

Editorial

Whither Speech Rhythm Research?

Scientists are not immune to repeating the same kind of methodological errors others have made in a related field of research before. There is a lot of truth in the saying that one does not learn from others' mistakes. Speech science in general, and speech rhythm research in particular, are no exception.

In the Preface to the *Bibliotheca Phonetica* volume *Speech and Speaker Recognition*, Manfred Schroeder [1985] says about the state of the art of automatic recognition of speech at the time: '... one of the main impacts of the computer has been to demonstrate the manifest inadequacy of superficial algorithms that take no account of context and meaning. The simple-minded computer per se was not the hoped-for cure-all, and speech recognition was in acute danger of withering in the laboratory rather than blooming in the field ...' Schroeder refers to Pierce [1969, 1970], who, one and a half decades earlier, launched a scathing attack on practitioners of ASR, which culminated in the assessment that: 'General purpose speech recognition seems far away. Special-purpose speech recognition is severely limited.' And he concluded that 'it would seem appropriate that before embarking upon such work, the worker should candidly ask and answer the following questions: Why am I working in this field? What particular thing do I hope to accomplish? Why is it worthwhile? Am I likely to succeed? How will I know whether or not I have succeeded? Where will success take or leave me?'

Since Schroeder's publication, tremendous advances have been made through the successful consideration of context and meaning in the machine recognition process. So, Pierce's objections need to be seen in a somewhat different light, but this does not mean that they have all been met, and in particular his questions are still as relevant as they were 40 years ago. And they have become more important than ever, since ready-made computer software for speech processing has pervaded all areas of speech science, including linguistics. There they may be applied in a conveyor-belt fashion, and sometimes even blindly, as standard tools to linguistic, phonological and phonetic questions by researchers who do not necessarily understand what these algorithms can or cannot do, but look at the representation of observable speech phenomena by metrics as the order of the day.

One area in speech science, which has had scholars' attention since antiquity and has been subjected to measurement for about 70 years, has seen an explosive surge of interest in the past decade: the study of rhythm in speech and language. The development of rhythm metrics as indices for the algorithmic classification of languages into either discrete classes, such as stress or syllable timing, or scalar weightings between these extremes, parallels the development of algorithms for automatic speech pattern recognition (admittedly, with a good deal less algorithmic know-how and sophistication on the part of linguists). All these metrics are based on segmental duration patterns,

which capture temporal structures in the phonologies and phonotactics of different languages, but they cannot encompass rhythm because they lack the essential ingredient of rhythm, namely prominence and its recurring patterns over time, triggered by physical properties among which vowel and consonant durations are but some besides several others. The various duration metrics have thus been applied to language data without a clear conception of what speech rhythm is and how it may differ between languages in the way they bundle its physical exponents. Moreover, the duration measures are almost exclusively obtained from text reading by native speakers of a language who are not screened for their rhythmic proficiency, although listening to Martin Luther King or Barack Obama, on the one hand, and announcers in airports or mumbling university lecturers and students, on the other, can convince anybody (before taking any measurements) that speakers of the same geographical, social and stylistic variety of a language do not speak with one tongue when it comes to rhythmicity. But production data obtained from subjects willing to have their text readings recorded are entered indiscriminately into samples for analysis. This shows another deficiency of these metrics: they disregard the listener. Furthermore, rhythm research based on these duration measures does not ask the question as to what is the function of rhythm in speech communication, which may be seen in facilitating the transmission of meaning from speaker to listener by recurrent sound patterns. If this functional perspective had guided experimental analysis, researchers would have started their data collection with nursery rhymes, which may be expected to embody speech rhythm in its clearest and most basic form since they are also accompanied by regular body movement, for example knee jiggling.

All these aspects of rhythm in speech and language were known 70 years ago, but they have been covered up by empirical measurement in the very restricted field of segment duration, and subsequently been lost sight of. At this point another quotation from Pierce's 'Letter to the Editor' of *JASA* shows the parallelism between speech recognition algorithms and rhythm metrics even more pointedly – the references to speech recognition just need to be replaced by those to rhythm:

The typical recognizer gets it into his head that he can solve 'the problem'. The basis for this is either individual inspiration ... or acceptance of untested rules, schemes, or information. ...

