Jian, H. (2004) "On the syllable timing in Taiwan English", in B. Bel / I. Marlien (eds.) *Proceedings of Speech Prosody 2004*, Nara (Japan) 23-26 March 2004, ISCA Special Interest Group on Speech Prosody (SproSIG), 247-250.

Keller, E. (ms.) "From sound to rhythm expectancy".

Keller, E. / Port, R. (2007) "Speech timing: Approaches to speech rhythm", in J. Trouvain / W. J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken 6-10 August 2007, Saarbrücken: Universität des Saarlandes, 327-329.

Krull, D. / Engstrand, O. (2003) "Speech rhythm – intention or consequence? Crosslanguage observations on the hyper/hypo dimension", *Phonum* 9: 133-136, http://www.ling.su.se/fon/perilus/2003_10.pdf (last accessed 30/06/2009).

Lee, C.S. / McAngus Todd, N. (2004) "Towards an auditory account of speech rhythm: application of a model of the auditory 'primal sketch' to two multi-language corpora", *Cognition* 93/3: 225-254.

Lenneberg, E. (1967) *Biological foundations of language,* New York: Wiley.

Liberman M. / Prince A. (1977) "On Stress and Linguistic Rhythm", *Linguistic Inquiry* 8: 249-336 (re-printed in Ch.W. Kreidler ed., *Phonology: Critical concepts*, London-New York: Routledge, 2001, 152-244).

Lindblom, Bj. / Rapp, K. (1973) "Some temporal regularities of spoken Swedish", *Papers in Linguistics from the University of Stockholm* 21: 1-59.

Lloyd James, A. (1940) *Speech signal in telephony,* London: Pitman & Sons.

MacNeilage, P.F. / Davis, B.L. (1990) "Acquisition of speech production: Frames, then content", in M. Jeannerod (ed.) *Attention and Performance XIII - Motor Representation and Control*, Hillsdale: N.J.: LEA, 452-468.

MacNeilage, P.F. (1998) "The frame/content theory of evolution of speech production", *Behavioral and Brain Sciences* 21: 499-546.

Mairano, P. / Romano, A. (2007a) "Lingue isosillabiche e isoaccentuali: misurazioni strumentali su campioni di italiano, francese, inglese e tedesco", in V. Giordani / V. Bruseghini / P. Cosi (eds.) *Scienze vocali e del linguaggio. Metodologie di Valutazione e Risorse Linguistiche*, *Atti de 3° Convegno nazionale AISV (Associazione Italiana di Scienze della Voce*) Povo di Trento 29 November- 1 December, Torriana (RN): EDK, 119-134.

Mairano, P. / Romano, A. (2007b) "Inter-Subject Agreement in Rhythm Evaluation for Four Languages (English, French, German, Italian)", in J. Trouvain / W. J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken 6-10 August 2007, Saarbrücken: Universität des Saarlandes, 1149-1152.

Mairano, P. / Romano, A. (2008a) "A comparison of four rhythm metrics for six languages", Poster presented at the workshop *Empirical Approaches to Speech Rhythm (EASR)* (University College London, 2008).

Mairano, P. / Romano, A. (2008b) "Distances rythmiques entre variétés romanes", in A. Turculeţ (ed.), *La variation diatopique de l'intonation dans le domaine roumain et roman*, Proceedings of the International Symposium, Iaşi 20-22 October, Romania, 2008, Iaşi, Editura Universitatii *Alexandru Ioan Cuza*, 251-272.

Mairano, P. / Romano, A. (2010) "Un confronto tra diverse metriche ritmiche usando Correlatore", in S. Schmid / M. Schwarzenbach / D. Studer (eds.) *La dimensione italiana del parlato*, Proceedings of the *5° Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce* (Zurich 4-6 February 2009), Torriana (RN): EDK, 79-100.

Marotta, G. (1985) *Modelli e misure ritmiche: la durata vocalica in italiano,* Bologna: Zanichelli.

Martinéz Celdrán, E. / Fernández Planas, A.M. / Carrera Sabaté, J. (2003) "Castilian Spanish", *Journal of International Phonetic Association* 33/2: 255-259 (sound samples available on-line for the *IPA* members: http://web.uvic.ca/ling/resources/ipa/members).

Mehler, J. / Dommergues, J. / Frauenfelder, U. / Segui, J. (1981) "The syllable's role in speech segmentation", *Journal of Verbal Learning and Verbal Behavior* 20: 298-305.

Mendicino, A. / Romito, L. (1991) "*Isocronia* e *base di articolazione*: uno studio su alcune varietà meridionali", *Quaderni del Dipartimento di Linguistica dell'Università della Calabria*, Serie Linguistica 3: 49-67.

Miller, M. (1984) "On the perception of rhythm", *Journal of Phonetics* 12: 75-83.

Mok, P.P.K. / Dellwo, V. (2008) "Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English", in P. A. Barbosa / S. Madureira / C. Reis (eds.) *Proceedings of Speech Prosody* 2008 (4th Conference), Campinas 6-9 May 2008, Campinas (Brazil), São Paulo: Capes, Fapesp, CNPq - Campinas: IEL, Unicamp, Motorola, Editora RG, cdrom, 423-426.

Molinu, L. / Romano, A. (1999) "La syllabe dans un parler roman de l'Italie du Sud (variété salentine de Parabita – Lecce)", Proceedings of the *2èmes Journées d'Études Linguistiques Syllabes*, Nantes 25-27 March, Nantes: Acoustique, acquisition, Interprétation (A.A.I.), 148-153.

Nazzi, T. / Bertoncini, J. / Mehler, J. (1998) "Language Discrimination by Newborns: towards an understanding of the role of rhythm", *Journal of Experimental Psychology: Human Perception and Performance* 24.3: 756-766.

Nespor, M. (1993) *Fonologia* Bologna: Il Mulino.

Nooteboom, S. (1997) "The prosody of speech: melody and rhythm", in W.J. Hardcastle / J. Laver (eds.) *The Handbook of Phonetic Sciences*, Oxford: Blackwell, 640-673.

O'Dell, M. / Nieminen, T. (1999) "Coupled oscillator model of speech rhythm", in J. Ohala / Y. Hasegawa / M. Ohala / D. Granville and A. Bailey (eds.) *Proceedings 14th International Congress of Phonetic Sciences*, Berkeley University, vol. 2: 1075-1078.

Pamies Bertrán, A. (1999) "Prosodic Typology: On the Dichotomy between *Stress-Timed* and *Syllable-Timed* Languages", *Language Design* 2: 103-130.

Pike, K.L. (1945) *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Ramus, F. (2002) "Acoustic Correlates of Linguistic Rhythm: Perspectives", in B. Bel / I. Marlien (ed.) *Proceedings of the 1st International Conference on Speech Prosody*, Aix-en-Provence 11-13 aprile 2002, Aix-en-Provence: Laboratoire Parole et Langage, 115-120.

Ramus, F. / Nespor, M. / Mehler, J. (1999) "Correlates of linguistic rhythm in the speech signal", *Cognition* 73/3: 265-292.

Rhardisse, N. / Abry, C. (1995) "Mandible as syllable organizer", in K. Elenius / P. Branderud (eds.) *Proceedings of the 13th International Congress of Phonetic Sciences* (Stockholm, Sweden), Stockolm: KTH and Stockolm University 3, 556-559.

Roach, P. (1982) "On the distinction between 'stress-timed' and 'syllable-timed' languages", in D. Crystal (ed.) *Linguistic controversies*, London: Edward Arnold, 73-79.

Roach, P. (ed.) (2003) *A Bibliography of Timing and Rhythm in Speech* (University of Reading, Last updated 2nd April 2003), http://www.personal.rdg.ac.uk/~llsroach/timing.pdf (last accessed 30/06/2009).

Roach, P. (2004) "British English: Received Pronunciation", *Journal of the International Phonetic Association* 34/2: 239-245 (sound samples available on-line for the *IPA* members: http://web.uvic.ca/ling/resources/ipa/members).

Romano, A. (1999) *Analyse des structures prosodiques des dialectes et de l'italien régional parlés dans le Salento: approche linguistique et instrumentale*. *Thèse de Doctorat de l'Université Stendhal de Grenoble* (partially published in 2001, Lille: Presses Univ. du Septentrion).

Romano, A. (2003) "Accento e intonazione in un'area di transizione del Salento centro-meridionale", in P. Radici Colace / G. Falcone / A. Zumbo (eds.) *Storia politica e storia linguistica dell'Italia meridionale, Atti del convegno internazionale di studi parlangeliani*, Messina 22-23 May 2000, Messina-Napoli: Edizioni Scientifiche Italiane, 169-181.

Romano, A. (2010) "Speech Rhythm and Timing: Structural Properties and Acoustic Correlates", in S. Schmid / M. Schwarzenbach / D. Studer (eds.) *La dimensione italiana del parlato*, Proceedings of the *5° Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce* (Zurich 4-6 February 2009), Torriana (RN): EDK, 45-75.

Romano, A. / Mairano, P. / Pollifrone, B. (2010) "Variabilità ritmica di varietà dialettali del Piemonte", in S. Schmid / M. Schwarzenbach / D. Studer (eds.) *La dimensione italiana del parlato*, Proceedings of the *5° Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce* (Zurich 4-6 February 2009), Torriana (RN): EDK, 101-112.

