

LEARNING
KEEPING
AND
USING

LANGUAGE

Offprint

EDITED BY
M.A.K. HALLIDAY
JOHN GIBBONS
HOWARD NICHOLAS

Isochrony and 'Uncomfortable Moments' in Conversation

Peter Auer, Elizabeth Couper-Kuhlen and Aldo Di Luzio

In the years past, the old debate over the rhythmic organisation of language has been opened again on the level of conversational interaction; in particular, in their important book *The Counselor as Gatekeeper*, Erickson and Shultz (1982) argue that making a conversation work also implies finding and sustaining a common rhythm ("being in phase"). Our paper is largely in this tradition and treats rhythm, as Erickson and Schultz do, as a means of contextualizing language (in Gumperz' sense of this word). However, we think that Erickson and Shultz' hypothesis is too strong to be validated empirically, if rhythmic synchronization is defined in a more precise way; we will therefore put forward the weaker hypothesis that rhythmic synchronization in the sense of "isochrony" is achieved to varying degrees in ordinary conversation, and that varying its degree is a contextualization device that can be used by participants for local purposes 'below' the level of sustaining the interaction as such. We will look at some data in which one such local purpose is at stake, i.e. the conversational organisation of 'uncomfortable moments' where at least one participant's "face" is endangered.

The notion of isochrony is an old one, but it was Pike who formulated it in its most influential version in modern linguistics. In "The Intonation of American English", he states that in a language such as English, (1) the intervals between the prominent syllables are "approximately" even, (2) as a consequence, the duration of the individual syllables is a function of how many syllables occur in the "foot", and (3) other languages, such as Spanish, "may" have a different rhythmic organization, in which the syllable is the basic unit, not the foot.

This is an offprint from:

M.A.K. HALLIDAY, JOHN GIBBONS, and HOWARD NICHOLAS (eds)

Learning, keeping, and using language:

Selected papers from the Eight World Congress of Applied Linguistics

Sydney, 16-21 August 1987

John Benjamins Publishing Co.

Amsterdam/Philadelphia

1990

ISBN 1-55619-100-6 (set)/1-55619-104-9 (vol.I)/1-55619-105-7 (vol.II) (U.S.)

ISBN 90 272 2071 9 (set)/90 272 2073 5 (vol.I)/90 272 2074 3 (vol.II) (Eur.)

© Copyright 1990 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm or any other means, without written permission from the publisher.

Since then, most attempts at measuring isochrony in phonetic terms have failed.¹ In order to explain this failure, and to rescue Pike's hypothesis, it has been argued that the measurement techniques used were inadequate (cf. the theory of the "perceptual centre", Morton, Marcus and Frankish 1976; Fowler 1979; Hoequist 1983), or the data wrongly selected (e.g. because isochrony is only possible inside the intonational phrase, cf. Rees 1975, or inside the grammatical sentence, cf. Lehiste 1977). The most important argument against the use of phonetic data, however, is probably that rhythmic isochrony may be a matter of perception, not of acoustics.

Despite all this, Erickson and Shultz, when working on student/counselor interaction, state that

... points of emphasis mark a pattern of regular rhythm. This underlying *resultant rhythm* or *cadence* is slightly freer than a metronomic 'beat', but it can be reliably measured by a metronome or by other sorts of chronometric instruments. (1982:75)

That is, they take isochrony to be relevant on the level of instrumental measurement. (Of course, it remains open in this statement how much deviation from the strictly metronomic rhythm the authors would allow; however, their examples — e.g. (1982:111) — show that it is small indeed, approx. 0.04 seconds = 1 frame.)