We would expect a scientist to check the literature concerning ideas, schemes, or information. Perhaps some point is old and has been established or confuted by a clear, simple, definitive experiment. If there is no clear experimental evidence, it might be possible for a scientist to devise a clear, simple definitive experiment. So, a science of speech might grow, certain step by certain step.

The typical recognizer will have none of this. He builds or programs an elaborate system ... No simple, clear, sure knowledge is gained. The work has been an experience, not an experiment.

If researchers in speech rhythm were to ask themselves the first of Pierce's questions, some of them might have to answer candidly 'because I have jumped on the bandwagon of an in-topic which attracts some funding, and I am conducting a project, writing a dissertation, preparing papers for the ever increasing number of workshops and conferences to enhance my academic profile in my CV for job applications'.

As regards Pierce's questions 2 and 3, which refer to the ultimate goal of the research, it is very difficult to make out what particular thing the developers of duration-based rhythm metrics have been hoping to achieve. From about 1940 onwards, anglophone phoneticians described English impressionistically in comparison with either French in Europe (Lloyd James, Abercrombie) or Spanish in North America (Pike) as having a stress-timed as against a syllable-timed rhythm, i.e. isochronous stress-foot or syllable sequences, respectively. This rhythm categorization as a feature

of a whole language was quickly generalized: any language belonged to either one or the other class. So, experimental analysis set out to measure foot or syllable isochrony in languages allocated to these two classes, although it was known at the time (for example, from a Frenchman's investigation into the rhythm of English prose [Classe, 1939]) that isochrony in the acoustic signal is impossible and that the transmission of meaning structures interferes with rhythmic structuring so that it is equally impossible to find one type of rhythmicity permeating all speech in a language. Moreover, the adequacy of a stress-foot concept was not established separately for each language included in the investigation. Again, it was known at the time that it is not valid for French, although French certainly has sentence accentuation. This means that the investigators of the two isochrony types had no clear idea of what their object of measurement was, and the failure to prove isochrony of either type was to be predicted a priori. Thus, nothing was accomplished, there was not even an explanation of why the highly skilled practical anglophone phoneticians had the perceptual impression of French and Spanish being characteristically different from English in their rhythmicity, for which they used the labels syllable- and stress-timed. To have a clearly defined object of analysis the measurements would have had to be set in relation to perceptual thresholds of isochrony, i.e. the listener would have had to provide the yardstick, as was the case in Ilse Lehiste's research.

Speech rhythm then became regarded as an epiphenomenon of the phonotactic structure of a language, especially with regard to the complexity of consonant clusters and to vowel reduction in unstressed syllables, turning the either-or rhythm classes into a gradient between two prototypical manifestations. But the two categories were kept, and languages seen as more or less the one or the other. To metricize this scalar rhythmical exponency across languages new measures were taken, with regard to the phonological reorientation: the percentage of vowel durations and the variability of consonant durations, or the Pairwise Variability Index (as well as some derivatives) of vowel and consonant durations, in a speech sample. Quite apart from the fact that these metrics were frequently inserted into the preset stress- and syllable-timing dichotomy for the languages under investigation, thus leading to circular argumentation, it was to be expected, again a priori, that new data samples would lead to different results and thus different rhythm classifications. So, it was still not clarified what the investigation was to achieve nor in what way it was worthwhile.

Answers to Pierce's questions 4–6, which refer to the success of the research, follow from the preceding. In view of the opaque conception of rhythmical duration measures, and the disregard of available knowledge of timing structures in languages, it would have been clear from the outset that researchers were not in a position to succeed in determining what constitutes stress- or syllable-timed rhythm in different languages had they asked the question. And there was no way for them to know whether their metrics were successful representations because it was not ascertained, independently of the individual measurements, how adequate they were in capturing rhythm in speech and language. Thus, experimental rhythm research has left us wondering where we need to go.

