stimuli containing the same number of syllables and feet. The stimuli were presented in a pair-wise comparison task using the Praat experimental GUI [1]. Stimulus pairs

consisted of stimuli with an identical number of syllables. For each stimulus pair, the subjects had to answer which stimulus was perceived as faster. Each stimulus was presented twice, each time in different order, to factor out any bias to click the left rather than right button (or vice versa). 10 distractors were added in which shrinking and no shrinking conditions were combined in the sequences in order to prevent subjects from learning one particular pattern. A short training phase (3 pseudo-stimuli) preceded the actual experiment to familiarize subjects with the task.

Table 1: Schematic Overview of Stimuli in Various Conditions as Regular Expressions. The feet shown in brackets are repeated up to three times in the case of 2-syllabic feet, and up to twice in the case of 3-syllabic feet.

Condition	2-syllabic feet	3-syllabic feet
No Shrinking	(ba:ba) ⁺ – ba:ba:	(ba:baba) ⁺ – ba:baba:
shrinking	(baba:) ⁺ – baba:	(bababa:) ⁺ – bababa:

Table 2: Stimulus Durations (in Milliseconds); The durations shown in brackets are repeated up to three times in 2-syllabic feet, and up to twice in 3-syllabic feet.

Condition	2-syllabic feet	3-syllabic feet
No Shrinking	(430 250) ⁺ – 430 360	(410 255 255) ⁺ – 410-255-310
Shrinking	(200 500) ⁺ – 200 590	(190 280 450) ⁺ – 190 280 500

3.2. Results and discussion

16 subjects (6 m, 10 f) participated in the test. One female subject was exempted for showing an obvious left-clicking bias. The remaining subjects were aged between 17 and 67, with an average of 34.9. All subjects were native speakers of German. Two subjects reported a slight hearing disorder. 4 subjects had musical training. Results are depicted in Fig. 2. The overall results show a clear tendency to perceive the patterns in the time shrinking condition as faster than the ones in the no shrinking condition ($\chi^2=34.68$; $n=300$; $p<0.0001$). This effect is visible for both 2-syllabic ($\chi^2=17.42$; $n=180$; $p<0.0001$) and 3-syllabic feet ($\chi^2=17.63$; $n=120$; $p<0.0001$). Subjects did not react differently to 2-syllabic or 3-syllabic feet. One pattern (sequences of three binary feet) showed no significant result, though a tendency into the predicted direction.

However, the responses show strong differences between subjects. While 10 subjects clearly perceive the predicted effect, 3 subjects are unable to hear a tempo difference and 2 subjects

even perceive those stimuli as faster were shrinking is suppressed. It is unclear whether this can be attributed to a real perception phenomenon, a misunderstanding of the instructions or to a random search for a pattern. Both subjects who reported a hearing problem are able to perceive the time shrinking phenomenon. Musical training seemed to reduce the impact of time shrinking and enabled listeners to better perceive tempo based on objective durations.

Given that the effect is stable in both longer and shorter sequences, it seems not to be explicable any longer by a mere factoring out of the last syllable due to final lengthening expectations. The differences in perceived tempo seem to be better explained by the specific rhythmic order of the syllables. Since we have not assessed the exact perceived tempo, we do not know whether the magnitude of the effect increases with the number of feet/syllables per phrase (as would be expected).

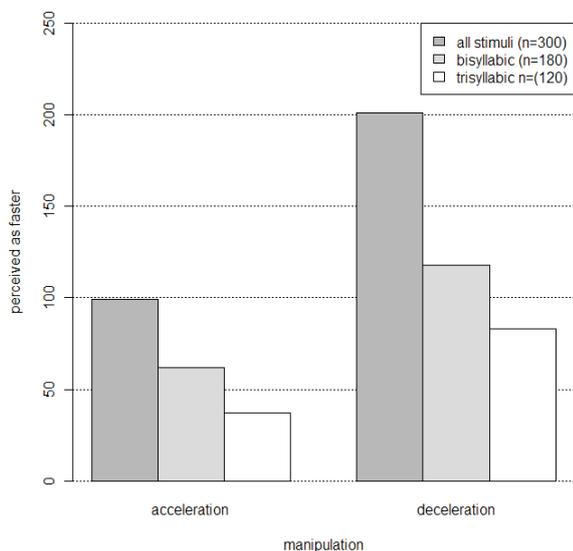
4. TIME SHRINKING PHENOMENA IN NATURAL LANGUAGE