A first look at our conversational data (British and American English, German) has not confirmed Erickson and Shultz' findings of a stable isochronous rhythm as the unmarked state of affairs; instead, we have noted rather frequent changes in rhythm and speaking rate, both within and between turns, and alternations between thoroughly arrhythmic and rhythmically rather well integrated stretches of conversation. We would therefore join those who take a more skeptical position in the phonetic and phonological literature on rhythm, i.e. that isochrony cannot be taken to be the "normal" state of affairs throughout an interaction. On the other hand, we are clearly able to hear conversational passages that exhibit the type of cadence mentioned by Erickson and Shultz. Our research therefore starts with this hypothesis: that the relevant units in which isochrony can/tends to occur, or tends to be suspended, are storable in conversation analytic terms (in addition to other presumably relevant units such as the grammatical phrase or the phonological phrase or sentence). Thus, although rhythmic integration in the narrow sense of isochrony may be something conversationalists strive for (and therefore, so to speak, an end in itself), there

should be linguistic and conversational structures that make it easier or harder to reach this goal. In fact, the data Erickson and Shultz deal with in the first place, i.e. question/answer sequences in routine counseling interactions, may be just one such environment: working through this almost questionnaire type of routine might have influenced rhythmic integration in a positive way, and their impression of isochrony throughout interactions may have been caused by the dominance of this conversational pattern.²

In this paper, we will be looking at one specific conversational environment in which matters of rhythmic integration seem to play an important role (as has also been noted by Erickson and Shultz, to whom we owe the expression "uncomfortable moments"): recipients' responses to potentially face-threatening turns, and first speakers' responses to these responses. We have selected passages for discussion in which there is a clear rhythmic organisation in the first speaker's turn (a question), and we start from the assumption that on the local level of turn transition, taking up this rhythmic organisation (in the answer) would be the preferred alternative to breaking it.

Before dealing with the data, a number of preliminary issues have to be dealt with. The first issue is the question of what counts as an isochronous rhythm and how one can hear it as such. Isochrony is only one aspect of rhythm: it concerns the regularity of intervals between the phonetically prominent syllables in real time. As such, it is independent of other parameters that influence our perception of rhythm as well, such as the number of unstressed syllables that intervene between the prominent ones, or the degree of prominence of the latter. (The perception of an isochronous rhythm presupposes a two-way division between phonetically prominent and non-prominent syllables.) One must not expect isochrony to integrate all prominent syllables; neither must the beats in the rhythm be confounded with the phonetic prominences.³ An isochronous rhythm may omit prominences; on the other hand, a beat may fall on a phonetic silence ("silent beat"). Isochrony is therefore, in the first place, a matter of perception; the following transcriptions of rhythm are consequently auditory transcriptions.⁴ What makes them reliable is the fact that the authors have arrived at them independently.

The second issue is the notion of "face-threat". Face-threatening situations come up in conversation in a complicated interplay between a culturally given repertoire of potentially face-threatening topics, or attributions to persons, or activities, and the co-participants' treatment of these topics,

attributions, activities, etc. as face-threatening. In other words, no situation is in itself face-threatening to the participants, but it is turned into a face-threatening situation by them. Although, as competent members of a culture, we are able to list potential face-threats, none of the issues at stake need necessarily become face-threats in interaction. For example, in our extracts, we will be dealing with a number of attributions to persons such as 'being unemployed' which are, in the culture in question, potential face-threats; however, the important thing to look at is how these potential face-threats are turned (or not turned) into actual face-threats. We want to show that one way this can happen is via a rhythmic break, or rhythmic disintegration in general.

Thus, it is hypothesized that rhythm, or not taking up another participant's established rhythm, works as a contextualization cue: it indicates that there is, or is not, a change in the interactional 'footing' or context — without, of course, identifying the new context by name. Co-participants use their cultural knowledge to infer from such a cue that a potential face-threat inherent in what is being said has become relevant for the interaction at hand. We insist on the negotiability of face-threatening situations, and we will try to show how "uncomfortable moments" are achieved or avoided conversationally by rhythmic means. (Needless to say, one contextualization cue seldom comes alone; there is redundancy in signalling changes of context. As a consequence, rhythmic changes are often accompanied by other indicators.)