For a start, we need a definition of what speech rhythm is and, building on it, hypotheses as to the physical variables that enter into the production and perception of rhythm. The labels stress-timed and syllable-timed and any form of isochrony are best abandoned: a classificatory rhythm scale for languages, from prototypical stress-timed to prototypical syllable-timed, should not be postulated at the outset, rather, rhythmic

analyses of speech performance in individual languages, and subsequent comparisons, result in a post hoc rhythm typology across the languages of the world. We should also look at speech rhythm, not as an epiphenomenon of phonological description, but as a feature in its own right, fulfilling a function in communicative interaction, where a speaker guides a listener more efficiently through the transmission of meaning by recurring prominence patterns. Speech rhythm is different from, and goes beyond, phonology-driven speech timing, imposing a performance structure for communicative effect. This means that the individual performers need to be taken into account in the investigation since they vary along a scale from poor rhythmic speakers to orators. This progression from the study of rhythmic speech action in individual languages to their comparisons and finally to rhythm classes reverses traditional analysis procedures, and with its anchoring in speech communication, meaning is brought back into the analysis.

The reason why we are working in this field is to elucidate prosodic grouping and its communicative function in speech interaction in the languages of the world. Results of this research are worthwhile because they allow us, for instance, to make empirically grounded statements as to rhythmic deficiencies and consequent poor intelligibility of native speakers of a language, or of foreign speakers between languages, e.g. a French speaker speaking English, or vice versa, and thus to devise viable plans, for example to improve the rhythmic proficiency in public announcements, including their automatic synthesis, or in foreign language teaching and learning. The outcome of this line of research will also provide an explanation as to why the skilled anglophone phoneticians proposed the rhythmic dichotomy in the first place. The stepwise progression starting from clear definitions is highly likely to succeed in the end, intermingled with failures from which we will be able to learn how to continue, and success can be measured against the initial hypotheses and the practical application in other areas of speech communication. This line of research will contribute to our understanding of the overarching question in communicative phonetic science: how do humans communicate by speech in the languages of the world?

Speech rhythm has been a topic at all the International Congresses of Phonetic Sciences since 1979. Initially, it was discussed under the heading of *Temporal Relations within Speech Units*, at a Symposium at Copenhagen, which was moderated by Ilse Lehiste. In her Introduction [Lehiste, 1979], she stresses the need to relate timing control in production to the role of timing in perception, and raises the question whether the temporal organization of utterances is determined by their syntax, or whether there exist rhythmic principles in production and perception that are at least partly independent of syntax. She refers to researchers who

have developed programs for generating the temporal structure of a sentence on the basis of segments and syntactic structure, without paying any attention to rhythm. ... in generative grammar, phonetic output is the last step in the generation of a sentence. An independent rhythm component simply has no place in the theory. ... the speech units are segments, phrases, clauses, and sentences. (And it is quite interesting to see them struggle with units not foreseen in the theory, like syllables and phonetic words.) Researchers who are not fully committed to this theoretical viewpoint operate with certain other units, such as speech measures or metric feet.

The paper by Klatt [1979] in this Symposium is an example of an approach that starts from an abstract linguistic description and moves to a sequence of segments, whose durations are conditioned by other segments and by syntactic constraints. On the other hand, Fujisaki and Higuchi [1979] present experimental data that support a model in which motor commands and the articulatory/acoustic realizations of

successive segments are programmed in such a way that the perceptual onsets of successive segments are isochronous. The papers by Huggins [1979] and by Donovan and Darwin [1979] are also concerned with speech rhythm. The former finds that a correct rhythmic pattern enhances intelligibility. The latter deals with the perceived rhythm of speech and confirms Lehiste's previous findings that listeners tend to hear utterances as more isochronous than they really are. The authors question the value of seeking direct links between syntax and segmental durations rather than indirect ones by way of an overall rhythmic structure as a perceptual phenomenon, which is not independent of intonation and is confined to language as against non-speech stimuli.

