ISOCHRONY, UNITS OF RHYTHMIC ORGANIZATION AND SPEECH RATE

K. J. Kohler

Institut für Phonetik und digitale Sprachverarbeitung
Kiel, Germany

ABSTRACT

The questions of isochrony, units of rhythmic organization and speech rate are discussed with regard to Sieb Nooteboom's keynote paper in this semi-plenary session.

1. GENERAL REMARKS

There is no doubt that, for an adequate assessment of any temporal effect in speech, we have to take a multiplicity of factors at different levels as well as their interactions into account, that we need to devise and test quantitative models to cope with the data, that statistical analyses of connected spech and of well controlled laboratory experiments complement each other, and that we need a better empirical foundation of the distinction between spontaneous and prepared speech. I also fully subscribe to the importance of phonetic explanation of the mechanisms responsible for speech timing beside the simple description of observable regularities. What I am going to take issue with concerns the topics of isochrony, of units of rhythmic organization and of the relationship between vowel reduction and increased speech rate.

2. ISOCHRONY IN SPEECH PRODUCTION

The strong isochrony hypothesis has been disproved. It has been shown for German [5] that with increasing articulatory complexity and number of syllables in rhythmical sequences of identical nonsense syllables compression to isochronous feet becomes less and less feasible because of the time constraints of articulatory movements; and even if compression is possible it results in the perception of increased spech rate in the case of achieved isochrony. But the lack of compression, i.e. the proportional expansion, also results in a change of overall tempo, this time a decrease. So, in order to stay within the same perceived rate of delivery the speaker has to compress, but this compression must not reach isochrony in this type of logatome syllable chains with unreduced vowels and consonant(s) (clusters).

The use of more natural syllable strings, which not only conformed to the phonotactics of syllables, but also to the rules of syllable chaining in German by selecting reduced vowels in unstressed positions of nonsense words, showed two complementary timing effects [5]:

(a) Disyllabic and monosyllabic feet of the same stressed syllable complexity (vowel quantity, consonant clusters) and within the same speech tempo have duration ranges that tend not to be statistically different, due to a shortening of the stressed syllable before /ə/, whereas polysyllabic feet, although also showing stressed-syllable compression, did not reach complete isochrony.

(b) The comparison of long vs.

short stressed syllables in 2- or 3-syllable feet ("Pahne" vs. "Pinne" or "Pahnige" vs. "Pinnige") yielded a complementary adjustment of the durations of the reduced unstressed vowels.

These data thus support a weak isochrony hypothesis at least for German, and other so-called stress-timed languages, e.g., English and Dutch, may be supposed to behave likewise. There is a tendency to compress as the number of syllables within the same frames increases, but this compression quickly approaches a ceiling when the number of syllables exceeds two. On the other hand, there is also the tendency to vary the durations of reduced unstressed vowels in opposition to the preceding stressed vowel, being a complementary aspect of a tendency to foot isochrony.

These data can, at first sight, also be referred to the word level because word and foot coincided in these experiments. But why should there be a tendency to make mono- and bisyllabic words of the same length by stressed-syllable compression and unstressed syllable compensation? There is nothing in the linguistic category of the word that could determine such a behaviour, whereas a superimposed rhythmic principle can easily explain it and a number of other phenomena:

(1) The ordering in German "mit Pfeil und Bogen" as against English "with bow and arrow" is not semantically, but rhythmically conditioned: the mono- rather than the disyllabic noun is put before the conjuction to get a more even sequence of foot durations than would be the case with the reversed order, and this grouping cuts across word boundaries.

(2) Articulatory reduction is at work irrespective of words and word boundaries [6]. Words may disappear altogether, and they may be treated as syllabic appendices to preceding words, even bridging phrase structure boundaries, e.g. in "Hast du einen Moment Zeit?" (*Have you got a moment to spare?*) ['haspm momɛn 'tsaɪt]. The deletion of [ə] in [dənən], derived from "du einen", follows the [ə] elision in "die geschnittenen Rosen" (*the cut roses*), although there is a phrasal boundary between "hast du" and "einen Moment". The reduction can go further to ['has məmɛn 'tsaɪt], where "du" and "einen" have disappeared from the phonetic surface, and the prestress syllable [mo] is also reduced. Finally, ['has mmɛn 'tsaɪt] can result, showing a further reduction of the unstressed part of the content word "Moment". All these processes are in keeping with the rhythmic principle to make feet as equal in duration as possible. Function words are obvious candidates to assist in this compression because they are unstressed in the unmarked case and signal redundantly coded syntactic functions rather than lexical meaning, but unstressed syllables of content words undergo the same reductions.