Romito, L. / Trumper, J. (1993) "Problemi teorici e sperimentali posti dall'isocronia", *Quaderni del Dipartimento di Linguistica dell'Università della Calabria*, *Serie Linguistica* 4: 89-118.

Russo, M. / Barry, W.J. (2008a) "Measuring rhythm. A quantified analysis of Southern Italien Dialects Stress Time Parameters", in A. Pamies / M.C. Amorós / J.M. Pazos (eds.) *Experimental Prosody*, *Language Design*, special issue 2: 315-322.

Russo, M. / Barry, W.J. (2008b) "Isochrony reconsidered. Objectifying relations between Rhythm Measures and Speech Tempo", in P.A. Barbosa / S. Madureira / C. Reis (eds.) *Proceedings of Speech Prosody* 2008, 4th Conference on Speech Prosody, Campinas 6-9 May 2008, Campinas (Brazil), São Paulo: Capes, Fapesp, CNPq - Campinas: IEL, Unicamp, Motorola, Editora RG, cdrom, 419-422.

Sachs, C. (1953) *Rhythm and tempo: a study in music history,* London: Dent.

Schmid, S. (1997) "A typological view of syllable structure in some Italian dialects", in P.M. Bertinetto / L. Gaeta / G. Jetchev / D. Michaels (eds.), *Certamen Phonologicum III*, Torino: Rosenberg & Sellier, 247-265.

Schmid, S. (2001) "Un nouveau fondement phonétique pour la typolo-

gie rythmique des langues", in *Écrits pour le 10ème anniversaire du laboratoire d'analyse informatique de la parole (LAIP)*, Lausanne (ms).
Schmid, S. (2004) "Une approche phonétique de l'isochronie dans quelques dialectes italo-romans" in T. Meisenburg / M. Selig (eds.) *Nouveaux départs en phonologie*, Tübingen: Narr, 109-124.
Schmid, S. (2008) "Measuring the rhythm of Italian dialects", Poster presented to the workshop *Empirical Approaches to Speech Rhythm (EASR)* (University College London 2008).
Schmid, S. (ms.) "L'ipotesi delle classi ritmiche e i dialetti italo-romanzi".
Scott, D. / Isard St.D. / de Boysson-Bardies, B. (1986) "On the measurement of rhythmic irregularity: a reply to Benguerel", *Journal of Phonetics* 14: 327-330.
Sock, R. / Löfqvist, A. / Perrier, P. (1996) "Kinematic and acoustic correlates of quantity in Swedish and wolof: a cross-language study", ESCA (editor) Proceedings of the 1st ETRW (ESCA Tutorial and Research Workshop) on Speech Production Modeling: *From Control Strategies to Acoustics*, Autrans (France), 20-24 May 1996, Grenoble: ICP (Institut de la Communication Parlée), 81-84.
Vayra, M. / Avesani, C. / Fowler, C. (1984) "Patterns of temporal compression in spoken italian", in M. van den Broecke / A. Cohen (eds.) *Proceedings of the 10th International Congress of Phonetic Sciences* (Utrecht, The Netherlands, 1983), Dordrecht: Foris, vol. 2, 541-546.
Volín, J. (2005) "Rhythmical properties of polysyllabic words in British and Czech English", in J. Cermák / A. Klégr / M. Malá / P. Šaldová (eds.), *Patterns*, Praha: Kruh moderních filologu, 279-292.
White, L. / Mattys, S.L. / Series, L. / Gage, S. (2007) "Rhythm Metrics Predict Rhythmic Discrimination", in J. Trouvain / W. J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken 6-10 August 2007, Saarbrücken: Universität des Saarlandes, 1009-1012.

A Time-delay Approach to Speech Rhythm Visualization, Modeling and Measurement

Petra Wagner
Universität Bielefeld Fakultät für Linguistik und Literaturwissenschaft

ABSTRACT: The main objective of this paper is the definition of psycho-acoustically confirmed acoustic correlates of rhythmic structure across various languages, accents and speaking styles. Within this quest, *speech rhythm is regarded as being characterized by grouped sequences of beats, which are characterized by their prominence structure*. Duration is identified as the major acoustic correlate of speech rhythm organization: it signals both beginnings and ends of rhythmic groups at different hierarchical levels of rhythmic organization and is the most robust cue to perceptual prominence. In order to visualize the impact of relative durations across rhythmically salient beat transitions, e.g. at phrase boundaries or stresses, time-delay plots are introduced. The method is evaluated quantitatively both statistically and using a KNN classifier. The visualization technique reveals different relative timing patterns for French and English, which provide prototypical cases for the classic distinction between stress and syllable timing. An analysis of relative timing across several languages shows that the traditional classification into stress and syllable timing falls way too short. An appropriate quantitative model of speech rhythm must be multidimensional and take into account psycho-acoustic facts.

1. Introduction

During the last decade, phonetic rhythm research has concentrated on the development of rhythm metrics based on typological findings discriminating between so-called stress timed and syllable timed languages (for a critical review see Arvaniti 2009 or Barry *et al.* 2003 among others). Unfortunately, more intuitive descriptions of rhythm as an impression of *prominence patterns* have somewhat been neglected by these approaches: they do provide metrics that can distinguish between languages based on their phonotactic structure. However, they do not answer how these differences relate to rhythmic impressions, e.g. to the perception of iambic or trochaic patterns. In the approach taken here, rhythm is described as a *perceptual impression of a structure evolving over the course of time*. This rhythmic structure is described as a sequence of fundamental rhythmic events (= *beats*), which are perceived as being more or less stressed, e.g. they are varying in prominence and they are grouped in a particular fashion, e.g. as iambic, trochaic or dactylic patterns. In this paper, it is first investigated, how such perceptual impressions of rhythmic grouping and structure are influenced by psychoacoustic phenomena and how they can be linked to meaningful acoustic events fulfilling important functions in speech communication. In a second step, conclusions are drawn from these insights with respect to a useful visualization strategy for phonetic research. A literature review and complementary own investigations lead to the conclusion that relative duration patterns are the major acoustic cue to rhythm: Duration provides the most robust acoustic correlate of perceptual prominence and the shape of relative durations can distinguish between the end and the beginning of rhythmic groups. In a further step, time-delay plots are introduced as a means of visualizing rhythmic properties related to stress and syllable timing, but also to phenomena such as iambic vs. trochaic patterns, rhythmic alternation and isochrony. The approach reveals the way that various languages make systematic differences in order to signal relevant boundaries, e.g. foot or phrase boundaries. However, these differences do not comprise a simple dichotomous system based on ‘stress vs. syllable timing’: Each language may employ one strategy out of many for each boundary type, e.g. phrase or foot boundaries. An analysis of duration patterns across feet furthermore reveals a new explanation for the persistent perceptual classification into so-called ‘syllable timed’ and ‘stress timed’ languages, the former sounding more ‘isochronous’ than the latter. Responsible may be a psychoacoustic phenomenon known as *time shrinking* which leads to the perception of isochrony in decelerating sequences, which are very typical for a syllable timed language such as French.

2. Rhythm as a Sequence of Structured and Grouped Beats

In line with musicology, psychology and metrical theory (e.g. Woodrow 1951; Fraisse 1982), rhythm is in this paper regarded as *a sequence of fundamental rhythmical events (= beats) which are characterized by their prominence structure, e.g. a (more or less regular) order of stronger and weaker beats perceived as groups.*

The distinction between grouping and structure is necessary, because prominence patterns alone are rhythmically underspecified. Given a strictly alternating structure (cf. Fig. 1), e.g. *weak -strong- weak-strong-weak...*, listeners may hear either a trochaic pattern or an iambic pattern starting with *anacrusis*. Further optional grouping strategies for this sequence may be possible as well, e.g. a listener could perceive it as an initial *amphibrach* (*weak-strong-weak*) followed by trochees. Thus, a full rhythmic specification cannot be carried out without taking into account grouping.