The topic was continued under various headings and in various forms: an Oral Session *Temporal Organisation of Speech* at Utrecht in 1983, a Symposium *Rhythm and Metrics* and an Oral Session *Metrical Theory* at Tallinn 1987, an Oral Session *Timing and Rhythm* at Aix-en-Provence in 1991, a Poster Session *Stress and Timing* at Stockholm in 1995, Poster Sessions *Intonation: Rhythm I, II, III* at San Francisco in 1999, an Oral Session and a Symposium (organized by Pier Marco Bertinetto) *Stress and Rhythm*, and an Oral Session *Prosody: Rhythm and Phrasal Structure* at Barcelona in 2003, an Oral Session *Stress and Rhythm* at Saarbrücken in 2007. The papers in these sessions do not form homogeneous packages of rhythm research but are a mix of questions of segmental timing, stress, intonation, and rhythm under the labels of stress- and syllable-timing.

But the Aix organizers also put on a Special Session *Timing and Rhythm* with an invited speaker and five invited discussant papers: Sieb Nooteboom, 'Some Observations on the Temporal Organisation and Rhythm of Speech'; Fredericka Bell-Berti, 'Comments on "Some Observations on the Temporal Organisation and Rhythm of Speech"'; Rolf Carlson, 'Duration Models in Use'; Gunnar Fant, 'Units of Temporal Organization: Stress Groups versus Syllables and Words'; Gunnar Fant, Anita Kruckenberg and Lennart Nord, 'Temporal Organization and Rhythm in Swedish'; Klaus J. Kohler, 'Isochrony, Units of Rhythmic Organization and Speech Rate'. In this Special Session, the concepts of speech timing and speech rhythm, as well as units of timing – words versus rhythmic stress groups – were clearly differentiated and controversially discussed. Nooteboom confined the field to the timing of words, whereas Fant and Kohler argued for both meaning-oriented timing and independent rhythmic structuring.

This line of discussion was taken up again in the opening Plenary Session at the Barcelona ICPHS with the invited lecture by Yoshinori Sagisaka, 'Modeling and Perception of Temporal Characteristics in Speech', and the paper by the invited discussant Klaus J. Kohler, 'Domains of Temporal Control in Speech and Language: From Utterance to Segment'. Timing, duration, and rhythm were defined as separate concepts in the analysis of speech, where rhythm refers to recurring effects of temporal control over longer stretches than the abstracted segment – syllable, foot – without these domains having to be turned into durational (or even isochronous) units, and this control is additional to, and interacts with, lexical and syntactic structuring. Chunking through pitch patterns, especially pitch accent sequencing, can create larger grouping domains which widen the temporal margins of regularity in production and perception and thus allow greater durational variance without disrupting the rhythmic flow. It was argued that isochrony is not a feature of rhythmic patterning, that rhythm analysis needs to include pitch as a grouping parameter, and that languages differ in the rhythmic control of speech by coordinating short-term vocal tract opening-closing gestures differently with more long-term pitch grouping.

More recently, various oscillator-based mechanisms were proposed to account for rhythmicity. So, Eric Keller and Robert Port organized a Symposium *Speech Timing: Approaches to Speech Rhythm* at the Saarbrücken ICPHS, which included papers by Plinio Barbosa ('How Prosodic Variability Can Be Handled by a Dynamical Speech Rhythm Model'), Michael Connolly Brady and Robert Port ('Quantifying Vowel-Onset Periodicity in Japanese'), Eric Keller ('Waves, Beats and Expectancy'). These papers point to vowel onsets as focal speech events whose distribution in time can be attracted to oscillatory systems of various kinds. In this connection, P(erceptual) centres are reintroduced in the discussion [for a recent overview of P-centres see also Pfitzinger, 2001].

