AN APPRAISAL OF RHYTHM AS A COORDINATOR OF TURN-TAKING

M. Bull, matthew@ling.ed.ac.uk
Department of Linguistics, University of Edinburgh, Scotland

ABSTRACT

This paper investigates the notion that perceptual isochrony may be used by participants in a dialogue as a method of timing and coordinating turn-taking. Results from the two experiments reported here provide no evidence for this notion. Instead it would appear that isochronic processes are at best heavily masked by pragmatic and cultural factors, and some form of linear cognitive processing.

INTRODUCTION

It has long been recognised that speech possesses a certain degree of rhythmicity. However, the extent, source and uses of this rhythmicity have over the years been subject to a great deal of debate. Earlier research [1, 8, 11] suggested that relatively strict productive isochrony existed. But these observations may have reflected a different and more plausible process, namely perceptual isochrony. Indeed, it is now beyond contention that listeners tend to impose a regularity on the rhythmic structure of an utterance even where no acoustically measurable regularity exists [4, 6, 7, 9]. The arguments in favour of a cognitive mechanism which organises raw acoustic data into perceptual chunks also appear not unreasonable [10].

Claims have also been made regarding the roles that isochrony might play in the interaction between speaker and hearer [5]. For example, Couper-Kuhlen has specifically cited turn-taking as a possible area for the use of perceptual isochrony. The coordination of turn-taking is extremely fine [12], and it has been suggested by Couper-Kuhlen that a purely linear model of the timing of turn-taking is insufficient to account for this. Instead, a hierarchic model is proposed, where there exists an unmarked case of turn-taking in which a hearer (H) would pick up on the rhythmic structure of a speaker's (S) utterance, such that the first relatively prominent syllable of H's turn would coincide with the rhythmic beat set up by S. Notice that this predicts that the duration of an inter-turn interval (ITI) would be a function of the perceived interval between stressed or prominent syllables in S's turn. However, Couper-Kuhlen's hypothesis remains largely speculative.

In a pilot study [3] I found that when recordings of exchanges with altered ITI's were presented to subjects, they could distinguish between ITI's which were longer or shorter than in the original recording, or of the same length. Having established that such differentiation was at least possible, the two experiments reported here were carried out to ascertain preferred ITI's. I reasoned that if the values chosen by subjects clustered about one or more points this would provide evidence that some very powerful mechanism was at work - one which could have been based on rhythmic principles.

Results from both experiments reported here do not seem to support this hypothesis, however. Instead, the picture appears to involve far more factors than can be accounted for by a rhythmic principle, and the role of isochrony in turn taking is at best one of several possible cues to the end of a turn.

EXPERIMENT I

Method

Twenty-five exchanges, consisting of a turn lasting a few seconds, a natural ITI, and a second turn lasting a few seconds, were selected from the HCRC Map Task Corpus [2], a corpus of spoken task-oriented dialogue. The exchanges chosen did not contain any major disfluencies, and appeared to involve a minimal amount of thinking time on the part of the hearer in the recording. The ITI's occurring in the original recording (hereafter Original ITI's) were well spread between 0 and about 1000ms.

The Original ITI's for each exchange were then altered, being replaced with ITI's between 0 and 1000ms, generated semi-randomly. The artificially generated ITI's consisted of low-volume 'noise', and 20ms of speech adjoining this noise were acoustically tapered to prevent a noticeable click which might otherwise have acted as an unwanted cue to the subjects.

The altered exchanges were presented to each of the subjects in a randomly generated order, over headphones. A transcript was provided to facilitate comprehension.

The subjects were instructed to listen to the presented exchange and to modify the ITI from the duration they first heard (hereafter the Start ITI) until they felt that the ITI was the length it would be in a natural discourse environment (Finish ITI). Subjects were able to make the ITI longer or shorter by pressing keys which altered the duration of the ITI either by 50ms or 150ms as often as they wanted, hearing the exchange with the altered ITI each time a key was pressed. They were also able to repeat the exchange with the unaltered ITI.

In designing the experiment it had been considered that subjects' responses might be influenced by Start ITI's. To account for this possibility, the subjects were split into two groups of fifteen, each group being given a different set of semi-randomly generated Start ITI's.

Results & Discussion

A multiple regression analysis was carried out on the data, indicating that a significant proportion of the variance in Finish ITI's could be accounted for by Start ITI's and Original ITI's (R^2 = 0.3766, F (2, 747) = 225.673, $p < 0.0001$). The results clearly show that Start ITI's had a significant influence on Finish ITI's ($\beta = 0.61$, $p < 0.001$). Original ITI's had no significant effect, however ($\beta < 0.01$, $p = 0.98$).