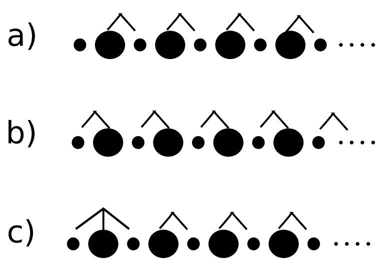


Figure 1. Rhythmic structure and rhythmic grouping are independent. Various grouping strategies are applicable to the identical strictly alternating sequence of strong and weak beats: a) shows a trochaic grouping with an initial anacrusis, b) depicts an iambic grouping and c) is an example of an initial amphibrach followed by a series of trochees

There seems to be wide consensus (e.g. Allen 1975; Lerdahl and Jackendoff 1983; Cummins 2002) that rhythmic structure may exist on several hierarchical levels, the lowest level being a fundamental level of rhythmic beats. These are grouped into rhythmic feet, which are again grouped into higher-level structures, e.g. rhythmic phrases. This hierarchical structure shows much similarity to Metrical-Phonological descriptions of rhythmic patterns, i.e. it resembles metrical grids if the focus of description lies on prominence relations, and it resembles metrical trees if the focus lies on grouping structure (cf. Fig. 2). Rhythmic grouping and rhythmic structure are closely related, since the relation between stronger and weaker beats causes grouping processes – as they are depicted in metrical trees. A prominent beat can signal the beginning (s-w) or the end (w-s) of a group, i.e. in a strictly alternating rhythm, perceptual grouping would make the difference between an iambic and a trochaic rhythm. As argued above, if grouping is ignored in a description, prominence patterns are rhythmically ambiguous – as it is the case with metrical grids. The decision, whether a rhythm is iambic or trochaic (or dactylic, spondeic etc.) may sometimes depend on the level of analysis, i.e. languages differ in their preference of marking the beginnings or ends of groups on various structural levels. E.g. in English and many other Germanic languages, there exists a tendency for final (nuclear) stress on phrase level but for initial stress on compound level (Chom-

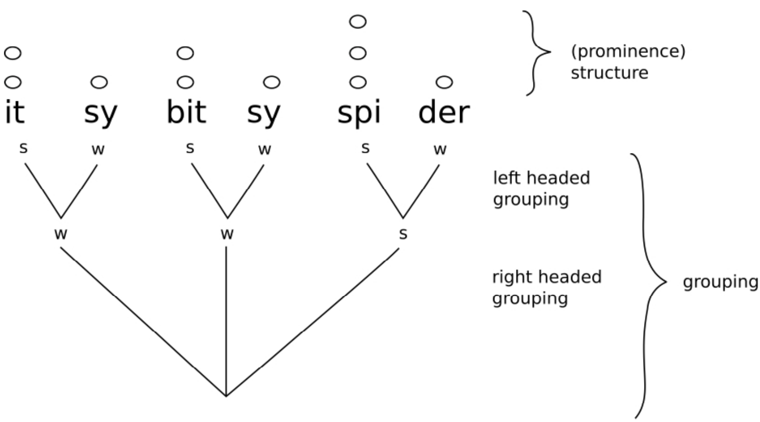


Figure 2. While metrical grids (top) illustrate rhythmic prominence structure, metrical trees (bottom) illustrate rhythmical grouping at various hierarchical levels

sky and Halle 1968). Thus, if we want to build a model able to identify rhythmic prosodic structure in acoustic events, we need to answer the following questions addressed in the subsequent two sections:

a) *What are the fundamental beats in speech*?
b) *What are the acoustic correlates of beginnings and ends* (= *perceptual grouping*) *in rhythmical groups and the perceptual prominence of individual beats* (= *structure*)?

3. Fundamental Beats in Speech

The syllable seems to be a good candidate when searching for a unit resembling *fundamental beats* in speech. It has been shown that there exists a 1:1 correspondence between the number of syllables contained in an utterance and the number of potential accompanying claps (Köhlmann 1984). However, it has to be taken into account that a syllable does not correspond to an *upbeat* in the musical sense, meaning a point where one would almost necessarily clap a hand or tap a foot while listening to a piece of music. Such *accented*

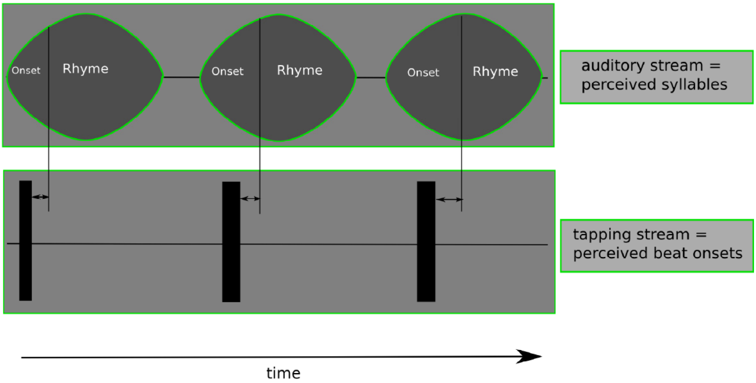


Figure 3. In the tapping paradigm, subjects' taps are registered and aligned with a simultaneous speech stream. The tapping occurrence is taken as evidence for p-center location

positions usually organize grouping at a higher level, i.e. they mark beginnings or ends of rhythmic groups. Thus, a syllable is rather the equivalent of the fundamental beat in a musical measure, e.g. in a 3/4 (waltz) measure the fundamental beat would be a quarter note, but only every third quarter note would demand an accent (and a foot tap). Certainly, this notion falls short with regards to languages organized in morae. It is likely, that these have a more complicated correspondence between fundamental beats in speech and perceived rhythm and need to take into account syllable duration or phonological quantity.

The perceptual starting point of a fundamental beat in speech probably corresponds to the so-called perceptual or p-center (e.g. Marcus 1981; Pompino-Marschall 1989; Scott 1993) which roughly corresponds to the vocalic onset in a syllable. Several suggestions have been made concerning a precise location of the perceptual center, the various strategies all proving similarly successful (Janker 1996). There are two experimental paradigms to investigate the p-center: One is called the tapping paradigm where the asynchrony between the moment of a finger tap and a syllable onset is measured (e.g. Allen 1972; see Fig. 3). The second paradigm uses lists of syllables with variable onset complexities (e.g. *ba, spa, ba, spa*...) where subjects are asked to order these in a perceptually equidistant fashion (Morton *et al.* 1976; see Fig. 4). It is found that subjects arrange the list in such a way that the time span between individual syllables is objectively variable. Instead of equalizing the duration between the syllables, the time span between the vocalic onsets – i.e. the p-centers – is kept identical. The exact location of the thus identifiable p-centers depends on syllable onset complexity: the longer the onset duration, the earlier the p-center is perceived. There seems to be a consensus that the p-center impression is triggered by a strong change in the intensity envelope–coinciding with the vocalic onset–but is also influenced by onset duration. However, it is yet unclear whether the p-center concept can in fact be transferred to running speech and how it relates to phonological units such as the syllable or the mora.

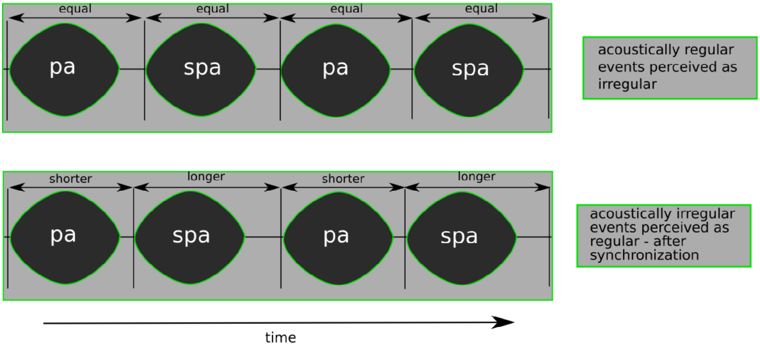


Figure 4. In the adjustment paradigm of p-center detection, subjects are asked to synchronize syllables of different onset complexities with a perceptual equidistance

4. Rhythmic Grouping and Structure in Speech

There are various cognitive and acoustic factors that have an impact on rhythmical grouping and structure in speech. Such factors ought to be considered in any model of speech rhythm and are discussed in the following.

4.1. Tempo Influences

It has long been known (Bolton 1894; Wundt 1911) that the perception of rhythm does not necessarily depend upon the objective presence of a rhythmic pattern in the acoustic signal. If listening to so-called *isochronous pulse trains*, i.e. sequences of acoustically identical beats presented at equal distances, listeners tend to perceive these as beat groups consisting of two, three or four beats, with the first beat of each group being perceived as somewhat stronger than the rest (cf. Fig. 5). The number of beats fitting into one group seems to be constrained by a particular window size, often referred to as the *window of temporal presence* which has been identified as being 400-600ms long (Fraisse 1982) and corresponds to the length of a typical foot in

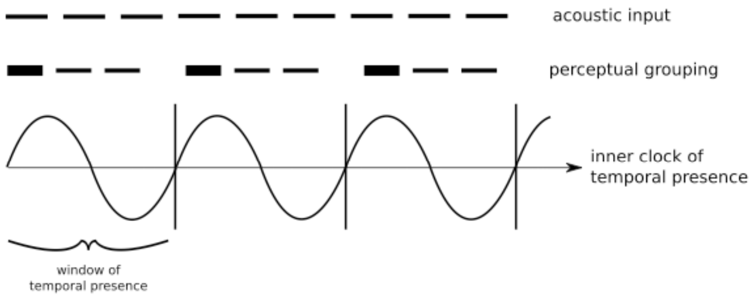


Figure 5. Objectively (acoustically) isochronous pulse trains are perceived as groups of beats with the first beat being perceived as more prominent than the remaining beats in the group

many of the world‘s languages (e.g. Dauer 1983; Roach 1982). Consequently, an increase in tempo (i.e. decreasing the interval between two successive beats) automatically leads to a perception of fewer stresses, since more syllables fit into one window of temporal presence. This circumstance may be at least partly responsible for findings that a language‘s inherent tempo is a very good indicator to distinguish between syllable and stress timing, with syllable timed languages being spoken faster (Dellwo to appear), containing more syllables in a foot (Dauer 1983) or slow speech rates causing beat insertion (Schreuder 2006). Thus, speech tempo plus the size of the window of temporal presence may be responsible for a large amount of known rhythm perception phenomena.

