

No Time to Lose? Time Shrinking Effects Enhance the Impression of Rhythmic "Isochrony" and Fast Speech Rate

Petra Wagner and Andreas Windmann

Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld, Germany

petra.wagner@uni-bielefeld.de, andreas.windmann@uni-bielefeld.de

Abstract

Time Shrinking denotes the psychoacoustic shrinking effect of a short interval on one or several subsequent longer intervals. Its effectiveness in the domain of speech perception has so far not been examined. Two perception experiments clearly suggest the influence of relative duration patterns triggering time shrinking on the perception of tempo and rhythmical isochrony or rather "evenness". A comparison between the experimental data and duration patterns across various languages suggest a strong influence of time shrinking on the impression of isochrony in speech and perceptual speech rate. Our results thus emphasize the necessity of taking into account relative timing within rhythmical domains such as feet, phrases or narrow rhythm units as a complementary perspective to popular global rhythm variability metrics.

Index Terms: prosody, perception, rhythm, timing, tempo

1. Introduction

Most approaches to speech rhythm classification rely on global acoustic metrics based on the variability of rhythmically relevant phonetic entities such as syllables, feet, narrow rhythm units or subsyllabic entities correlating with phonotactic complexity (e.g. [1, 2, 3, 4, 5]). Many of such metrics have shown to be very useful in the classification of languages into traditional rhythm classes (i.e. stress, syllable and mora timed), but have also helped identifying L2 influences on speech production, distinguishing between rhythmically different dialects or in the differential diagnosis of pathological speech (e.g. [6, 7, 8]). While plenty of evidence exists that listeners can indeed distinguish between the languages' timing properties thus described, it is yet unclear in what way global timing variability is linked to the perception of *rhythmical patterns*, i.e. a series of more or less stressed beats. However, such a notion of rhythm appears to be more intuitive and shows a closer connection to other domains of rhythmical production and perception, e.g. music or dance. Unfortunately, the link between psychoacoustic findings of rhythmical pattern perception and speech has so far been explored only to a very limited extent.

This paper investigates the influence of a psychoacoustic effect known as *time shrinking* in the domain of speech perception. It is shown that time shrinking may be at least partly responsible for the well-known effects of perceptual isochrony [10] in some of the languages called "syllable-timed" but also for the opposite effect of stronger variability or alternation in some of the languages called "stress-timed".

2. The Time Shrinking Effect

The psychoacoustic effect of time shrinking describes the phenomenon that a comparatively long interval is perceived as sub-

stantially shorter when following a comparatively short interval. Sasaki and colleagues [11] could show that this effect is also present in series of three intervals. Another finding was that time shrinking has the power to propagate across intervals, i.e. a "perceptually shrunk" intermediate interval may still have a shrinking effect on the last interval in the series. If the series begins with a long interval, time shrinking is blocked. In the presence of a strictly alternating rhythm, time shrinking is blocked as well. The time shrinking effect could lead to the perceptual impression that a series of stimuli decreasing in time is faster than a series of objectively isochronous stimuli, despite both series having the same average and total duration. It may also lead to the perception of a higher degree of isochrony in an objectively decelerating series of acoustic intervals (cf. Figure 1).

While it is still unclear whether time shrinking is effective in speech perception, several hypotheses can be derived from these results. Given a language preferring decelerating sequences (e.g. *short, longer, very long*), these ought to be perceived as

- faster (H1)
- more even/more isochronous (H2)

compared to a language preferring accelerating sequences (e.g. *long short short*). These two hypotheses are tested in two perception experiments, both of them leading to important implications for the perception of speech rhythm.

3. Experiment 1: Time Shrinking and Tempo Perception

In order to explore the effect of time shrinking in speechlike stimuli, various sequences of the syllable /ba/ were recorded spoken by a female native speaker of German. In order to investigate the presence of time shrinking on variable lengths of interval sequences, one 2-syllabic /baba/, one 3-syllabic /bababa/ and one 4-syllabic /babababa/ sequence was recorded. The recordings were scaled to an average intensity of 60 dB and f0 was flattened to 169Hz throughout the utterance, corresponding to the speaker's average. Each recording was then modified in duration according to three conditions described below. All prosodic modifications were carried out using TD-PSOLA [12]. This procedure resulted in three different versions for each original stimulus and nine stimuli in total. Durations were manipulated in such a way that the total duration of each 2-, 3- or 4-syllabic stimulus had an identical absolute duration. This results in an identical articulation rate (syllables/sec) for all stimuli containing the same number of syllables. The three following manipulation conditions are used (see Table 1 for a description of the duration patterns).

