THE PATTERN OF PAUSES IN MONOLOGUES: A CASE IN MANDARIN CHINESE

I-li Yang

楊懿麗*

摘　要

假如語言是心靈的窗戶，那停頓就是透視說話內心歷程的門洞。說話中的停頓可以有許多不同的義意：它可能是音與音之間必有的現象 (articulatory pause)，它也可能只是為了換氣 (breath pause)，或許是為了決定說話的內容或字彙 (hesitation pause)，有時也會因為語句結構的關係 (juncture pause)，也很可能是為了達到某一種說話的效果 (rhetorical pause)。基於這些原因，遂有本研究。

本文探討國語口語中停頓的現象。除了對說話中的停頓做次數、長短、及位置等的分析之外，還研究相關的兩大主題：說話速度和說話單位，以便了解停頓在說話中所扮演的角色。

本文利用四段自然口語來分析停頓現象，把所得的結果，拿來和國外的研究作比較，看看說話中的停頓現象，有多少是跨語言所共有的現象，有多少是因語言或個人所造成的差別。因此本文先就相關的文獻做一番徹底的介紹，再提出本研就分析的結果，然後進行比對和討論。

研究結果，發現我們的四個研究對象在說話中所產生的停頓比例，較其他研究的發現為高，尤其是講者 A，因此他的說話速度 (SR and AR) 都是難得一見的慢（每秒只有 1.97 及 3.28 字）。同時我們也發現，國語的說話速度 (SR)，和其他語言一樣，與停頓有密切的關係－說話慢者停頓必較多較長。再者，本研究發現多數的停頓出現在子句或片語之前，但是停頓的長短和語句之結構並沒有很一致的關係。只有當連續的語句構成句子時，這種關係才明顯－即句子前的停頓通常較長，而子句或片語或單字前的停頓都較短。最後，我們用停頓界定出來的語句，讓我們看到了相當長的語句（含兩個以上的 IU），對於這些語句的時間特性，更加明白。也有助於我們對於語言產生現象之理解。

Abstract

Pauses are a common phenomenon in speech. There are pauses between the transition of phonetic segments; there are also pauses for breathing. Sometimes, we pause in order to decide on the right sentence structure or the right word; at other times, we may want to impress our listeners. These facts about pauses are what motivate us to take up the present study.

The purpose of the present study is to explore the matter of speech rate and speaking unit via an analysis of the pauses occurring in spontaneous speech. By the pattern of pauses, we analyze the frequency, the duration, and the location of pauses. In so doing, we hope to understand the process of speech production.

The data of the present paper are four speech samples from four native speakers of Mandarin Chinese. They are experienced speakers and are all delivering their speech in the most natural way. We use Macintosh SoundPro to analyze our speech samples

* 作者為本校語言所教授

and put all our results of analysis into statistical analysis.

Our most important results include:

1.Although articulation rate remains quite constant and is a function of pauses, our four samples have a relatively higher percentage of pausing compared to those found by other researchers. This phenomenon is especially acute with Speaker A who not only pauses a lot but also delivers his speech very slowly. Several reasons are given to account for it.

2.Our pause pattern does not have an exact correspondence with syntactic structures. Only the pause pattern of Speaker B, C, and D exhibits the influence of syntactic structure.

3.From Pause-defined units, we find that all the four speakers produce their speech in a highly systematic way; their discourse is structured by topics. Long pauses are often used to signal major semantic divisions--topics and subtopics; shorter pauses are used to transit from sentence to sentence or from clause to clause, etc. Only on some occasions when the speaker wants to emphasize a point or when he is at a loss, will a long pause be required, otherwise speech is often very fluent.

1. Introduction

The study of speech pauses, pausology, has been researched for some time since the 1950's (Rochester 1973, O'Connell & Kowal 1983, Yang 1996). It has been found that pauses are the determinant factors for articulation rate (as distinct from speech rate, see below for more detail) (Goldman-Eisler 1956, 1961a, 1968) and that they are significant indicators of speech activities--specifically planning and execution (Goldman-Eisler 1968). People do not pause during speech just for taking breath; instead, pauses, all kinds of them, are inadvertently used to help speech produced. Pauses occur not only between major discoursal and syntactic boundaries but also within clauses and phrases (Boomer 1965, Goldman-Eisler 1972, Butcher 1981, Stenström 1986). Pauses are speculated to be attributable more to the content or semantic aspect of speech production than to syntactic or other automatic aspects of language processing (Goldman-Eisler 1968, Butterworth 1975). People pause to "find the next focus," i.e., to decide "what to talk about next" sometimes and "how to talk about what they have chosen" at other times (Chafe 1980, P. 171).