4.2. Time Shrinking

Another interesting effect with respect to the perception of rhythmical groups has been called *time shrinking* (Nakajima *et al.* 1991; ten Hoopen *et al.* 1993) and relates to the phenomenon that in a decelerating sequence of events, the later (longer) event is perceptually shortened. This shrinking effect results in a perceptual impression of relative evenness, which is not present in the acoustic signal (cf. Fig. 6). Time shrinking can propagate across several beats, i.e. even a *perceptually* *shrunken beat* exerts a shrinking effect on a subsequent one (Sasaki *et al.* 2002). However, time shrinking can be blocked in the presence of strict alternation. Thus, listeners try to equalize the durational characteristics of events within decelerating sequences. If this tendency can be shown to be effective in speech as well, it may have an important impact on the perception of rhythmic structure since it would mean that languages preferring groups of decelerating sequences are perceived as more even and showing less prominence variations than languages preferring groups of accelerating sequences.

In order to test the possibility of the time shrinking effect in language-like data, Wagner and Windmann (2009) carried out two perception experiments based on stimuli consisting of 2-, 3- and 4-syllabic sequences of the syllable *ba,* in which they systematically varied the order of long and short syllables but kept the total durations equal across stimuli consisting of an identical number of syllables. The stimuli were manipulated in three conditions, a 'deceleration condition' where each syllable was slightly longer than a previous one, an 'acceleration condition' where the sequence started with a long syllable followed by one or several isochronous shorter ones, and an

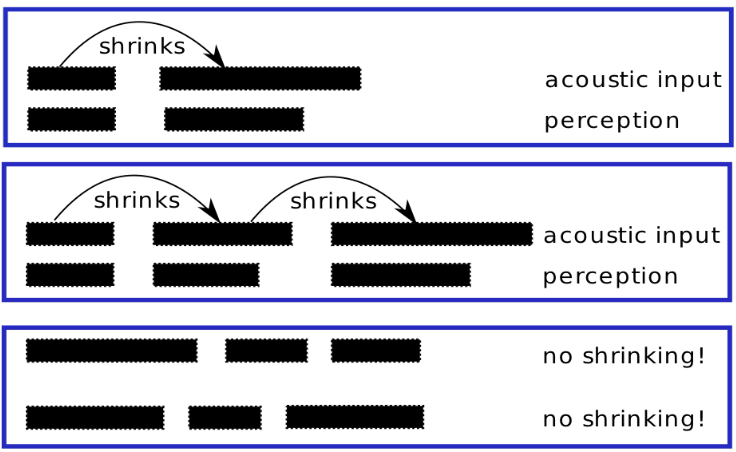
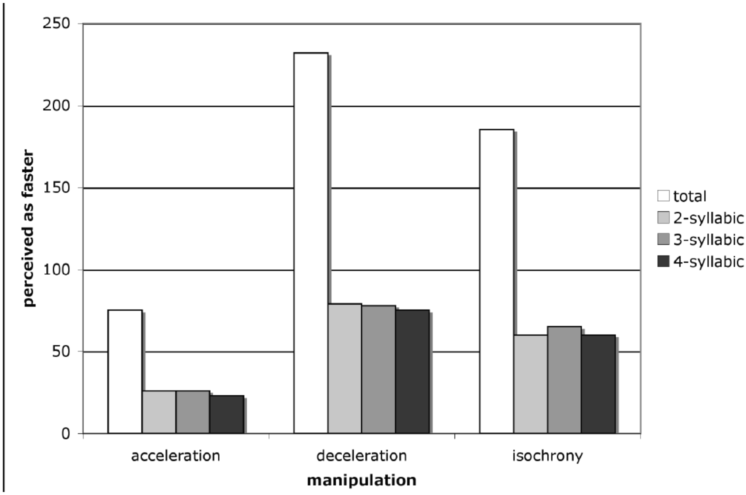


Figure 6. Time shrinking leads to a perceptual shortening of a longer event preceded by a short event. Time shrinking can propagate across several events, i.e. a perceptually shortened event can still shrink a following one



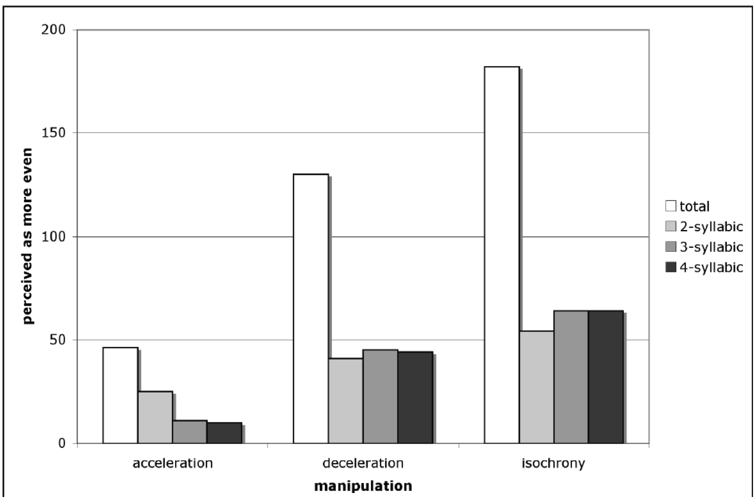


Figure 7

Figure 7. Main results of the study investigation the impact of time shrinking on speech rhythm perception. Syllable sequences consisting of strictly decelerating sequences and those starting with a short syllable followed by longer isochronous sequences are systematically perceived as shorter and more even than accelerating sequences

'isochrony condition' where an initial short syllable was followed by one or several longer isochronous syllables. In pair wise comparison tasks, they could show that the stimulus order indeed had an influence on listeners' perception of tempo and perceptual evenness or 'isochrony': listeners show a strong tendency to perceive accelerating sequences as longer and more variable than decelerating ones that are implemented in both the 'deceleration' and the 'isochrony' condition. It appears that time shrinking is a factor influencing rhythmic perception hitherto overlooked in speech and language rhythm research. Fig. 7 gives an overview of their main experimental results.

4.3. Disentangling Acoustic Parameters of Rhythmic Structure and Grouping

The studies cited above all relate rhythm to duration patterns. In speech, though, rhythmic structure may not only consist of a particular sequence of longer and shorter events but is most likely also influenced by further acoustic parameters that have been indentified as shaping the *perceptual prominence* of speech events. Good candidates for rhythm related acoustic properties besides duration certainly are

- **Overall intensity**, indicating *loudness* or **spectral intensity**, indicating *vocal effort* (e.g. Slujter and van Heuven 1996; Kochanski *et al.* 2005)
- Presence of fundamental frequency excursions (**pitch accents**) (e.g. Fry 1958; Jessen *et al.* 1995), their **excursion** (e.g. Gussenhoven 2004) **and shape** (e.g. Kohler 1991; Ladd and Morton 1997).

Despite the undeniable impact of pitch and intensity on the perception of prominence and despite evidence that the usage of the various prominence lending acoustic cues is language dependent (Barry *et al.* 2007), there is increasing cross-linguistic evidence that duration is the most stable acoustic cue in signaling prominence: Batliner *et al.* (2007) argued that duration is most reliable, since pitch accents tend to be accompanied by lengthening phenomena but not vice versa. The same can likely be said for intensity. Also, Tamburini (2006) and Tamburini-Wagner (2007) found duration to be the most important cue of perceptual prominence for English and German in an acoustic model calculating prominence as a weighted sum of various acoustic correlates.

In shaping speech rhythm, acoustic parameters such as duration, pitch accent (shape) and intensity certainly stand in a complex interaction. It has long been claimed (e.g. Woodrow 1951; Allen 1975) and also experimentally verified (Hay and Diehl 2007) that lengthened beats tend to be interpreted as the ends of a rhythmic group, while an increase in loudness or intensity tends to be interpreted as the beginning of a group – independent on the native language. A problem with these findings is, that they are not easily transferrable to speech data. First of all, pitch excursions have not been systematically investigated yet, but they may play a role in rhythmic grouping as well – the reason for this is their important role in shaping perceptual prominence (see above). Another problem is that in speech, pitch excursions are rhythmically ambiguous, since they can signal both the end (as *boundary tones*) and the beginning of a group (as *pitch accents*). The same ambiguity holds for local increases in duration, which can function as ending signals (*final lengthening*) or beginnings (as *force accents,* Kohler 2005). Of course, there may be cases when accents also indicate ends, such as in an iambic poem. It has been shown, though, that in (German) iambic poems, accentual lengthening is more pronounced than in trochaic ones (Bröggelwirth 2007). Thus, an influence of final lengthening may be at work even here. The only rhythmically unambiguous acoustic feature seems to be intensity. An increase in intensity is unlikely to signal the end of a rhythmical group in speech unless it correlates with lengthening.

In order to replicate Hay and Diehl's results, extend them to pitch excursions and in order to test whether the ambiguities of rhythm correlates in speech can be resolved, a pilot study was carried out based on non-speech event sequences. The following hypotheses were tested:

– H1: A moderate increase in duration is interpreted as the beginning of a group. A strong increase in duration is interpreted as the end of a group.