(3) In verse, the rhythmic principle is regularised as in / Humpty / Dumpty / sat on a / wall / / Humpty / Dumpty / had a great / fall / / All the king's / horses and / all the king's / men / / Couldn't put / Humpty to/gether a/gain / This is only possible because there is an underlying rhythmic principle in speech, which triggers the tendency towards isochrony independent of the chaining of words.

Nooteboom refers to the Swedish data from read speech in Fant and Kruckenberg [2] in support of his dismissal of isochrony as a factor in speech production. But continuous texts, i.e. accidental corpus rather than systematic experiment-

al materials can neither prove nor disprove such a rhythmic principle because in connected speech a great number of timing factors operate, and they may easily override tendencies towards isochrony, as was demonstrated in Kohler [4]. The reference to Fant and Kruckenberg actually runs counter to Nooteboom's statement that "statistical studies on corpuses of connected speech obscure real regularities: there remains a need for testing specific ideas with well controlled materials in laboratory experiments". It was precisely this methodological prerequisite that determined the experimental design followed in the Kiel studies of speech timing, which devised language materials and data collection procedures in a hypothesis-driven fashion to systematically test and possibly reject the isochrony assumption [5]. But the reference is also at odds with Nooteboom's statement that "the systematic effects on speech sound durations of anyone particular factor can only reliably be assessed when we take the effects of many other factors, on different levels of speech processing, into account". In continuous, ad hoc texts the many different factors and their interactions cannot be reliably separated.

3. WORDS OR STRESS GROUPS AS UNITS OF TEMPORAL ORGANISATION?

Nooteboom provides a very clear and categorical answer to this question: "Words are important units for temporal organization of speech, stress groups are not." And he adduces four reasons:

(a) Speech pauses always occur at word boundaries, never at stress group boundaries that do not accidentally coincide with word boundaries, and boundary phonemes of emphasized or informative words, not boundary phonemes of emphasized or informative stress groups tend to show increased duration and reduced coarticulation.

It is very important to define what is meant by pauses: pauses for syntactic and semantic structuring are usually placed at word boundaries, hesitation pauses can be anywhere in the syllable chain, and pauses for emphasis can also be inside words before the stressed syllable, as in "I warn you: don't for ... get." (compare this with "I warn you: don't forfucking-get."). Even if it could be argued that in this case the pause or an inserted swear word occurs after a stripped-off prefix and therefore at a linguistic not a rhythmic boundary, examples such as "po ... tato and to ... mato" refute this as well. And in verse a slowing down can produce pauses at any stress group boundary irrespective of word boundaries. Thus "Humpty Dumpty" can be read with pauses at all the foot boundaries including "to/gether" and "a/gain". Instead of having pauses at foot boundaries the boundary phonemes in all the examples quoted may be lengthened and detached from their environment for special emphasis. It is the stressed syllable that gets the extra prominence either in order to highlight the word it is in for semantic reasons, or the foot it is in, for rhythmic reasons.

(b) Beside phrase-final lengthening, there is word-final lengthening, which could not be explained in terms of isochronous intervals.

This finding does not justify the exclusion of the stress group as a timing unit. Of course, it is possible to have word-induced duration control, as was shown for German "eine gezeigt" ([*has*] *shown one*) vs. "einige zeigen" ([*will*] *show some*) in Kohler [3, 4]. This is due to the content structuring

of speech, but its occurrence does not cancel out rhythmic structuring; on the contrary, the latter can obliterate word-related timing as the same investigations demonstrate.

(c) Nooteboom refers to unpublished data by van Santen that show considerable duration effects of the number of syllables and the stressed-syllable position in words, but not in stress groups.

The data in Kohler [3, 4] point to effects in both units.

(d) A principle of economy compels us not to introduce more units than are necessary to account for our data, and there are no publications where it has been convincingly shown that data cannot be explained without recourse to stress groups.

First of all, economy is certainly a useful criterion in phonetic data description, but when it comes to explaining the observed phenomena we are bound by what can give us the deepest insight into the widest possible empirical domain, and economy is secondary to this consideration because why should speech timing, or any other phonetic event in human production and perception processes, be entirely governed by economy. Secondly, the occurrence of greater compression in German "eine gezeigt" ([*has*] *shown one*), as against "eine zeigen" ([*will*] *show one*), in some recorded data [3], points to a timing factor that cannot be equated with the word as the only relevant rhythmical unit.