One important issue in the study of pauses is the relation between pauses and speech rate. The speed of speech depends to a great extent on the frequency and the duration of pauses. It has been found that articulation rate or the absolute speech rate is rather constant, independent of differences in speakers, situations, etc. (Goldman-Eisler 1954, 1956, 1961a, b).

It has also been found that the distribution and the duration of pauses are not random. Pauses of different lengths at different places signal different processing demands in language production. Typically, longer pauses occur at pre-paragraph and pre-utterance

positions and are presumably in charge of global planning, while shorter ones occurring at constituent breaks or before major lexical items within utterances are supposed to take care of local problems, e.g., the dissolution of constituent structure or the selection of lexical items (Goldman-Eisler 1967, 1968; Butcher 1981).

Presumably, the distribution and the duration of pauses illuminate the underlying process of language production. Pauses and speech occur sequentially in a systematic manner. It seems clear that the production of speech emerges in perceivable units -- encoding units -- with pauses as one of the major delimiters. Many different kinds of encoding units have been proposed -- from words to phrases to clauses, but of special significance is the tone or intonation unit.

Pauses have been found to be determined by many factors--linguistic (cognitive), affective, and communicative (or even physiological). A pause may occur simply because of a semantic or syntactic demand, or a limitation on memory span, or an anxiety out of a task situation, or a rhetorical need of the speaker, or a different breathing situation. Therefore, the occurrence of pauses (their distribution and duration) is subject to great variations, variations caused by differences in individuals, in situations/tasks, and perhaps in many other factors such as sexes and languages and cultures.

In a word, pauses are important clues to the real-time underlying processes of speech production. From pauses, we come to understand how one is putting forth what he wants to say. It is for this reason that the present study is taken up. As a first attempt, this study aims to disclose the pause pattern of Mandarin Chinese and to see whether it agrees or disagrees with those found for other languages. Furthermore, the present writer would also like to know what the pause pattern obtained can say about the structure and the process of speech production in Mandarin Chinese, namely how speech is produced as defined by pauses. In other words, this paper would address the following issues:

1. What kind of pause pattern can be found in Mandarin speech?
2. What can the pattern say about the unit of speech production in Mandarin Chinese?

It is hoped that the results of this study can help develop a theory of speech production based on an understanding of the pause pattern in the Chinese language.

2. Definition of Terms

1. Pauses. In this paper, pauses refer to silent pauses only, i.e., the period of time when no phonation is being made. Hesitation phenomena such as false starts,

filled pauses, and repeats are not included.

2. Pause pattern. This refers to the occurrence of pauses--their frequency, distribution, and duration. In order to understand the pause pattern, the following measures are adopted.

A. Total speech time: the duration of speaking (including pauses) (ST)

B. Number of pauses (NP)

C. Number of syllables in each utterance (NS)

D. Pause time (PT)

E. Duration of articulation (excluding pauses) (AT)

F. Speech rate = NS/ST or PT+AT

G. Articulation rate = NS/AT

H. Percentage of pause duration to total speech time, PP = PT/(AT+PT) $\times$ 100

3. Utterance. An utterance in this paper is defined as a sequence of speech delimited by pauses lasting more than 0.1 s.[1] This utterance is considered as a unit of speaking. And it is hypothesized that this unit is comparable to a tone unit in that they are all processing units, defined differently though. An utterance in this paper is pause-defined, but the usual tone unit is defined prosodically, pauses being one of the cues. It is hypothesized that such pause-defined units have their own structure, and from the way they are structured, the process of speech production could be revealed.

3. Related Literature

As the present paper is interested in the pause pattern and the structure of the speaking unit in Mandarin speech, only the studies concerning the two issues will be reviewed.

3.1. Studies on Pause Pattern

Most early studies on pauses endeavored to describe the following three issues: pauses and the rate of speaking, the frequency and the duration of pauses, and the distribution of pauses over utterances. Let us first take up the issue of pauses and speech rate.

[1] 0.1 s is adopted as the cutoff point for reasons to be discussed in Section 4.

3.1.1. Pauses and the Rate of Speaking

Using interviews of eight subjects (five patients and three staff members at the Mandsley Hospital), Goldman-Eisler (1956) obtained the following statistics.