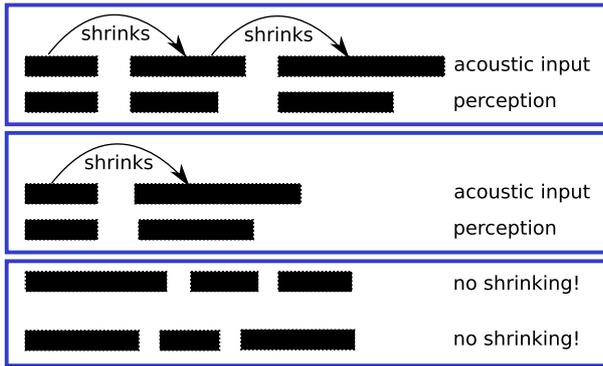


Figure 1: Schematic overview of time shrinking effects. Time shrinking occurs in decelerating sequences and can propagate across several intervals, but is blocked by initial long intervals and in the presence of strict alternation.

- Condition "acceleration": Initial lengthening plus isochrony or alternation
- Condition "deceleration": Strict deceleration, very short initial syllables
- Condition "isochrony": Objective isochrony plus final lengthening

Table 1: Stimulus structure and duration patterns (in milliseconds) for the various manipulation conditions.

Condition	Group 1 2-syllabic	Group 2 3-syllabic	Group 3 4-syllabic
acceleration	/ba:.ba/ 410-250	/ba:.ba.ba/ 410-255-255	/ba:.ba.ba.ba/ 410-230-300-250
deceleration	/ba.ba:/ 190-470	/ba.ba:.ba:/ 190-280-450	/ba.ba:.ba::ba::/ 190-250-300-450
"isochrony"	/ba.ba:/ 230-430	/ba.ba.ba:/ 240-240-440	/ba.ba.ba.ba:/ 250-250-250-440

If the psychoacoustic findings are transferrable to speech-like data, time shrinking ought to be blocked in the acceleration condition either by a sufficiently long initial interval and/or by the presence of an alternation in the 4-syllabic case. However, deceleration ought to lead to a strong time shrinking effect. In the bisyllabic stimulus of the "isochrony-condition", the initial syllable was chosen longer than in the deceleration condition. Else, both conditions would have been identical. This was done in order to see whether a longer initial syllable results in a less strong shrinking effect. For the 3- and 4-syllabic stimuli, time shrinking ought to be less effective in the isochrony condition and only take place between the penultimate and the last syllable in the series. The duration patterns were adjusted in such a way that they provided a sufficient perceptual contrast between the different conditions, that they could be naturally occurring and sounded "speech-like".

4. Experiment 1: Time Shrinking and Tempo Perception

In order to test our H1 that time shrinking leads to a perception of faster tempo in speech-like stimuli, a perception experiment was carried out.

4.1. Method

All stimuli belonging to the same group (based on numbers of syllables) were compared in a pairwise manner and subjects had to state which of the two stimuli appeared as faster. Each group consists of 3 stimulus pairs resulting in 9 stimulus pairs in total. To control for order effects during presentation, each pair was presented in both possible sequences, leading to 18 stimulus pairs. The stimuli were presented via headphones and subjects had to state, whether they perceive the first or the second stimulus as faster by clicking a button on a computer screen. Subjects were allowed to listen to each stimulus pair up to three times before making a decision. In order to find out whether subjects were biased with respect to clicking the left or the right button, pairs of identical stimuli were included in the test. This procedure resulted in 9 additional stimulus pairs. Including one repetition, subjects had to judge 54 stimulus pairs. The experiment started with an initial training phase in order to accustom the subjects with the task. The training phase consisted of 4 stimuli selected out of the entire stimulus pool quasi-randomly (one identical stimulus pair was chosen deliberately) which were also presented twice.