- H2: An increase in intensity is interpreted as the beginning of a group
- H3: An increase in fundamental frequency does not have a clear impact on perceptual grouping

For the perception task, 4 beat patterns were created all of which were alternating in one acoustic property. The basic event was a 220Hz, 200ms, 60dB sine wave. The pause between successive stimuli was kept constant at 50ms. The first pattern (intended trochaic) varied the duration between successive stimuli with a ratio of 1.5:1, resulting in an event chain of *300ms, 200ms, 300ms, 200ms...* The second pattern (intended iambic) varied the duration between successive stimuli with a ratio of 2:1, resulting in an event chain of *400ms, 200ms, 400ms, 200ms,...* A third pattern (intended trochaic) varied the intensity between successive acoustic events at 10dB, resulting in an alternating pattern where every event was twice/half as loud as the immediately surrounding ones. A last stimulus chain alternated in fundamental frequency (440Hz: 220Hz = 1 octave). This was expected to be rhythmically ambiguous. 10 musically trained and 2 phonetically trained native speakers of German were asked to listen to each pattern as long as necessary to identify it as either iambic or trochaic – the unlimited time was introduced in order to lessen the bias of the initial order of the stimulus sequences. Subjects were allowed to say the pattern was rhythmically unidentifiable. The results (cf. Fig. 8) clearly confirm hypotheses H1 and H2, namely that relative durations can differentiate between an iambic and a trochaic grouping impression. The results also confirm the listeners‘ identification of intensity increases as a beginning and their inability to link an increase in fundamental frequency to rhythmical grouping. Anecdotal responses from the musically trained subjects can be interpreted in such a way that they are influenced by their musical experience: Bassists tend to align low fundamental frequency with group beginnings, while sopranos, tenors and violinists tend to do the opposite. While this pilot study needs further data to be conclusive, it indicates that fundamental frequency increases prominence while not shaping

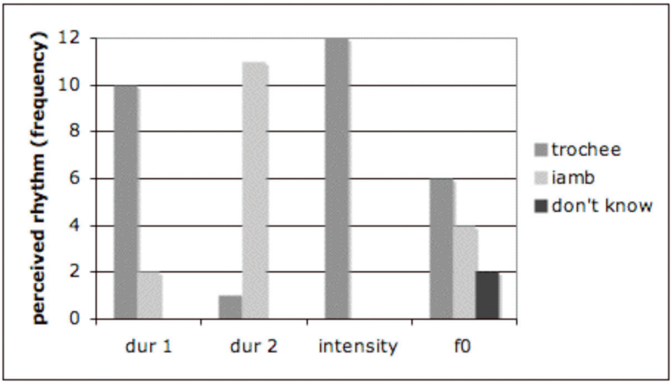


Figure 8. The results of the perception study. “dur 1” denotes an alternating short-long pattern with a moderate durational variation (intended trochaic), “dur 2” denotes an alternating short-long pattern with a stronger durational variation (intended iambic), “intensity” denotes an alternating loud-soft pattern and “f0” denotes an alternating high-low pattern

rhythmical grouping on its own. Intensity is perceived as unambiguously marking the beginning of a group. Duration appears to be the only acoustic factor able to signal both iambic and trochaic grouping.

4.4. Conclusions for Speech Rhythm Modeling

It can be concluded, that duration is the fundamental acoustic correlate of rhythmical grouping. Besides, it appears to be the most robust cue to perceptual prominence across a number of languages. However, duration is functionally ambiguous since it signals both the end and the beginning of a rhythmic group. While relative timing appears to be a major cue able to distinguish between beginnings and ends, other acoustic dimensions such as fundamental frequency and intensity may help in this disambiguation process as well, e.g. by increasing prominence. Thus, a future model of rhythmic structure must take into account the full range of acoustic rhythm correlates.

In order to link rhythmic structures to their prosodic function and in order to detect it in speech, it makes sense to postulate hypotheses concerning the location of important group boundaries. Such hypotheses may of course be language, speaker and speaking style specific. The next section will present a method able to unfold the link between prosodic function and rhythm expressed as relative timing patterns in various languages and speaking styles. Rhythm related typological features such as stress timing and syllable timing will be illustrated as well.

5. Modeling and Visualization of Speech Rhythm Phenomena

With having identified relative duration as the major acoustic correlate of speech rhythm, a new visualization method is introduced that is able to capture the central characteristics of rhythm in speech. In this approach, local and global durational variation as quantified by traditional rhythm metrics are captured as well as relative timing.

5.1. Capturing Rhythmic Phenomena with Time-delay Plots

The basic idea of time-delay plots is simple and has already been used in Nonlinear Time Series Analysis in order to describe nonlinear dynamic systems such as water dropping (Baier 2001).

In a time-delay plot (cf. Fig. 9), the duration of any beat *i* is plotted on the x-axis, and the duration of the subsequent beat *i+1* is plotted on the y-axis. In order to compare relative durations across utterances spoken at different speaking rates, beat durations are z-normalized by:

$$z_i = \frac{x_i - mean(S)}{stdev(S)}$$ for all values contained in the data series $S = x_1, x_2, x_3, \ldots$

Normalization results in values where ‘0’ denotes the mean beat duration and ‘1’ and ‘-1’ denote the negative and positive standard deviation of the examined data series. Beat durations are calculated based on the duration between the beginning of the perceptual center according to Marcus (1981) and the end of the syllable rhyme. This way, the beat duration is primarily determined by the syllable rhyme itself, which is also the locus of a syllable’s prominence (or moraic structure) according to phonological theories. The time delay plots created by this method unveil many rhythm related properties of speech:

– Any sequence of locally isochronous beats (as measured by the PVI metric, cf. Low *et al.* 1999) is plotted near the

9a

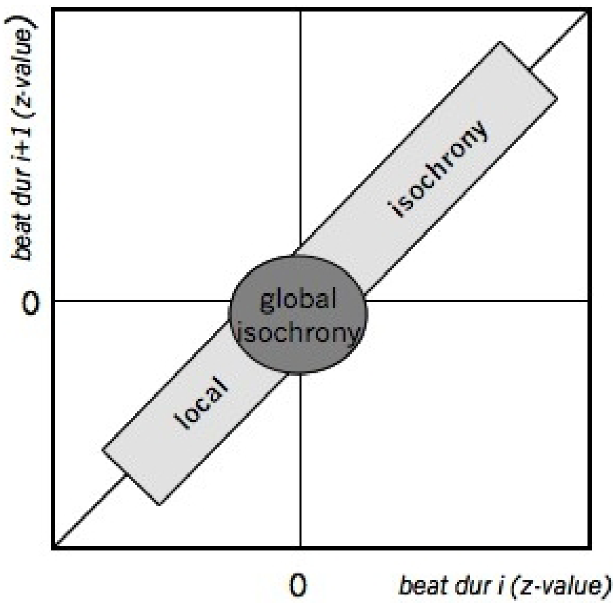
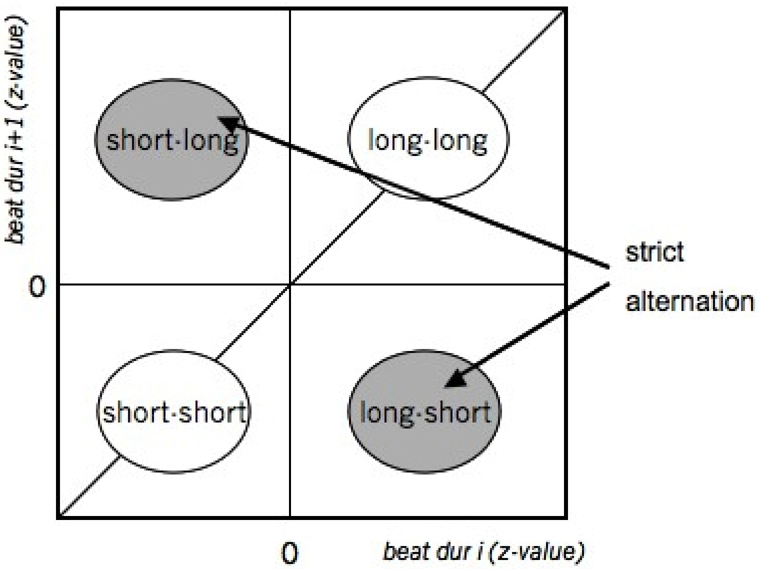


Figure 9. The upper graph 9a illustrates the visualization of global and local isochrony in time-delay plots, the lower graph 9b illustrates the visualization of relative timing. A time-delay plot with concentrations in the upper left and lower right indicates a tendency of strict alternation

9b



bisecting line between the x- and y-axis, i.e. whenever $beatdur_i \approx beatdur_{i+1}$.

– A tendency towards global isochrony (as traditionally postulated for syllable timing) would result in a cluster around the {0,0}-co-ordinate since it should lead to beat durations concentrating around the mean value.
– A strictly alternating rhythm would lead to one cluster in the lower right (long-short) quadrant and another cluster in the upper left (short-long) quadrant.
– Sequences of two short beats are plotted in the lower left quadrant. Sequences of two long beats are plotted in the upper right quadrant.