Let us start with the following three extracts, because of their parallel interactional (sequential) structure. All of them are from the opening phase of conversation between a radio DJ and a caller (Radio Piccadilly phone-in show, recorded in Manchester, 1980).

- (1) BdI: Ju.
 DJ: /hi Jùdith /
 J: /hello /
 DJ: /hello Judith /
 /hów àre you/
 J: oh /fine /
 DJ: /góod /
 J: /wèlcome to the /
 /programme what do you/
 /dó in life /
 /Júdith /

- J: /ém /
 /well I work for /
 /Bóots the /
 /chémist /
 (2) JM (B)
 DJ: / hèllò /
 /Jóhn /
 J: hèl- /
 /ló /
 DJ: hèlló hów are /
 /yòu /
 J: uh not so /
 /bàd Dàve /
 DJ: [/thánks /
 [/whát d/you do in life / (faster) /
 /Jóhn /
 J: /úh wéll I'm /
 /òff sìck at /
 /présent /
 DJ: /how long've you been off /
 [/work /
 J: [/oh: I've been òff quite /
 /while nòw /
 DJ: /háve you (.) /
continuation (no rhythmic transcription)
 DJ: any: signs of going báck to work yet
 J: [hm
 no: I don't think so [no
 DJ: [h
 ah: well how d'you pass your day?
 ((etc.))
 (3) G.M. 15.0
 DJ: /hí Gary /
 G: /hí /
 DJ: /hí /
 /hów àre you /
 G: /nòt so bad /

- /tháinks
- DJ: /góod /
/whèreabouts in/
/Bólton d'you /
/wòrk
- G: úh I /dón't I'm unem-/
/plóyed — wèll a/
/stúdent — — /
/parttime /
- DJ: a stú:dent — uh /párt time unem-/
/plóyed /
- G: /yéah /
- DJ: fíne — ókày —
/hòw long you been unem-/
/plóyed
- G: úh (1.5) èight mónth
- continuation (no rhythmic transcription):*
- DJ: mhm — as lóng as thát
- G: yes
- DJ: góodness gracious when was that then (.) tell me (.) very
quickly [I
- G: [well I'm taking an
- DJ: [A level course actually
can't remember
- ah: (.) sensible man (.) so while you're unemployed y- you're try-
ing to get yourself an A level are you?
- G: yeah
- DJ: ah great well that sounds that sounds sensible (.)
((continues with caller's plans for the future))

Passages with a consistent isochronous rhythm are written in slashes; every first syllable in a line constitutes a beat in the rhythm. Sometimes speakers accelerate the rhythm by producing the next beat a fraction of a second too early, but then stay in rhythm (cf. in (3), DJ in lines 4 and 7); in these cases, there is a rhythmic shift, but before and after it the talk is isochronous. Note how isochrony selects the proper beats out of a larger number of phonetically prominent syllables (marked by ' and `); for

instance, there is a clash between two syllables of equal prominence in extract (2) in *off sick*, but only one of them (*off*) is 'on the beat'.

On the basis of these extracts, it can be shown how a potential face-threat is turned into an interactionally uncomfortable moment by participants. Let us take for granted here that being out of work is a potential face-threat for an adult person in the north of England in 1980. Accordingly, the DJ's initial question (i.e. *what do you do in life?, whereabouts in Bolton do you work?*) is a crucial one for the callers in extracts (2) and (3), but not for the caller in (1), who has a job as a shop assistant. However, although extracts (2) and (3) group together on the level of content, we argue that (1) and (2) contrast with (3), because only the latter leads into what can be termed an uncomfortable situation, whereas both in (1) and (2) participants fail to display any orientation whatsoever to an interactionally relevant face-threat. This means that in (1), there is no potential or actual face-threat, in (3) both a potential and an actual face-threat, and in (2) a potential but not an actual face-threat for the caller.