4.2. Results and Discussion

Subjects were 10 native listeners of German (aged between 20 and 64; 5 female, 5 male) without any reported hearing loss. The results clearly indicate an influence of the various stimulus manipulations on the perception of tempo (cf. Figure 2). As expected, stimuli that were manipulated according to the deceleration condition were identified as faster in the vast majority of cases. The isochrony condition was rated faster less often than "pure" deceleration, but stimuli of the acceleration condition were only rated faster in rare cases. Listeners' judgments differed significantly from chance between groups (χ^2 , $p < 0.0001$) and the ranking can be replicated for all three groups of 2-, 3- and 4-syllabic stimuli. There is no significant difference between results within the different groups. This indicates that the effect of time shrinking has an impact on speech-like acoustic signals. It is striking that the isochrony condition lost in almost all cases against the deceleration condition, even though these were fairly similar and subjects often claimed not being able to properly distinguish between the two. An evaluation of the identical stimulus pairs showed a slight bias to press the left button — thus, the control for order of stimulus presentation shows being important. Of course, one might argue that the effect of time shrinking is better explained as an effect of filtering out final lengthening in speech perception tasks. While this may explain the slower perception of the acceleration condition, it does not explain the faster impression of the deceleration over the isochrony condition (with the exception of the bisyllabic stimulus) as is shown in Table 2. We conclude that time shrinking is an effect present in the processing of speech-like stimuli leading to the perception of a higher articulation rate and therefore accept our H1.

Table 2: Comparing the various conditions for the perception of tempo.

compared to	% perceived as faster		
	acceleration	deceleration	isochrony
acceleration	—	90	90
deceleration	10	—	25
"isochrony"	10	75	—

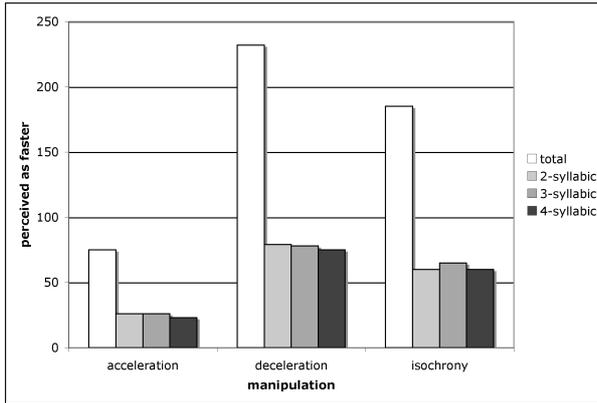


Figure 2: Number of stimuli perceived as faster across the three groups built out of different numbers of syllables.

5. Experiment 2: Time Shrinking and Isochrony

A second experiment was carried out to investigate the impact of time shrinking on the perception of isochrony or rather, "perceptual evenness" (H2).

5.1. Method

The basic procedure was identical to experiment 1. This time, subjects were asked to rate which stimulus in each pair sounded "more even" (German: "ebenmäßig"). Prior to each experiment, the concept of "evenness" was discussed with each subject in detail, because most subjects found it difficult to link an impression of evenness to an auditory stimulus. In fact, several subjects asked whether evenness was supposed to denote "rhythmical regularity" or "isochrony". In this experiment, identical stimulus pairs had been removed from the stimulus pool in order to avoid fatigue. Thus, subjects rated 36 stimulus pairs including 1 repetition.

5.2. Results and Discussion

10 subjects (different from experiment 1) participated in the experiment (native speakers of German, aged between 25 and 40, 6 male, 4 female). The results show a clear difference between the acceleration condition and the other two conditions (χ^2 , $p < 0.0001$) across the groups (cf. Figure 3). Accelerating sequences are clearly perceived as less even or "less isochronous". The isochrony condition seems to be perceived as slightly more even than the deceleration condition. This tendency is significant if the data is pooled across all groups (χ^2 , $p < 0.01$), but not present when individual groups are compared. This finding provides evidence that a perception of "evenness" can indeed be increased by a presence of objective isochrony — however, the order effects (acceleration vs. deceleration) have a much higher impact. No differences can be found within groups with the exception of the 2-syllabic stimuli: The "less even" impression of acceleration is less clear (χ^2 , $p < 0.05$) for the 2-syllabic stimuli. Maybe not surprisingly, subjects had difficulties applying the concept of "evenness" on 2-syllabic "series".