5.2. Revealing Rhythmic-Prosodic Functions

Of course, the described visualization method remains comparatively uninteresting for research in language and speech, unless revealed patterns can be linked to rhythmic-prosodic communicative functions. Therefore, functionally different prosodic transitions are illustrated differently in rhythm plots:

- **Transition Type 1:** Transitions to lexically stressed beats (= potential foot beginnings) are marked "+"
- **Transition Type 2:** Transitions to unstressed beats are marked "**o**"
- **Transition Type 3:** Transitions to phrase final beats are marked "**x**"

In order to test the basic applicability of the method, two languages are examined, which are usually regarded as belonging to distinct rhythm classes, namely English (as the prototypical 'stress timed' language) and French (as the prototypical 'syllable timed' language). The data was taken from the BonnTempo database (Dellwo *et al.* 2004) and comprises read speech of various speech rates (very slow to very fast). For reasons of clarity, the shown plots are based on material by three randomly chosen different speakers (approximately 450 beats per speaker) per language. All quantitative follow-up examinations are based on the entire database.

Figure 10 shows the results for English and French. It is evident that the three different transition types show much less overlap in English compared to French. The strict separation between transitions to stressed syllables and unstressed syllables in two distinct regions of the plot indicate a tendency towards alternation in English. Besides, final lengthening is (not surprisingly) more pronounced with respect to a relative increase in duration. In French, however, hardly any distinction between the Transition Types 1 and 2 is visible. The lack of word final lengthening in French stands in contrast to the common notion of French being an 'iambic' or 'trailer timed' language. One reason for this may be that French-lengthening effects are confined to the ends of phrases or stress groups. However, a slight change in the plot unveils that the impression of French having word final lengthening is not entirely without empirical foundation: Instead of plotting the transition to the supposedly lexically stressed (= word final) beat, the transition from the word final to the following

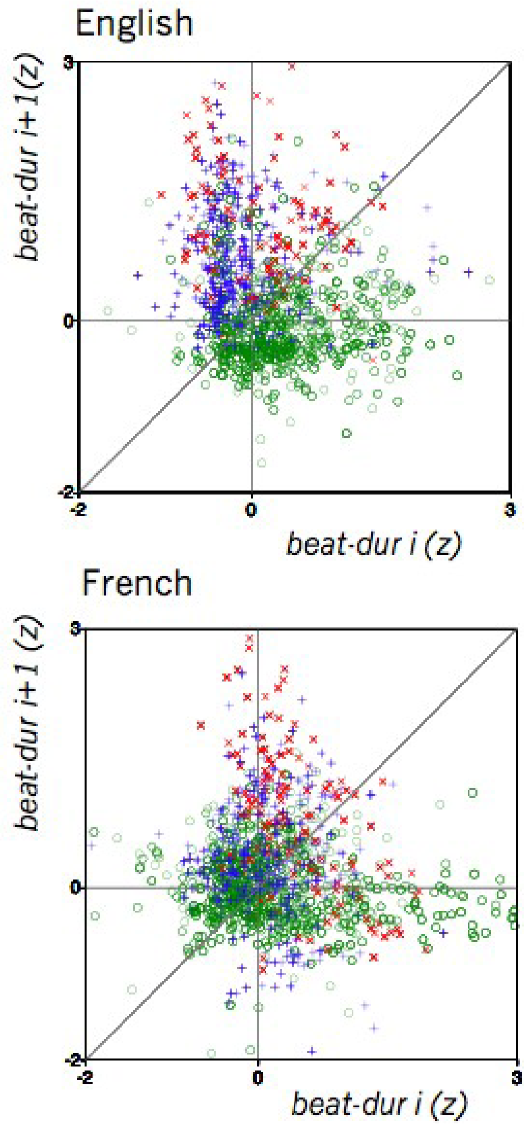


Figure 10. Shown are the time-delay plots for English and French. Transitions Type 1 are marked "+", Transition Type 2 is marked "o" and Transition Type 3 is marked "x". While there is little overlap between Transitions Types 1 and 2 in English, French shows hardly any differentiation between them. Phrase final lengthening, however (Transition Type 3), stands out slightly in both English and French

word initial beat is plotted (cf. Fig. 11). Now it is evident that the presumably 'stressed' word final beat in French is indeed lengthened, at least relative to the subsequent word initial beat. Word boundary marking in French thus appears to be not determined by final lengthening but rather by group initial shortening. Still, from the still considerable amount of overlap in transition types one can conclude that this shortening effect is not as stable as the stress induced lengthening found for English. These impressionistic results show that time-delay plots provide a robust, simple and straightforward tool to reveal formerly known and unknown systematic tendencies of relative timing and link them to their communicative function.

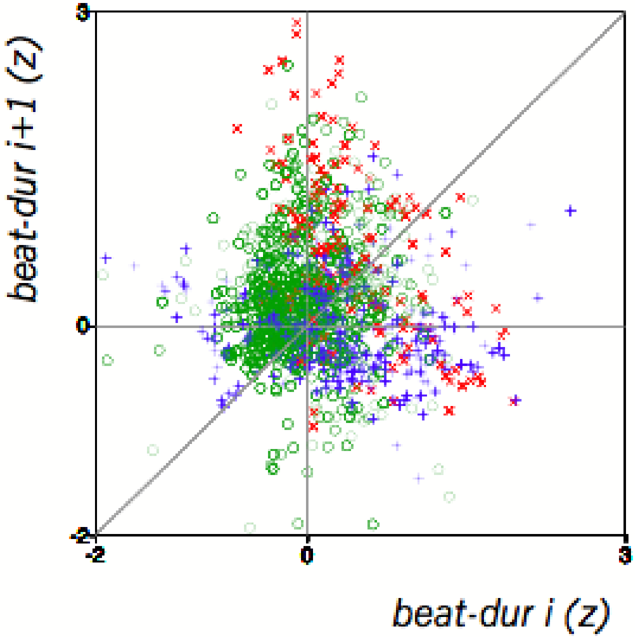


Figure 11. The time delay plot shows relative timing events for French. Transitions from lexically stressed to post stressed beats are marked "+", all other transitions are marked "o" unless they are of Transition Type 3 which is marked "x". Instead of lengthening induced by lexical stress, French seems to be characterized by post stress shortening effects

Table 1. Results of the one-factorial ANOVA comparing variances of the three transition types (differences of adjoining beat durations, z-score normalized) for English and French

English		
Transition Type	*Mean (z-score)*	*Variance*
1 (to stressed)	0.79	0.7
2 (to unstressed)	–0.39	1.72
3 (to phrase final)	1.02	0.78
French		
1 (to stressed)	0.19	0.36
2 (to unstressed)	0.24	0.42
3 (to phrase final)	0.4	1.5

5.2. Quantitative Interpretation of Time-delay Plots

In this section, suggestions concerning the quantitative interpretation of time-delay plots are made. First, durational differences for each transition $beatdur_{i+1}$-$beatdur_i$ are calculated. That way, decelerating sequences are assigned positive values and accelerating sequences are assigned negative values. Then, a one-factorial ANOVA is calculated, revealing whether there exist significant differences between the three transition types initially defined. The results (cf. Tab. 1) confirm the visual impression that in English, the three transition types show highly significant differences, $F = 223$; $p < 0.0001$), while for French, no clear differences between the groups' variances are traceable ($F = 1{,}44$; $p = 0.24$ (n.s.).

In order to get a better understanding of the nature of the differences between the various transition types for each language, we calculate the mean values for each transition type in both dimensions of the time-delay plot, thus identifying the most typical transitions for each language. Then, the Euclidian Distance is calculated between two points $P = (p_x, p_y)$ and $Q = (q_x, q_y)$ representing typical transitions:

$$distance = \sqrt{(p_x - q_x)^2 (p_y - q_y)^2}$$

We now have an additional means to identify whether our first visual impressions were correct, although it should be kept in mind that our calculations measure distance as it would be done with the help of a ruler – it is no adequate metric of perceptual distance! However, it provides us with a useful tool in our purpose of building a visualization method for rhythmic structures in speech. In addition to our two prototypical languages, Euclidian Distances were calculated for three other languages contained in the BonnTempo database, one of which has been traditionally classified as stress timed (German), another one as syllable timed (Italian), while a third language (Polish) has been notoriously hard to place within this dichotomy. Fig. 12 reveals that the Euclidian Distances differentiate between stress timing and syllable timing languages especially with regards to the difference between Transition Types 1 and 2 (distance between transitions to stressed and to unstressed beats), with stress timed German and English showing notably higher distances than syllable timed French and Italian. Polish, however, looks 'even more syllable timed' than French on this scale. The calculation furthermore shows that syl-

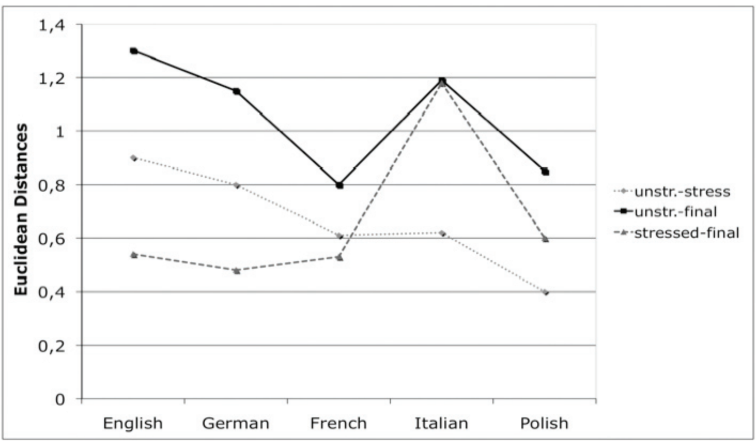


Figure 12. The graph shows the Euclidian Distances between transitions to unstressed beats – to stressed beats (Transition Types 1 and 2), to unstressed beats – to final beats (Transition Types 2 and 3) and to stressed beats – to final beats (Transition Types 1 and 3) for various languages

lable timed Italian has equally pronounced final lengthening tendencies as stress timed English and German, while Polish shows a behavior in between French and Italian in this respect. The calculations based on time-delay plots show that a model that places language rhythm in between the two rhythmic categories *stress timing* and *syllable timing* falls too short. Languages may share certain rhythmic characteristics and behave differently with respect to others. Instead of trying to squeeze speech and language rhythm into an over simplistic dichotomous model, it should rather be described in a multi-dimensional fashion.