What is responsible for this is the smooth rhythmical development in (1) and (2) as compared to (3). In both cases, there is rhythmic continuity in the transition between the DJ's question and the caller's response; in particular, the caller in (2) answers in the rhythmical pattern initiated by the DJ's *what d'you do in life John* by picking up the prominences on *what* and on *John*, and by producing his utterance in such a way as to make sure that the prominences on *uh* and on *sick* are isochronous with *what* and *John*. (The same holds for extract (1) where *em, well, Boots* are in an isochronous rhythmic pattern with preceding *do* and *Judith*.)

In contrast, the caller's response in (3), *uh I don't I'm unemployed etc.* is, although rhythmical in itself, clearly out of the isochronous pattern established by the DJ in his preceding question (*whereabouts in Bolton d'you work?*). The utterance following the question breaks rhythmic continuity, just as in (2) it produces continuity. Things do not get any better when the DJ reformulates (or maybe, misformulates) the caller's response (*a student (.) part-time unemployed*); this utterance is out of rhythm with what precedes and temporarily comes to a halt, until a new rhythm is proffered with *part-* and *-ployed* which G picks up and ratifies with *yeah*. But in the following exchange, when the DJ tries to resume an interactional rhythm (*how long you been unemployed* contextualizes a new sub-topic by a marked increase in speaking rate, judged by number of syllables in real

time and between prominences), G again fails to take up this pattern, for his *uh* is not on the beat established by the DJ's *how* and *ploy-*. Thus, rhythm is repeatedly thrown off and this 'coming out of phase' contextualizes the situation as it turns into an uncomfortable one.

Now how can we show that (3) is indeed taken to be more face-threatening than (2)? The weaker evidence for this is to be found in the opinion of external observers, who we assume — to the extent that they are members of the same speech community — will be sensitive to the awkwardness (or smoothness) of a participant's reaction to a potential face-threatening utterance. To substantiate this assumption we designed a small pilot study in which we asked a group of 18 native and near-native speakers to rate these situations (and others involving potential face-threats) according to degree of awkwardness or uncomfortableness. Their reactions point clearly in the direction of our interpretation: for example, on a 'Scale of Awkwardness' ranging from 1 (*Very*) to 7 (*Not very*), excerpt (2) — the rhythmically integrated one — averaged 6 (=quite smooth), whereas excerpt (3) — the rhythmically non-integrated one — averaged 3.8 (=more awkward than smooth). Thus there is initial external evidence that rhythmically integrated responses to potential face-threats are less awkward and produce less uncomfortableness in 'eavesdropping' hearers than rhythmically non-integrated ones. But obviously, we cannot exclude the possibility that matters of content influenced the judgements.

More subtle, but also, in a way, stronger evidence for our claim that (3), but not (2) is a face-threatening situation, can be taken from the conversational development of the two stretches of talk themselves, which shows participants' orientation to the 'uncomfortableness' of the latter. As for (2), we observe a very smooth development of the call's "first topic" (i.e., the caller's unemployment). The DJ expands this topic by a number of questions directly addressing the issue (*how long've you been off work, any signs of going back to work, how d'you pass your day*), and although the answers to the first two questions both present a rather gloomy picture of the caller's professional situation, the way they are dealt with by both parties does not disrupt the flow of speech.

Things are quite different in extract (3). There is a good deal of conversational work both parties invest in smoothing out an awkward situation; and doing so, they demonstrate to the analyst their orientation to this awkwardness. Note that the caller provides two categorizations for this present status in the answer to the DJ's initial question: (a) that he is unemployed

and (b) that he is a student. The second one may be seen as the caller's preferred alternative (obviously, the less face-threatening one). However, the DJ chooses the first one for the topical continuation of the conversation (*how long you been unemployed*) and then attempts to shift away from this unpleasant subject (*when was that ...*). Thus, the DJ's orientation to the awkwardness of the situation may be inferred from his eagerness to move to a different topic. At the same time, the caller forces a different topical development, insisting on the second categorization offered above, i.e. being a student; only now does the DJ recognize the positive face value of this categorization for the caller and provide the appropriate feedback. The caller's orientation to the awkwardness of the situation can be inferred from his insisting on the categorization as a student against (and in overlap with) the DJ's topic-shifting question — a rather rare instance of a caller rejecting the DJ's dominant conversational position in the radio show.