So, in addition to the development of rhythm metrics, there has been extensive discussion of the rhythm issue in relation to timing, prosody, and syntactic-semantic grouping in speech for three decades. In this research setting, it was timely and highly laudable on the part of Volker Dellwo (UCL), Elinor Payne (Oxford), Petra Wagner (Bonn), and Laurence White (Bristol) to organize a workshop on speech rhythm in its own right: *Workshop on Empirical Approaches to Speech Rhythm*, which took place at University College London on 28 March 2008. It was advertised in a Call for Papers:

Empirical studies of speech rhythm are becoming increasingly popular. Metrics for the quantification of rhythm have been applied to typological, developmental, pathological and perceptual questions. The prevalence of rhythm metrics based on durational characteristics of consonantal and vocalic intervals (e.g. ΔV , ΔC , %V, nPVI-V, rPVI-C, VarcoV and VarcoC) indicate the need for agreement about their relative efficacy and reliability. More fundamentally, it remains to be demonstrated whether such metrics really quantify speech rhythm, a controversial and elusive concept. We invite abstract submissions for a limited number of ... oral presentations, and for poster presentations. We welcome abstracts that address any or all of the following questions: What is speech rhythm? How should we measure speech rhythm? Which rhythm metrics are most effective and reliable? What can rhythm metrics tell us? What are the limitations of rhythm metrics?

Since this workshop programme looked very promising in view of the turbulent history of rhythm research sketched above, *Phonetica* opened its pages for a small number of the most rewarding presentations, expanded and written up as journal papers, which, after review by two reviewers, were judged to address the above scientific questions most successfully. This Thematic Issue on *Rhythm in Speech and Language – From Theory to Data* is the result of a multi-tiered selection process, uniting seven papers centred on a common goal – new directions in rhythm research that investigate phonetic forms and communicative functions of recurrent prosodic grouping patterns, beyond segmental timing, in production and perception of different languages. The title of the issue reflects the progression from speech to language and from theory to data. Rhythmical performance in different speaking styles needs to be investigated within various individual languages, in parallel and according to the same hypotheses, before whole languages can be compared as to their rhythmic similarities or differences. This methodology needs to be built on a theory of rhythm and on ensuing hypotheses about its forms and functions in speech communication. The goal is thus not the physical substantiation of a linguistic typology but the elucidation of communicative interaction by speech in languages. This is the answer to the questioning title of Pierce's 'Letter to the Editor' of *JASA*, adapted to speech rhythm.

The discussion is opened with Fred Cummins' paper 'Rhythm as an Affordance for the Entrainment of Movement', which starts from a general account of rhythm in humans as coordinated behaviour in a wide variety of situations. So, rhythm is in the active bodily performance, not as a specific property of the signal, but as an affordance

to entrain other movement. Speech rhythm can occupy only a marginal place in this conception of rhythm, although the discussion of experimental evidence for speech cycling and synchronous speech shows that it also has the potential to entrain movement. And nursery rhymes accompanying movement, such as knee jiggling, are further support. This approach questions the validity of traditional metrics for insightful statements about speech rhythm, and it also points to nursery rhymes as a primary database before the analysis of read prose texts or of spontaneous speech is attempted, particularly since in the latter, the transmission of meaning takes precedence over rhythmic performance.

The paper by Klaus J. Kohler, 'Rhythm in Speech and Language: A New Research Paradigm', first gives a historical overview of rhythm research during the past 70 years, but leaves a more detailed discussion of the current rhythm metrics to the papers by Amalia Arvaniti and Francis Nolan. Then parameters that create rhythmicity are introduced: recurring patterns of fundamental frequency, syllabic duration, syllabic energy, spectral dynamics generate a regular recurrence of waxing and waning prominence profiles across syllable chains over time. A new paradigm for rhythm research is proposed, which takes variability in productive and perceptual performance as its point of departure to establish flexible rhythmic patterns. It gives the listener the key role in deciding on what constitutes rhythmic regularity, which is considered scaled according to speaker, speaking situation and style and has a guiding function for the listener. Rhythm is thus not a fixed typological prominence pattern for groups of languages but is variable within each language. But it is also determined by the language in that the potential rhythmical parameters are bundled differently in the languages of the world depending on how they are tied at other linguistic levels.