Tracing time shrinking effects in natural speech and testing them cross-linguistically is a difficult task, due to the complex interaction of phone and syllable rate on tempo perception (cf. Introduction). In [12] it was argued that time shrinking may be partly responsible for the often claimed impression of French being more syllable isochronous than English, since French timing patterns do license time shrinking far more often. Further evidence for the impact of time shrinking on speech tempo perception can be gained from music and poetic speech: Scholars of Western poetry claim that decelerating iambs and anapaests create a fast, driving impression, while accelerating trochees and dactyls create a more relaxed atmosphere [3]. Similarly, it is noticeable that modern rap artists prefer iambic or anapaestic meter, often even overriding native stress patterns. Interestingly, the preference for final stress in rap seems to be fairly independent of the artists' native language – if time shrinking is language-independent, it could help rap singers of any language to produce the impression of higher tempo without the need of increasing vocal effort.

Western poetry also reports several instances where a preference of iambic or anapaestic meters may be connected to time shrinking, e.g. Edmund Spenser's iambic pentameter was perceived as a novel, more dynamic style at his time. Classic Greek

literature used anapaestic meter for military marches, due to their dynamic, forward pushing impression.

Figure 2: Overview of Results of the Perception Study. Stimuli in the deceleration conditions that triggers time shrinking were perceived as faster in 2-syllabic and 3-syllabic foot sequences.



While it would be overly simplistic to reduce the usage of poetic or musical styles to a single psycho-acoustic phenomenon, it is conceivable that this phenomenon plays a role in helping artists to convey a particular atmosphere and emotion.

5. CONCLUSIONS

We have demonstrated that time shrinking effects can be traced in sequences containing a higher number of syllables per utterance and persists independently of the number of syllables per foot, and independently of the number of feet per phrase. Thus, time shrinking has an impact on speech tempo perception. H1 is therefore accepted. However, a minority of listeners does not seem to be affected by it. It seems likely that patterns inducing time shrinking are deliberately employed by poets and singers in order to create a vivacious impression. We have thus empirical support for H2.

6. REFERENCES

[1] Boersma, P., Weenink, D. 2011. Praat: doing phonetics by computer. <http://www.praat.org>

[2] Bruhn, H. 2000. Kognitive Aspekte der Entwicklung von Rhythmus. In Müller, K., Aschersleben, G. (eds.), *Rhythmus. Ein interdisziplinäres Handbuch*. Bern: Huber, 227-244.

[3] Burdorf, D. 1997. *Einführung in die Gedichtanalyse*. Stuttgart/Weimar: Metzler.

[4] Charpentier, F., Moulines, E. 1989. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Proc. of Eurospeech* Paris, 2013-2019.

[5] Dauer, R. 1983. Stress timing and syllable timing reanalyzed. *Journal of Phonetics* 11, 51-62.

[6] Dellwo, V. 2010. *Influences of Speech Rate on the Acoustic Correlates of Speech Rhythm*. Doctoral dissertation, Universität Bonn. URN: urn:nbn:de:hbz: 5-20033.

[7] Lehiste, I. 1977. Isochrony reconsidered. *Journal of Phonetics* 5, 253-263.

[8] Nakajima, Y., ten Hoopen, G., van der Wilk, R. 1991. A new illusion of time perception. *Music Perception* 8, 431-448.

[9] Pfitzinger, H.R. 1999. Local speech rate perception in German speech. *Proc. of ICPhS* San Francisco, 893-896.

[10] Roach, P. 1998. *Some languages are spoken more quickly than others*. In Bauer, L., Trudgill, P. (eds.), *Language Myths* Penguin, 150-158.

[11] Sasaki, T., Suetomi, D., Nakajima, Y., ten Hoopen, G. 2002. Time shrinking, its propagation and Gestalt principles. *Perception and Psychophysics* 64(6), 919-931.

[12] Wagner, P., Windmann, A. 2009. No time to lose? Time shrinking effects enhance the impression of rhythmic 'isochrony' and fast speech rate. *Proc. of Interspeech 2009* Brighton, 1523-1526.