When comparing the results between the different manipulation conditions, both isochrony and deceleration are clearly perceived as more even than the acceleration condition, while isochrony is perceived as more even than deceleration in the

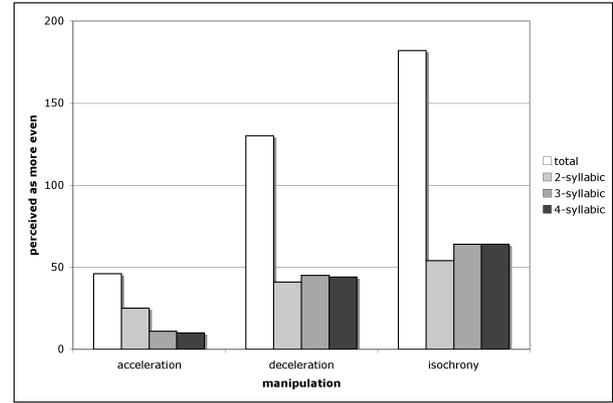


Figure 3: Number of stimuli perceived as "more even" across the three groups built out of different numbers of syllables.

vast majority of cases. This is a reversed effect of experiment 1, where deceleration was clearly perceived as faster than isochrony. From these results we can deduce that the impressions of tempo and perceptual isochrony are somewhat related. However, deceleration is a stronger trigger of tempo increase (due to time shrinking), while isochrony is a stronger trigger of evenness (time shrinking is less effective). Acceleration seems to block the perception of evenness, even in the presence of a series of objective isochrony, as it is the case for the 3-syllabic stimulus. Based on these results, H2 is partly accepted, since it is unlikely that time shrinking is the only effect creating an impression of evenness. Objective isochrony, when presented in a configuration also triggering time shrinking, provides the most reliable impression of evenness.

Table 3: Comparing the various conditions for the perception of evenness.

compared to	% perceived as more even		
	acceleration	deceleration	isochrony
acceleration	—	73	87
deceleration	27	—	63
isochrony	13	37	—

6. Discussion: Implications for Natural Speech Rhythm

The potential impact time shrinking effects on the perception of speech rhythm are obvious. If languages create rhythm groups in such a way as to prefer series of decelerating over accelerating ones, this ought to lead to different auditory impressions. While a tendency of deceleration seems to be a universal phenomenon across languages on phrase level, its impact is less understood on lower levels of rhythmic hierarchical organization, i.e. foot level. For Germanic "stress timed" languages, both decelerating and accelerating feet are possible in poetry (e.g. iambic and trochaic), but accelerating feet are assumed to be the standard cases. If this is true, time shrinking could provide a partial explanation for Germanic languages sounding slower and less isochronous than so-called syllable-timed languages. For French, a final lengthening on foot level is often assumed but difficult to trace in empirical data. However,

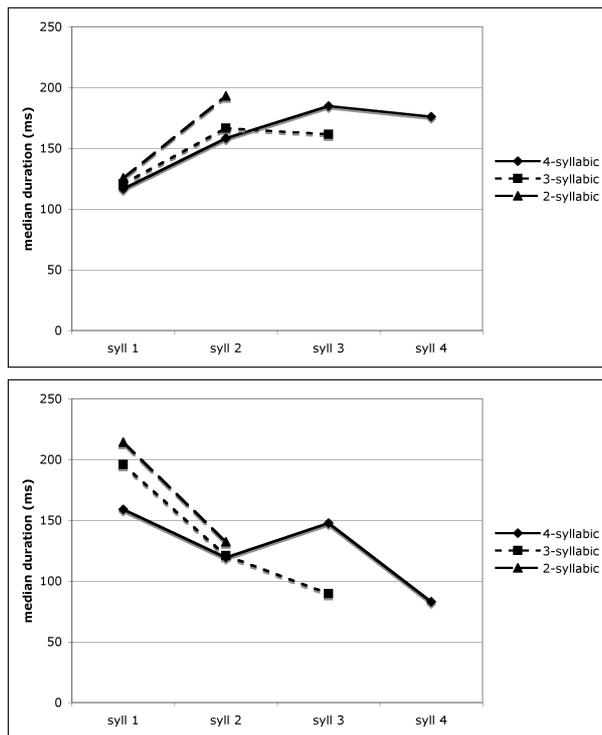


Figure 4: *Deceleration tendencies in French (top) compared to acceleration in English (low) in 2-, 3- and 4-syllabic feet (median durations).*

when counting the final syllable of content words to be foot final, a deceleration effect can be detected in French sequences (data taken from the BonnTempo corpus [13], cf. Figure 4). More importantly, the foot-initial syllable tends to be very short and may thus trigger time shrinking as well. This may be at least partly responsible for the perceptual impression of French sounding "isochronous" and fast. The link between tempo and rhythm has recently been found to be crucial for a language rhythm typology [14]. It is possible, that time shrinking plays an important role in connection with this finding.