A comparison was also made between the choices made by the two groups of subjects. Each group was presented with exchanges which had a significantly different set of Start ITI's ($r = 0.02$, $p = 0.924$). One would have expected this difference to have been eliminated if subjects were to rely on a common rhythmic mechanism for determining ITI's. A simple regression analysis of the relationship between Finish ITI's of group A, and Finish ITI's of group B showed no significant correlation between the Finish ITI's of the two groups of subjects ($r = 0.035$, $p = 0.505$), indicating further that Start ITI's must have had a significant bearing on the results.

Finally, one pattern which emerged was the tendency for the Finish ITI's to be grouped within a tighter range than the Start ITI's (Start ITI mean = 499.2ms, sd = 293.75ms; Finish ITI mean = 521.06ms, sd = 248.93ms). It was found that Start ITI correlated with Start ITI less Finish ITI ($r = 0.574$, $p < 0.001$). That is, relatively large or small Start ITI's tended to yield greater alterations on the part of the subjects than did less extreme Start ITI's. Although the effect was not massive, it was large enough to conclude that subjects tended to choose Finish ITI's smaller in range than Start ITI's, and hence tended toward some 'average' ITI.

EXPERIMENT II

Method

The aim of Experiment II was to ascertain the effect that dialogue context has on the task in Experiment I. Instead of hearing only a few seconds of speech either side of the turn transition, subjects were presented initially with approximately ten seconds of speech either side of the turn transition, the exact amount depending on the details of each dialogue. A transcript of each dialogue was provided. The turns immediately surrounding the target transition point were highlighted, so that subjects knew which turn transition in the dialogue to pay attention to. Having heard the entire dialogue, subjects then heard only those turns immediately around the target transition, as in experiment I. Note that the Start ITI in both the first presentation of the dialogue and the first presentation of the target exchange were identical. At no stage was the Original ITI heard.

Results & Discussion

A multiple regression analysis was again carried out, again indicating that a significant proportion of the variance in Finish ITI's could be accounted for by Start ITI's and Original ITI's (R^2 = 0.4519, F (2, 747) = 307.919, $p < 0.0001$). Start ITI's had a significant influence on Finish ITI's ($\beta = 0.65$, $p<0.001$), and that Original ITI's had a

very small, yet significant, influence (β = 0.07, p = 0.01). It would seem, therefore, that given greater context subjects were influenced to a small extent by the Original ITI's, which at no stage did they hear. This suggests that an increased amount of context may yield a greater number of cues to some idealised ITI. In particular, the increased amount of conversation heard by the subjects may have given them a better impression of the rhythmic structure.

A comparison was again made between the choices made by the two groups of subjects. A simple regression analysis of the relationship between Finish ITI's of group A, and Finish ITI's of group B showed no significant correlation between the Finish ITI's of the two groups of subjects (r = 0.1, p = 0.053), although the correlation only just missed the standard 95% significance level. Therefore, while this result indicates as in Experiment I that Start ITI's must have had a significant bearing on the choice of Finish ITI, it also backs up the findings from the multiple regression analysis that Original ITI's had a greater influence than in Experiment I.

As in Experiment I, a tendency emerged for the Finish ITI's to be grouped within a tighter range than the Start ITI's (Start ITI mean = 499.2ms, sd = 293.75ms; Finish ITI mean = 476.38ms, sd=265.34ms).

It was also found that Start ITI correlated well with the difference between Start ITI and Finish ITI (r = 0.498, p<0.001). So, similar evidence emerged as in Experiment I that subjects tended, if only to a small degree, to choose Finish ITI's which were grouped about some 'central' range of ITI's.

GENERAL CONCLUSIONS & DISCUSSION

Experiments I and II lead one to conclude that subjects were primarily influenced by factors other than perceptual isochrony to judge the 'ideal' ITI in exchanges.

Experiment II did reveal a very small yet significant relationship between Original ITI's and Finish ITI's, indicating that a greater degree of contextual information has a small effect on the choice of Finish ITI. This finding is not surprising if one assumes that context plays an important role in all aspects of language. The problem here is to decide whether a greater amount of context would give subjects a better sense of the rhythmical structure of a dialogue, which would be necessary for the notion of rhythm-as-coordinator. But according to this notion, the only rhythmic structure that ought to be necessary are the few beats occurring before the turn transition. These were present even in the low-context situation in Experiment I, where Original ITI's had no significant effect on Finish ITI's. It should also be emphasised that any contextual effect was substantially smaller than the effect of Start ITI's.