Note that all three of the extracts discussed so far have in common the fact that the caller prefaces his/her answer with *uh*, *eh* or *em*. These have been referred to as 'fillers' or 'floorholders' in the literature and they are usually put in the category of delaying devices associated with dispreferred turn-shapes. However, a close analysis of rhythm at transition places where these fillers occur shows that some [as in (1) and (2)] are clearly integrated into the overall rhythmic pattern established by the first speaker and continued by the second speaker, whereas others [as in (3)] are not integrated, i.e. they throw the established rhythm off (although they may serve to set up a new rhythmic pattern peculiar to the second speaker). This suggests that it is not merely the presence of such fillers which contextualizes an actualized face-threat but more importantly whether they are rhythmically integrated or not. Where the filler causes no rhythmic break in the overall pattern established and continued, there is no actualization of a potential face-threat, although on the lexical level there may be an ostensible delay.

The hypothesis that rhythmic integration vs. non-integration is ultimately more significant for the actualization of a face-threat than lexical content is substantiated by the following extract:

- (4)
- | | | |
|----|-----------------------|---|
| D: | /while I talk to Mrs/ | |
| | /Wagstaff | / |
| | /hello sir | / |
| W: | [/madam | / |
| | /he — | / |

D: /hello Dick /
 /that was a clanger /
 /wasn't it /
 W: it /
 /was .hhh /
 D: /^ /
 /go on then /
 W: / I just wanted to: /
 D: /have you forgiven me/ /
 W: /yes I have /
 D: /oh good /

Note that W's immediate agreement to D's *that was a clanger* is perfectly integrated into the overall rhythmic pattern. Although agreement with a coparticipant's self-denigration is typically a highly dispreferred activity, here the rhythmic configuration of the turn prevents its actualization as a real face-threat, an interpretation which is borne out not only by our informants' judgments [this excerpt received an average of 5.8 (=quite smooth) on the Awkwardness Scale] but also by the fact that D subsequently has no trouble maintaining the established rhythm, although laughter intervenes. In other words, put in everyday language, D handles this situation very very smoothly. Smooth management of a potentially face-threatening situation is, we maintain, essentially a rhythmic matter.

The preference in verbal interaction for smoothness or, rhythmically speaking, for beat coordination is obvious not only in cases where potential face-threats are averted via rhythmic integration but also in cases where there is an ultimate rhythmic break-down — but only *after a coordinated start*. Returning to our 'what do you do in life' examples, consider, for instance:

(5)
 DJ: /tell us what you/
 /do in life /
 /Vera /
 C: /eh /
 /well I'm /
 /just e:m you know a wedded --
 /housewife /
 /^ put it /
 /that way /

Here Vera is obviously so 'tuned in' to the DJ's rhythm that her initial fillers coincide precisely with the beats which we would expect based on the pattern established. However, this rhythmic coordination disappears as her turn progresses. The arhythmic configuration of her search for an appropriate way to construct an answer to the DJ's question contextualizes the fact that this is, at least for Vera, a face-threat.

In sum, as excerpts (1), (2), and (3) illustrate, it is not *em*'s and *uh*'s which contextualize the actualization of a face-threat but their rhythmic non-integration. Across QA sequences 'keeping the beat' appears to be a preferred activity, as evidenced by the fact that it happens spontaneously, before the (dispreferred) arhythmia of word search (excerpt 5) sets in. The contextualizing power of rhythmic isochrony vs. anisochrony is reflected in the fact that even objectively dispreferred responses, which under normal circumstances potentialize significant face-threats, can be neutralized by rhythmic integration (excerpt 4). We hope to have shown with these examples that rhythm in conversation is a powerful means of contextualization, whose intricacies are just beginning to be understood.