In the paper 'Rhythm, Timing and the Timing of Rhythm', Amalia Arvaniti continues the historical survey of recent rhythm metrics in greater detail, and argues that the metrics are unreliable predictors of rhythm, as they produce no more than a crude measure of timing. Timing is different from rhythm, and equating the two has led to circularity and a psychologically questionable concept of speech rhythm. Rhythm research is to be based on the same principles for all languages, namely grouping and prominence, for which evidence is provided. So, this paper reaches the same type of conclusions as the preceding one.

The paper by Francis Nolan and Eva Liina Asu, 'The Pairwise Variability Index and Coexisting Rhythms in Language', retraces the origins, and the rationale, of the Pairwise Variability Index of vowel and consonant durations, characterizing it as 'irredeemably trivial and disproportionately successful', its success being gauged 'in terms of its widespread adoption and application' (due to its easy calculation), rather than 'in terms of how much [it] has advanced our understanding of the nature of speech rhythm'. The original nPVI was to capture the point-to-point duration variability, through a data sample, of vowel or consonant segments, normalized for speech rate. It is important to note two aspects of the initial application of this metric. First, it was to compare the temporal organization in two varieties of English, RP and (educated) Singapore English, both based on the same lexicon and grammar, but differing in their phonetic accent variety, more particularly in their timing, which had been impressionistically described as Singapore English sounding more syllable-timed than RP. This common language provided a referential basis for the comparison of two groups of speakers, expected to cluster round significantly different PVIs for vowels and consonants, in parallel to the impressionistic observations, and such a difference was interpretable as

a rhythmic difference. Secondly, PVI was also obtained for energy variability. Thus, the two measures combined were able to point to different prominence profiles in the two varieties of English. All later applications of the PVI (and its derivatives) for segment durations were not similarly constrained, and the PVI lost its information value in rhythm research. The authors of this paper want to move the PVI concept as a rhythm metric into a new direction by calculating PVI for syllable and for foot durations. They argue that stress-timing and syllable-timing are not end points of a continuum but orthogonal dimensions, and they present data from Estonian, English, Mexican and Castilian Spanish to support their hypothesis, well aware of the problems of language-specific foot definitions for the compared languages. Filling the PVI concept with new, more adequate empirical measures, this paper bridges the theoretical and methodological orientation of the first three papers and the experimental data analyses of the following three contributions in this Thematic Issue.

In their paper ‘Do Rhythm Measures Reflect Perceived Rhythm?’, William Barry, Bistra Andreeva and Jacques Koreman present normalized syllabic PVI analyses (against percent vowel duration) of readings of nursery rhymes in different metres and of prose utterances with varied accentual patterns, produced in textual and iterative ‘dada’ fashion in Bulgarian, English and German. The PVI measure turns out to be sensitive to metrical differences but not to the temporal distribution of accents in utterances with the same metrical structure. The analysis of the production data was followed by a perception experiment to examine the relative perceptual weight of four physical parameters correlated with the perception of prominence: F_0 , duration, vowel quality and energy. Pairs of 8-syllable trochaic lines were constructed from an excerpted foot of an English ‘dada’ rendition of a poetic stanza, manipulated in the values of one, two or all four parameters, and concatenated to the 8-syllable sequences. The pairs were presented to Bulgarian, English and German listeners for judgement as to which sequence was more strongly rhythmical. The main finding is that perceived rhythm is not just a product of durational structuring, though it occupies a high position in the rhythm index hierarchy. Change in F_0 is also a strong factor. The important aspects of this new data analysis in rhythm research are: (1) complementing production by perception experiments, (2) collecting production data from nursery rhymes, (3) calculating syllabic nPVI, and (4) including other physical parameters besides duration, particularly F_0 , as contributors to the perception of prominence profiles in rhythm judgements. So, this is an example of the new paradigm of rhythm research set out in the preceding papers.