Of course, the question need to be asked how many conclusions one can draw from nonsense stimuli for the perception of natural speech. Also, it ought to be kept in mind that the stimuli used in this experiment were presented to the subjects in isolation, thus, they were most certainly interpreted as phrases. It remains to be shown that the effect remains active on foot level in natural speech.

7. Conclusions

We could show that duration patterns within metrical groups have a strong impact on the perception of isochrony and tempo. These effects could to a large extent be explained with the help of time shrinking. It is striking, that both the impression of isochrony and tempo can be linked to a distinction between typological rhythm classes as well. Furthermore, patterns triggering time shrinking can be traced in languages often labelled as perceptually isochronous and fast (i.e. "syllable timed"). While it may be an oversimplification to identify time shrinking as the major influence on perceptual speech rhythm, its effect has hitherto been completely overlooked. In the design of objective

rhythm metrics, well-established psychoacoustic effects need to be taken carefully into account. In the future, perception experiments with native listeners of languages other than ("stress timed") German ought to be carried out in order to determine the universality or language dependency of the effect.

8. Acknowledgements

A thousand thanks go to Volker Dellwo for pointing out the weird relationship between rhythm and tempo — and for the many hours he spent building the BonnTempo database. Thanks also to the audiences at the recent rhythm workshops in London, Paris and Osnabrück where some of the initial ideas for this work were discussed. Above all, we would like to thank everybody who voluntarily participated in our boring perception experiments.

9. References

- [1] Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3):265–292.
- [2] Low, E. L., Grabe, E., and Nolan, F. (1999). A contrastive study of prosody and lexical stress placement in Singapore English and British English. *Language and Speech*, 42(1):39–56.
- [3] Dellwo, V. (2003). Rhythm and speech rate: A variation coefficient for ΔC . In *Proceedings of the 38th Linguistic Colloquium*, Budapest, Hungary.
- [4] Jassem, W., Hill, D., and Witten, D. (1984). Isochrony in English speech: Its statistical validity and linguistic relevance. In Gibbon, D. and Richter, H., editors, *Intonation, Accent and Rhythm. Studies in Discourse Phonology*, pages 203–225. Walter de Gruyter, Berlin, New York.
- [5] Asu, E. and Nolan, F. (2006). Estonian and English rhythm: A two-dimensional quantification based on syllables and feet. In *Proceedings of Speech Prosody*. OS1-5.0229.
- [6] Bond, Z., Markus, D., and Stockmal, V. (2003). Prosodic and rhythmic patterns produced by native and non-native speakers of a quantity language. In *Proceedings of the XVth International Congress of the Phonetic Sciences*, 527–530.
- [7] White, L. and Mattys, S. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35(4):501–522.
- [8] Liss, J., White, L., Mattys, S., Landsford, K., Lotto, A., Spitzner, S., and Caviness, J. (to appear). Quantifying Speech rhythm abnormalities in the dysarthrias. *Journal of Speech, Language and Hearing Research*.
- [9] Hay, J. and Diehl, R. (1999). Effect of duration, intensity and f0 alternations on rhythmical grouping. *Proceedings of the XIVth International Congress of the Phonetic Sciences*, 245–248.
- [10] Pike, K. (1945). *The Intonation of American English*. University of Michigan Press, Ann Arbor.
- [11] Sasaki, T., Suetomi, D., Nakajima, Y., and ten Hoopen, G. (2002). Time-shrinking, its propagation, and gestalt principles. *Perception and Psychophysics*, 64(6):919–931.
- [12] Charpentier, F. and Moulines, E. (1989). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. In *Proceedings of EUROPEECH*, 13–19, Paris: ENTS.
- [13] Dellwo, V., Aschenberger, B., Wagner, P., Dankovicova, J., and Steiner, I. (2004). BonnTempo-corpus and BonnTempo-tools: A database for the study of speech rhythm and rate. In *Proceedings INTERSPEECH 2004*, pages 777–780, Jeju Island, Korea.
- [14] Dellwo, V. (to appear). *The influence of speech rate on speech rhythm*. PhD thesis, Universität Bonn.