Furthermore, van Dommelen has shown [1] that in a falling, as against a level, F0 contour, combined with vowels from a duration continuum spanning the quantity opposition /a/ vs. /a:/ in German, the perceptual quantity switch occurs at greater vowel durations, provided the syllables are embedded in a rhythmic sequence, irrespective of word boundaries, e.g. "Er hat As ['as] / Aas ['a:s], Assen ['asən] / aßen ['a:sən], Masse ['masə] / Maße ['ma:sə] verstanden." (*He understood ace/carrion, aces (dat.) / (they) ate, mass/ measurements.*) This result is replicated when the stimuli of the disyllabic word pairs are presented in isolation, but it is, in this position, reversed to shorter vowel durations for the monosyllabic pair, in keeping with Lehiste's findings [7]. What is important here is simply the presence or absence of a following unstressed syllable without reference to word division. A possible explanation for these opposed effects can be sought in a perceptual syllable lengthening in falling F0, which changes the rhythmic patterning of syllable sequences and thus the speech rate; the quantity assessment of the physical vowel duration then occurs against a slower tempo frame and therefore appears shorter. If there is no rhythmic frame surrounding the test syllable, especially no following unstressed syllables, there is no independent tempo assessment, and the perceived lengthening affects the vowel directly.

The answer to the question of rhythmical units in speech should not be an "either or", but a "both and". Words are certainly important units for the temporal organisation of speech, but stress groups are as well, and the two interact. In verse the rhythmic principle dominates, in continuous connected, spontaneous speech the word (content) aspect gets more prevalent, but the rhythmic principle never disappears. Just as words have to be put into a segmental frame, so they also have to be fitted into a rhythmic one, and both segments and timing are affected by the content structur-

ing of utterances.

4. VOWEL REDUCTION AND INCREASED SPEECH RATE

Nooteboom rules out that vowel shortening due to a higher speech tempo leads to vowel reduction, by referring to data in van Son and Pols [9]. But these data were obtained from a "single highly experienced professional speaker", and I therefore think that such a categorical exclusion of vowel reduction in increased speech rate of spontaneous speech is unjustified. What is essential here is that, given the need of speakers to be understood under different speech production conditions, phonetic variation, including vowel spectra in different tempo frames, can be located along a hyper-hypo scale to guarantee sufficient discriminability with as little effort as is necessary in the particular communicative situation [8]. So, speakers can execute precise movements to reach targets irrespective of speech rate if they put in the necessary effort to achieve increased discriminability for listeners, but they may also slur if they think the effort is not worthwhile, and it is the latter attitude that eventually results in language change.

5. REFERENCES

[1] VAN DOMMELEN, W. A. (1991), "F0 and the perception of duration", *Proc. XIIth Intern. Congr. of Phonetic Sciences*

[2] FANT, G. and KRUCKENBERG, A. (1989), "Preliminaries to the study of Swedish prose reading style", *STL, QPR*,2, 1-83.

[3] KOHLER, K. J. (1983), "Prosodic boundary signals in German", *Phonetica*, 40, 89-134.

[4] KOHLER, K. J. (1984), "Temporal control at the utterance level in German"; in VAN DEN BROECKE and COHEN (eds.) "*Proc. of the Xth Inter. Congr. of Phonetic Sciences*"; pp. 197-200, Dordrecht: Foris Publications.

[5] KOHLER, K. J. (1986), "Invariance and variability in speech timing: from utterance to segment in German", in PERKELL and KLATT (eds.) "*Invariance and Variability in Speech Processes*"; pp. 268-289, Hillsdale, New Jersey, Lawrence Erlbaum.

[6] KOHLER, K. J. (1990), "Segmental reduction in connected speech in German: phonological facts and phonetic explanations"; in HARDCASTLE and MARCHAL (eds.) "*Speech Production and Speech Modelling*", pp. 69-92, Dordrecht/Boston/London: Kluwer Academic Publishers.

[7] LEHISTE, I. (1976), "Influence of fundamental frequency on the perception of duration", *Journal of Phonetics*, 4, 113-117.

[8] LINDBLOM, B. (1990), "Explaining phonetic variation: a sketch of the H & H theory", in HARDCASTLE and MARCHAL (eds.) "*Speech Production and Speech Modelling*", pp. 403-439, Dordrecht/Boston/London: Kluwer Academic Publishers.

[9] VAN SON, R. J. J. H. and POLS, L. C. W. (1989), "Comparing formant movements in fast and normal rate speech", in TUBACH and MARIANI (eds.) "*Eurospeech 89*", *vol. 2*, pp. 665-668 Edinburgh: CEP Consultants.