Oliver Niebuhr’s paper ‘Fundamental Frequency-Based Rhythm Effects on the Perception of Local Syllable Prominence’ presents data from a set of perceptual experiments to show the influence of rhythmic patterning in a stimulus context on the decoding of target items that are differentiated by lexical stress. There are bisyllabic verb forms in German that are distinguished as either a verbal compound, with a separable particle, or a prefix verb, by means of first- or second-syllable stress, respectively. The two forms signal different lexical semantics. Such verb forms were joined to preceding or following trochaic or iambic syllable sequences in sentence frames of German words, resulting in meaningful everyday German utterances in which either lexical meaning of the verb forms was possible. The target verb form was given an F_0 rising-falling peak contour spanning the two syllables, with the peak maximum either in the first half of the vowel in the first or the second syllable to create clear strong-weak or weak-strong prominence profiles, or at the second vowel onset for an ambiguous

prominence profile. In all cases the first vowel had predominantly rising, the second falling pitch. The rhythmic patterns of the context section were created by rising-falling peak contours or by just rising or falling F_0 patterns on the prominent syllables. The hypotheses behind this stimulus construction were: (1) the rhythmic structure of the context section – trochaic or iambic – guides the perception of the prominence profiles on the target section and thus the decoding of one or the other verb form; (2) this effect is strongest for the ambivalent F_0 peak position on the target; (3) it is larger when the context precedes, but also present when it follows, and (4) the repetitive rising or falling F_0 pattern on prominent syllables of the context section guides the listener to the syllable of the target section that also contains either a rise or a fall, thus favouring the decoding of one or other verb form. All the hypotheses were confirmed by the data, thus lending support to the theoretical postulate that rhythm has a guiding function in speech communication because it creates expectancies and that speech rhythm is not just in acoustic parameters of the signal but is also a global top-down construction process. So, this paper provides first experimental evidence for a new functional approach to speech rhythm.

In their paper ‘On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence’, Oded Ghitza and Steven Greenberg take the guiding function in the rhythm of speech one step further to a higher processing level. They discuss data from word intelligibility experiments with semantically unpredictable English sentences, using psychophysical methods, to ascertain whether brain rhythmicity plays a role in the decoding of speech. TTS-generated speech stimuli were time-compressed by a factor of 3, resulting in <50% word intelligibility. The time-compressed waveforms were then segmented into 40-ms fragments, each followed by a silent interval. The length of the silent interval was varied in the range 0–160 ms, and the successive lengths of silence were either periodic or aperiodic. The word error rate as a function of mean silence duration was U-shaped, improving over a range of 20–120 ms. The highest intelligibility occurred for 80-ms silences inserted periodically. At this mean duration, the word error rate was lower for periodic than for aperiodic insertion. This intelligibility peak may be seen in relation to wave-form energy fluctuations in the range of 6–12 Hz, which is similar to the syllabic rate of natural speech and which corresponds to the frequency range of theta brain rhythms. The authors interpret their behavioural findings as reflecting a decoding process of speech that exploits an underlying synchronization mechanism with theta brain oscillations as a pacemaker and thus shows the best decoding performance when the energy fluctuation of the external stimulus coincides with the internal brain oscillation in frequency and periodicity. The reference of speech rhythm to optimal synchronization with brain oscillations as a guiding function for the listener gives the concept of rhythm oscillators a physical grounding, although the behavioural context of these experimental data needs to be supplemented by the physiological context of brain-imaging experiments with comparable speech material. The question then also needs to be asked whether there are cortical oscillators for the production of rhythmicity in speech. This leads us back to the questions raised in the first paper of this Thematic Issue to form a full circle for a new paradigm of rhythm research.

It is hoped that this *Phonetica* publication will kindle renewed interest in prosodic form and communicative function of speech rhythm with a new, and hopefully successful, theoretical and methodological perspective. On behalf of the journal, I would like to thank, in the first instance, the organizers of the Workshop for launching and running

a meeting dedicated to speech rhythm, for inviting the authors of the first four papers to present their work in London, and for providing a platform for the selection of papers for a Thematic Issue in *Phonetica*. My thanks then go to the four reviewers from the Editorial Board, William Barry, Nina Grønnum, Vincent van Heuven, Yi Xu, as well as to the four reviewers from outside, Ilse Lehiste, Sieb Nooteboom, Hartmut Pfitzinger, Rachel Smith, all experts in the field of speech timing and rhythm, who invested a great deal of time, effort and expertise in the reviewing and in the compilation of this special issue. And last, but not least, I thank the authors for submitting and revising their papers on the reviewers' suggestions. May we soon be able to say, in a reversal and adaptation of Manfred Schroeder's remark: 'Rhythm research is not withering in the phonological laboratory, on the contrary, it is blooming in the field of communication research.'