5.3. Automatic Classification of Transition Types

The Euclidian Distances compared in the previous section are based on mean values not properly reflecting the variability contained in the data. Besides, they do not indicate whether the language specific distinctions are systematic enough to allow for a generalization. Listeners can use a stable and systematic relationship between relative timing and transition type as a perceptual cue when processing acoustic input. In order to

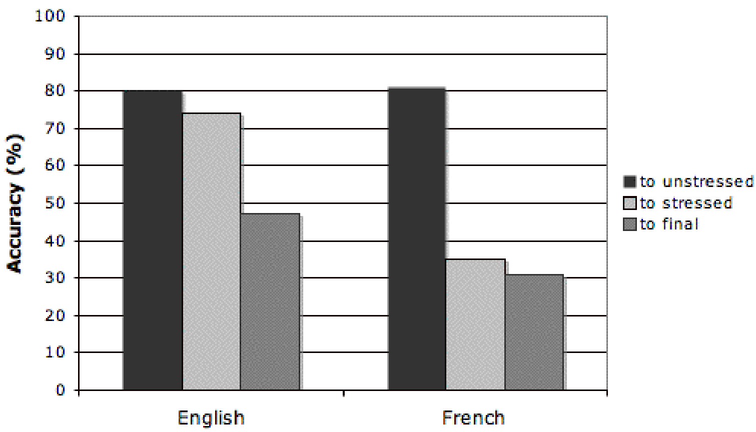
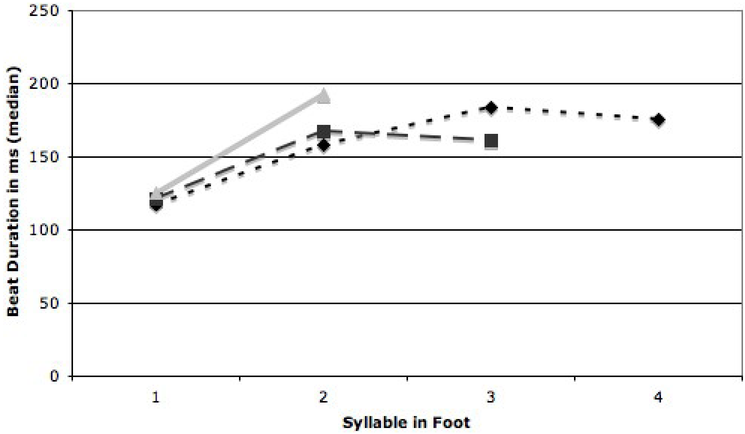


Figure 13. The results of the KNN-classification for the three different transition types

test whether such a systematic relationship exists in the data, a K-Nearest Neighbor (KNN) classification (see Dasarathy 1991 for an introduction) with *k=5* was performed for the two prototypical languages, English and French. The KNN-classifier was used because it also builds on Euclidian Distances thus providing a method that can be straightforwardly linked to the visual interpretation of time-delay plots. The material used for classification was again taken from the BonnTempo database (roughly 2400 transitions per language) and it was tested whether the Transition Types 1–3 were distinguishable for the classifier. From the data, 66% were chosen randomly for training and 33% were used for testing. The hypothesis was, that the classifier could learn to distinguish between the different transition types for English, but not for French with its rather blurred time-delay plots. The results (cf. Fig. 13) confirm this hypothesis quite clearly: while the classifier reaches 73% overall accuracy for English, it perform a lot worse for French with overall 63% accuracy. Furthermore, the prediction is clearly above chance level (based on the relative amounts of transitions in the data), in all transition categories in English. For French, the classification is above chance level as well, but performs worse for all transition types, much worse for the transitions to stressed beats. The comparatively high accuracy for transitions to unstressed beats seems to be mostly the result of the large amount of such transitions in the data itself.

The time-delay plot analysis indicated that French stress is rather characterized by a post stress shortening than a lengthening of the stressed syllable. Therefore, another KNN classifier aiming to identify post stress shortening is trained under the same conditions and on the same material. Indeed, this KNN shows a better performance (40% accuracy) in detecting post stress shortening compared to the previous classifier that had been trained in the detection of stress induced lengthening. Still, this result is still far worse than the KNN classification of the English data. From these results we certainly should not conclude that French behaves ‘arrhythmic’. Rather it can be concluded that its rhythm is organized in a different fashion.

14a



14b

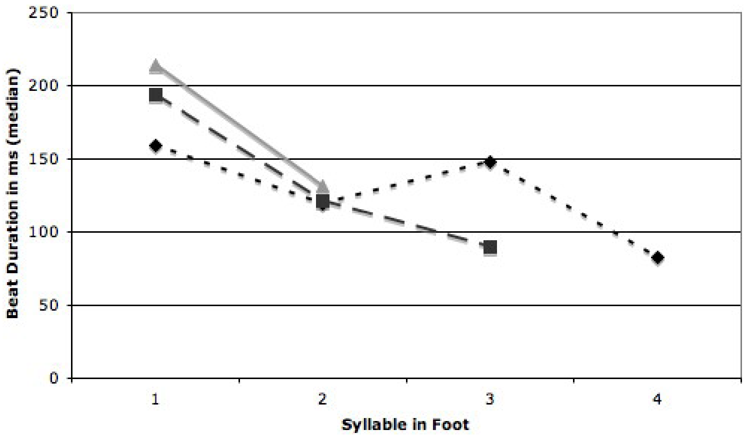


Figure 14. Shown are median beat durations across binary, ternary and quaternary feet in French (top) and English (bottom). The French decelerating patterns across groups can cause time shrinking phenomena, while the English accelerating patterns block time shrinking. The perceptual impression of French being "syllable timed" may be conditioned by time shrinking

5.4. Long Term Timing Regularities

While the time-delay plot analysis visually indicated that French timing is organized between the word final beat and the subsequent one, this tendency was only moderately confirmed by the KNN classification. In this section, another path is taken to investigate timing regularities in French and English. The phonetic foot is chosen as an important unit of rhythmical organization since it seems to be closely linked to what has been identified as window of temporal presence – the timing window constraining rhythmical grouping in isochronous pulse trains (cf. 4.1). We are now looking at long-term characteristics of rhythmical grouping in order to detect timing regularities listeners can use to form expectations concerning upcoming rhythmical events – such long term expectations can be called a language specific abstract *meter*. The timing patterns across feet of different sizes (cf. Fig. 14) clearly reveal characteristic differences. For once, French foot final lengthening is only evident for binary feet. In longer feet, the durational increase propagates subtly throughout the foot and lengthening is not restricted to the final syllable. In English, however, we clearly see a tendency towards binary alternation in quaternary feet, which is missing in French. This automatically leads to a predominance of trochaic and dactylic patterns in English. Furthermore, French 'trailer timing' or 'perceptual iambicity' seems to be not the result of foot final lengthening, but rather of a global deceleration tendency.

These long-term patterns provide an explanation for the often cited impression of French sounding 'isochronous', when taking into account the phenomenon of *time-shrinking* introduced in 4.2: in Wagner and Windmann (2009) it was confirmed that time shrinking leads to a perception of evenness and increased tempo in decelerating sequences as they are typical for French, but is blocked by accelerating or alternating patterns often found in English. Thus, time shrinking may at least partly account for the perceptual impression of French sounding more 'isochronous' and 'faster' than English.

6. Conclusions and Outlook

In the present paper, speech rhythm is regarded as being constituted by beats organized in terms of (hierarchical) rhythmical groups, which have an internal prominence structure. Relative durations were taken as a starting point to identify rhythmical patterns in speech based on psychoacoustic findings that duration is the fundamental acoustic parameter underlying rhythmical grouping and structure. Based on this, time-delay plots were introduced as a visualization method in order to identify language specific relative durations fulfilling communicative functions, e.g. lexical stress or phrase finality. Time-delay plots and subsequent statistical analyses indicated language specific differences in relative timing but also revealed similarities across various languages going beyond a dichotomous classification into stress- and syllable timing. The visual interpretations were evaluated with the help of a KNN-classification. Furthermore, long-term timing patterns across feet could be linked to a psychoacoustic phenomenon called time shrinking. Time shrinking provides an explanation for the famous perceptual impression characterizing French sounding more 'isochronous' than English.