One possible objection to these findings is that subjects, rather than altering the ITI's freely until they were completely satisfied that they had reached a 'natural' ITI, felt that they were under some time pressure and chose a value which was not vastly different from the Start ITI. Anecdotal evidence from the subjects would suggest however that this was not the case.

Also, an explanation for the lack of correlation between Original ITI values and Finish ITI values may have been caused by the non-spontaneous nature of the task. That is, that subjects were aware of an upcoming turn transition point through repeated exposure to an exchange, whereas in natural dialogue hearers would possibly be able to detect a turn transition point before it occurred through syntactic, pragmatic, intonational or other cues. Of course, if a rhythmic process were being used in natural dialogue, the results ought not be affected by repeated exposure, since in both natural and artificial situations the same rhythm would be timing the start of the second turn. But overriding these considerations is the observation that in both experiments the Original ITI was at best significantly less of a factor than Start ITI.

An interesting result emerged in the range of Finish ITI's for each experiment. While there were significant correlations between Start ITI's and Finish ITI's, there was evidence that longer Start ITI's produced Finish ITI's which were slightly shorter, and vice versa. Subjects seemed

to be sensitive to different ITI's, in that they were able to detect whether a given Start ITI was particularly long or short. They then attempted to adjust the ITI on this basis. The findings of [3] confirm this by showing a relatively broad tolerance of what constitutes a 'natural' ITI, yet with subjects' being able to recognise long from short ITI's.

Even if there exist rhythmic cues, the claim that they are used directly in timing entry to the floor does not seem to be borne out by this data. It is possible that there are rhythmic processes which facilitate the detection of turn-transition points, even if it is not in a precise enough manner to be detected by these experiments. However, these processes co-exist with a variety of cues signifying the imminent closure of a turn, and possible passing of the conversational floor.

One of the arguments used in favour of rhythmic coordination is that the mutually constraining principles of *earliest possible start* and *intelligibility* [12] would by default yield latching in conversation (latching being that situation where the close of a speaker's turn coincides with the start of a second speaker's turn). However, Couper-Kuhlen points out that this is not generally the case, and reasons that one of the causes for this is that the coordination process is determined rhythmically. But the reasons why participants in a conversation often overlap, or delay their entry to the floor by fractions of a second after the close of a previous turn are, as I hope to have pointed out here, apparently highly complex, involving cultural norms, pragmatic concerns and cognitive limitations, but not necessarily rhythmic factors. The evidence that perceptual isochrony plays anything more than a minor role in timing and coordinating turn-taking is currently thin.

REFERENCES

[1] Abercrombie, D. (1967), *Elements of General Phonetics*, Edinburgh: Edinburgh University Press.

[2] Anderson, A.H., Bader, M., Bard, E.G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S. & Weinert, R. (1991), "The HCRC Map Task Corpus", *Language and Speech*, vol. 34 (4), pp. 351-366.

[3] Bull, M. (1994), "Isochrony and the Rhythm of Conversation", Proceedings of the Edinburgh Linguistics Department Conference '94, pp. 5 - 16.

[4] Classe, A. (1939), *The Rhythm of English Prose*, Oxford: Basil Blackwell.

[5] Couper-Kuhlen, E. (1993), *English speech rhythm: form and function in everyday verbal interactions*, Amsterdam: John Benjamins.

[6] Darwin, C.J., & Donovan, A. (1980), "Perceptual studies of speech rhythm: isochrony and intonation", In *Spoken Language Generation and Understanding*, Proceedings of the NATO Advanced Study Institute. Dordrecht: D. Reidel.

[7] Donovan, A., & Darwin, C.J. (1979), "The perceived rhythm of speech", Ninth International Congress of Phonetic Sciences, vol. 1, pp. 268-274.

[8] Halliday, M.A.K. (1985), *An Introduction to Functional Grammar*, London: Edward Arnold.

[9] Lehiste, I. (1977), "Isochrony Reconsidered", *Journal of Phonetics*, vol. 5, pp. 253-263.

[10] Martin, J.G. (1972). "Rhythmic (hierarchical) versus serial structure in speech and other behavior", *Psychological Review*, vol. 79, pp. 487-509.

[11] Pike, K.L. (1945), *The Intonation of American English*, Ann Arbor, Mich.: University of Michigan Publications.

[12] Sacks, H., Schegloff, E.A. & Jefferson, G. (1974). "A simplest systematics for the organization of turn-taking for conversation", *Language*, vol. 50, pp. 696-735.