Kiel, February 2009
Klaus J. Kohler, Editor

References

- Barbosa, P.: How prosodic variability can be handled by a dynamical speech rhythm model. Proc. 16th ICPhS, Saarbrücken 2007, pp. 331–336.
- Bell-Berti, F.: Comments on 'Some observations on the temporal organisation and rhythm of speech'. Proc. 12th ICPhS, Aix-en-Provence 1991, vol. 1, pp. 238–241.
- Brady, M.C.; Port, R.: Quantifying vowel-onset periodicity in Japanese. Proc. 16th ICPhS, Saarbrücken 2007, pp. 337–342.
- Carlson, R.: Duration models in use. Proc. 12th ICPhS, Aix-en-Provence 1991, vol. 1, pp. 243–246.
- Classe, A.: The rhythm of English prose (Blackwell, Oxford 1939).
- Donovan, A.; Darwin, C.J.: The perceived rhythm of speech. Proc. 9th ICPhS, Copenhagen 1979, vol. II, pp. 268–274.
- Fant, G.: Units of temporal organization: Stress groups versus syllables and words. Proc. 12th ICPhS, Aix-en-Provence 1991, vol. 1, pp. 247–250.
- Fant, G.; Kruckenberg, A.; Nord, L.: Temporal organization and rhythm in Swedish. Proc. 12th ICPhS, Aix-en-Provence 1991, vol. 1, pp. 251–256.
- Fujisaki, H.; Higuchi, N.: Temporal organization of segmental features in Japanese disyllables. Proc. 9th ICPhS, Copenhagen 1979, vol. II, pp. 275–282.
- Huggins, A.F.: Some effects on intelligibility of inappropriate temporal relations within speech units. Proc. 9th ICPhS, Copenhagen 1979, vol. II, pp. 283–289.
- Keller, E.: Waves, beats and expectancy. Proc. 16th ICPhS, Saarbrücken 2007, pp. 355–360.
- Keller, E.; Port, R.: Speech timing: approaches to speech rhythm. Proc. 16th ICPhS, Saarbrücken 2007, pp. 327–329.
- Klatt, D.H.: Synthesis by rule of segmental durations in English sentences. Proc. 9th ICPhS, Copenhagen 1979, vol. II, pp. 290–297.
- Kohler, K.J.: Isochrony, units of rhythmic organization and speech rate. Proc. 12th ICPhS, Aix-en-Provence 1991, vol. 1, pp. 257–261.
- Kohler, K.J.: Domains of temporal control in speech and language: From utterance to segment. Proc. 15th ICPhS, Barcelona 2003, pp. 7–10.
- Lehiste, I.: Temporal relations within speech units: Summary of Moderator's introduction. Proc. 9th ICPhS, Copenhagen 1979, vol. II, pp. 241–244.
- Nooteboom, S.: Some observations on the temporal organisation and rhythm of speech. Proc. 12th ICPhS, Aix-en-Provence 1991, vol. 1, pp. 228–237.
- Pfitzinger, H.: *Phonetische Analyse der Sprechgeschwindigkeit*; PhD Diss., München, Forschungsber. Inst. f. Phonetik und sprachliche Kommunikation der Univ. München (FIPKM), 38, 2 (2001).
- Pierce, J.R.: Whither speech recognition? J. acoust. Soc. Am. 46: 1049–1051 (1969).
- Pierce, J.R.: Whither speech recognition? II. J. acoust. Soc. Am. 47: 1616–1617 (1970).
- Sagisaka, Y.: Modeling and perception of temporal characteristics in speech. Proc. 15th ICPhS, Barcelona 2003, pp. 1–6.
- Schroeder, M. (ed.): Speech and speaker recognition. *Bibl. Phonetica* 12 (Karger, Basel 1985).