Evidently, the time-delay analysis is useful beyond a mere typological classification. It is able to show interesting intra-language variation across speaking styles or speakers and may provide a robust tool for prosodic L2-analyses (cf. Wagner 2007). Furthermore, the approach looks promising with respect to enhancing automatic language recognition and speech synthesis. Of course, a long-term goal should also be the further development of the method by an integration of other rhythm related acoustic parameters like fundamental frequency or (spectral) intensity in order to capture perceptual prominence rather than pure duration phenomena.

References

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

Allen, J. (1975) “Speech rhythm: its relation to performance universals and articulatory timing”, *Journal of Phonetics* 3: 75-86.

Arvaniti, A. (2009) “Rhythm, Timing and the Timing of Rhythm”, *Phonetica* 66.1/2: 46-63.

Baier, G. (2001) *Rhythmus: Tanz in Körper und Gehirn,* Rowohlt: Reinbek.

Barry, W. J. / Andreeva, B. / Russo, M. / Dimitrova, S. and Kostadinova, T. (2003) “Do rhythm measures tell us anything about language type?”, in M.-J. Solé / D. Recasens / J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona 3-9 August 2003, Barcelona: Causal Productions Pty Ltd, 2693-2696.

Barry, W. / Andreeva, B. / Steiner, I. (2007) “The Phonetic Exponency of Phrasal Accentuation in French and German”, in H. Van Hamme / R.J.J.H. van Son (eds.) *Proceedings of Interspeech* 2007, Antwerp (Belgium), Universiteit antwerpen, Radboud University Nijmegen and Katholieke Universiteit Leuven: Causal Productions Pty Ltd, 1010-1013.

Batliner, A. / S. Seidl / B. Schuller / D. Seppi / T. Vogt / L. Vidrascu / N. Amir / L. Kessous / V. Aharonson (2007) “The impact of f0 extraction errors on the classification of prominence and emotion”, in J. Trouvain / W. J. Barry *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken 6-10 August 2007, Saarbrücken: Universität des Saarlandes, 2001-2004.

Bolton, T. (1894) “Rhythm”, *American Journal of Psychology* 6.1: 145-238.

Bröggelwirth, J. (2007) *Ein rhythmisch-prosodisches Modell lyrischen Sprechstils.* Doctoral disseration: Universität Bonn, http://hss.ulb.uni-bonn.de/diss_online/phil_fak/2007/broeggelwirth_joerg/

Cummins, F. (2002) “Speech rhythm and rhythmic taxonomy”, in B. Bel / I. Marlien (eds.) *Proceedings of the 1st International Conference on Speech Prosody*, Aix-en-Provence 11-13 April 2002, Aix-en-Provence: Laboratoire Parole et Langage, 121-126.

Chomsky, N. / Halle, M. (1986) *The Sound Pattern of English,* MIT Press: Cambridge.

Dasarathy, B.V. (1991) *Nearest Neighbor (NN) Norms: Patterns Classification Techniques*, Los Alamitos: IEEE Computer Society Press.

Dauer, R. (1983) “Stress timing and syllable timing reanalyzed”, *Journal of Phonetics* 11: 51-62.

Dellwo, V. (to appear). *The influence of speech rate on speech rhythm*. Doctoral Dissertation: Universität Bonn.

Dellwo, V. / Aschenberner, B. / Wagner, P. / Dankovicova, J. / Steiner, I. (2004) “Bonn-Tempo Corpus and Bonn-Tempo Tools: A database for the study of speech rhythm and rate”, in ISCA (ed.) *Proceedings* *of Interspeech 2004, – 8th International Conference of Spoken Language Processing,* 4-8 October, Jeju Island, Korea: ISCA, 770-780.

Fraisse, P. (1982) "Rhythm and tempo", in D. Deutsch (ed.) *The Psychology of Music,* New York: Academic Press, 149-180.

Fry , D.B. (1958) "Experiments in the perception of Stress", *Language and Speech* 1: 126-152.

Gussenhoven, C. (2004) *The Phonology of Tone and Intonation*, Cambridge: CUP.

Hay, J / Diehl R. (2007) "Perception of rhythmic grouping: Testing the iambic/trochaic law", *Perception & Psychophysics* 69.1: 113-122.

Janker, P.M. (1996) "Evidence for the p-center syllable-nucleus-onset correspondence", *ZAS Papers in Linguistics* 7: 94-124.

Jessen, M. / Marasek, K. / Schneider, K. / Classen, K. (1995) "Acoustic correlates of word stress and the tense / lax opposition in the vowel system of German", in K. Elenius / P. Branderud (eds.) *Proceedings of the 13th International Congress of Phonetic Sciences* (Stockholm, Sweden), Stockolm: KTH and Stockolm University 4, 428-431.

Kochanski, G. / Grabe, E. / Coleman, J. (2005) "Loudness predicts prominence: fundamental frequency lends little", *Journal of the Acoustical Society of America* 118: 1038-1054.

Kohler, K. (1991) "Terminal intonation patterns in single accent utterances of German: Phonetics, phonology and semantics", *AIPUK (Arbeitsberichte, Institut für Phonetik und digitale Sprachverarbeitung, Universität Kiel),* 25: 115-185.

Kohler, K. (2005) "Form and function of non-pitch accents", *AIPUK* 35a: 97-123.

Köhlmann, M. (1984) "Bestimmung der Silbenstruktur von fließender Sprache mit Hilfe der Rhythmuswahrnehmung", *Acustica* 56: 120-125.

Ladd, D.R. / R. Morton (1997) "The perception of intonational emphasis: Continuous or categorical?", *Journal of Phonetics* 25: 313-342.

Lerdahl, F. / Jackendoff, R. (1983) *Generative theory of tonal music,* Cambridge: MIT Press.

Low, E.L. / E. Grabe / F. Nolan (1999) "A contrastive study of prosody and lexical stress placement in Singapore English and British English", *Language and Speech* 42.1: 39-56.

Marcus, S. (1981) "Acoustic determinants of perceptual center (p-center) location", *Perception & Psychophysics* 30: 247-256.

Morton, J. / S.M. Martin / C. Frankish (1976) "Perceptual Centers (P-Centers)", *Psychological Review* 83: 405-408.

Nakajima, Y. / G. ten Hoopen / R. van der Wilk (1991) "A new illusion of time perception", *Music Perception* 8: 431-448.

Pompino-Marschall, B. (1989) "On the psychoacoustic nature of the p-center phenomenon", *Journal of Phonetics* 17: 175-192.
Roach, P. (1982) "On the distinction between 'stress timed' and 'syllable timed' languages", in D. Crystal (ed.) *Linguistic Controversies, Essays in linguistic theory and practice*, London: Edward Arnold, 73-97.
Sasaki, T. / Suetomi, D. / Nakajima, Y. / ten Hoopen, G. (2002) "Time-shrinking, its propagation, and gestalt principles", *Perception & Psychophysics* 64: 919-931.
Schreuder, M. (2006) *Prosodic processes in language and music.* Doctoral dissertation: University of Groningen, *GRODIL* 60 (= Groningen Dissertations in Linguistics).
Scott, S.K. (1993) *P-centers in speech: An acoustic analysis.* Doctoral Dissertation: University College London.
Slujter, A. / van Heuven, V. (1996) "Spectral balance as an acoustic correlate of linguistic stress", *Journal of the Acoustical Society of America* 100: 2471-2485.
Tamburini, F. (2006) "Reliable prominence identificiation in English spontaneous speech", in R. Hoffmann / H. Mixdorff (eds.) *Proceedings of the 3rd International Conference on Speech Prosody 2006*, Dresden, May 2-5 2006, ISCA Archive: http://www.isca-speech.org/archive/sp2006 (paper 019: http://www.isca-speech.org/archive/sp2006/papers/sp06_019.pdf).
Tamburini, F. / Wagner P. (2007) "On automatic prominence detection for German", in H. Van Hamme / R.J.J.H. van Son (eds.) *Proceedings of Interspeech* 2007, Antwerp (Belgium), Universiteit antwerpen, Radboud University Nijmegen and Katholieke Universiteit Leuven: Causal Productions Pty Ltd, 1809-1812.
ten Hoopen, G. / R. Hilkhuysen Y. Nakajima / F. Yamauchi (1993) "A new illusion of time perception-II", *Music Perception* 11: 15-38.
Wagner, P. (2007) "Visualizing levels of rhythmic organization", in J. Trouvain / W. J. Barry *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken 6-10 August 2007, Saarbrücken: Universität des Saarlandes, 1113-1116.
Wagner, P. / Windmann A. (2009) "No Time to Lose: Time Shrinking Influences the Perception of Tempo and Rhythmic *Isochrony*", in M. Uther / R. Moore / S. Cox (eds.) *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009),* Brighton 6-10 September 2009, Brighton: Causal Production, 1523-1526.
Woodrow, H. (1951) "Time perception", in S.S. Stevens (ed.) *Handbook of Experimental Psychology*, New York: Wiley and Sons, 1224-1236.
Wundt, W. (1911) *Grundzüge der psychologischen Physiologie*, Band 3, Leipzig: Wilhelm Engelmann.