We would like to finish with another extract from our phone-in show in which a participant's face is at stake (in this case, it is the DJ's face, for a change), and in which this fact is contextualized via rhythmic disintegration; it is also an extract that may be of special interest at a conference of linguists, for it deals with linguistic problems on the content level:

(6) Bd I, Nora I
 ((DJ reads out the riddle — "brainteaser"))
 DJ: you can
 /find rreference in /
 /ány Látin /
 /díctionary to a bri-/
 /gáde /
 N: is it /honee swa key /
 /mal ee pents /
 / / /
 / / /
 DJ: pardon?
 N: /onee / (new rhythm, slowly)
 /swa key/
 /mal ee pents /
 DJ: /good job / (new rhythm, faster)

N: /Sweeney's not on [the/
 [hh/
 DJ: /radio wi' /
 N: /this:
 [he hh
 continuation (no rhythmic transcription):
 DJ: well whatever you said no= (slowly)
 N: it's the motto of the guards brigade
 DJ: oh is it?
 N: hha=
 DJ: ah well never being in the guards of course I wouldn't know a
 thing like this hm no it's not that [but eh
 N: [well I thought that was Latin
 'n' it was
 he [he h h
 DJ: [well no doubt it is: =
 N: [yeah
 DJ: [I don't dispute that for one moment Nora but [eh
 N: [hm hm=
 DJ: =certainly not the answer we're looking for but a good answer
 N: [(well)
 [nice to see er (.) a little bit o' thought went into that (.)

Notes

1. Cf., among others, Shen and Peterson (1962), O'Connor (1965), Bolinger (1965).
2. It will also be noted that face-to-face data and the telephone data we have primarily considered so far may not be comparable in all respects.
3. The issue will be dealt with in more detail in future phases of our work. The relationship between rhythmical beats (in the sense of *arsis* and *thesis* in poetics) on the one hand, and phonetic prominence on the other, is a very important issue, but it has not been investigated in modern phonology in any detail.
4. This is not to say that instrumental measurements are of no interest; they are, on the contrary, on our current agenda. The problem is that there are most probably systematic biases in the perception of rhythm due to the monitoring of the articulation of the utterance in question which allow us to hear actually delayed or anticipated prominences as being on the beat nevertheless. One has to know exactly what these biases are in order to evaluate instrumental data.

References

- Bolinger, D.W. 1965. "Pitch Accent and Sentence Rhythm". *Forms of English: Accent, morpheme, order* ed. by I. Abe & T. Kanekiyo. Cambridge: Harvard University Press.
 Classe, A. 1939. *The Rhythm of English Prose*. Oxford: Basil Blackwell.
 Erickson, F. & J. Shultz. 1982. *The Counselor as Gatekeeper. Social interaction in interviews*. New York: Academic Press.
 Fowler, C.A. 1979. "'Perceptual Centers' in Speech Production and Perception". *Percept. Psychophys.* 25. 375-388.
 Hoequist, E. Jr. 1983. "The Perceptual Center and Rhythm Categories". *Language and Speech* 26. 367-376.
 Lehiste, I. 1977. "Isochrony Reconsidered". *Journal of Phonetics* 5. 253-263.
 Morton, J., S. Marcus & C. Frankish. 1976. "Perceptual Centers". *Psychological Review* 83. 405-408.
 O'Connor, J.D. 1965. "The Perception of Time Intervals". *Phonetics Laboratory Progress Report 2*, University College London. 11-15.
 Pike, K. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Press.
 Rees, M. 1975. "The Domain of Isochrony". *Work in Progress* 8, Department of Linguistics, Edinburgh University. 14-28.
 Shen, Y. & G.G. Peterson. 1962. "Isochronism in English". *University of Buffalo Studies in Linguistics, Occasional Papers* 9. 1-36.