Table 1. Means (M) Standard Deviations (SD), and Variation Coefficients (SD/M × 100) or V of Speech Rates (SR, in syllables per second), Percentage of Pauses (PP), and Articulation Rates (AR, in syllables per second). (Goldman-Eisler 1956, p. 138)

	SR			PP			AR			NU
	M	SD	V	M	SD	V	M	SD	V	
SI	4.3	0.78	18.1	4.4	4.01	91.1	4.5	0.75	16.6	26
Co	3.9	0.53	13.6	19.3	10.4	53.8	4.9	1.12	22.8	53
He	3.7	0.44	11.8	27.9	18.2	65.2	5.0	1.03	20.6	11
SII	3.3	1.09	35.8	29.8	9.7	32.5	4.7	0.54	11.5	15
Jo	3.3	0.93	28.2	34.3	12.5	36.4	5.0	0.69	13.3	29
Mu	2.8	0.78	27.8	43.6	12.9	29.6	5.2	1.06	20.4	54
Pea	2.7	0.45	16.6	53.2	15.0	28.2	5.9	1.48	25.1	46
BI	2.3	0.81	35.2	47.6	14.3	30.0	4.4	0.60	13.7	33
*M	3.3	0.71	21.5	32.5	14.92	45.9	4.95	0.45	9.1	

* The mean was added to the table by the present writer.

One remarkable finding along with these figures is that articulation rate remains rather constant for the eight subjects as a group and within each of them (ranging from 4.4 to 5.9 syllables), while the total speech rate varies in accordance with the lengths of pauses, again in both situations (the group as a whole and within each individual). The occurrence of pauses contribute to slower rate of speaking (Compare Pea and BI with SI and Co). In other words, speech rate is a function of pauses.

There are also statistics on speech rate and the like obtained for other languages. For example, in a cross-cultural study of speech rate, Osser & Peng (1964) obtained an average of 595.7 phonemes per minute for American English native speakers, and 572.5 phonemes per minute for Japanese native speakers, or 9.9 phonemes per second for English and 9.5 phonemes per second for Japanese respectively. If a syllable in both languages can be conceived as consisting of three phonemes (CVC), then Osser & Peng's results will be 3.3 syllables and 3.2 syllables per second. Not only did they find no significant difference between the speech rates for the two languages, but also

their results fall relatively safe within the range obtained by Goldman-Eisler. Osser & Peng did not analyze articulation rate, though.

Some statistics have been found for French and German too. The following table summarizes the findings of three studies: Grosjean & Deschamps (1975) on French, Raupach (1980) on German and French, and Duez (1982) on French again. Both G&D and Raupach used college students as subjects to tell a cartoon story, but Duez used three types of speech: political speeches, political interviews and casual interviews.

Table 2. Summary of Results of Three Studies on SR, PP, and AR

			SR	PP	AR
G&D		French	2.57	41.29	4.45
Raupach		German	2.57	37.9	4.09
		French	1.94	44.6	3.69
Duez	PS		3.26	37.3	5.2
	PI		4.26	25.3	5.7
	CI		4.06	22.0	5.2

In the table, there seems to be some disparity between the figures regarding French obtained in the three studies. The differences between G&D's and Raupach's may be attributable to three factors: (1) In G&D, a pause is treated as one only when it is longer than 250 ms, but the cut-off point in Raupach is 300 ms. Therefore, a number of shorter pauses are not counted in Raupach's study, resulting in longer duration for articulation time, hence lower speech rate. (2) Raupach used description as task for collecting speech, but G&D used both description and interviews. Previous research (Goldman-Eisler 1961b) has found that more pauses occur in descriptions than in interviews. (3) The sample size is not large enough, thus individual peculiarities may contribute to the difference.

As to the difference between Duez and G&D, a similar reason can be found: Duez has treated silence, filled pauses, false starts, repeats and lengthened syllables as "pauses" and their duration is excluded from articulation time. This will certainly increase articulation rate.

If we compare the figures in Table 2 with those in Table 1, we can have the following observations:

1. Despite the differences in the methods of speech collection and data analysis, articulation rate, or the absolute speech rate, does seem to remain quite constant from speaker to speaker, from situation/task to situation/task, and from language to language.

The mean is 4.82, range 3.69-5.7, mode 5.0 and 5.2, and median 4.9.[2]

2. Speech rate is a function of pauses, whether the pauses concerned refer to silence only or to other types of hesitations. More pauses mean slower rate of speaking.

3. Yet from the percentage of pauses, we do find that speech rate differs in different situations and different languages. Political speeches and cartoon descriptions seem to produce more pauses, and French seems to be spoken faster (having less pausing) than the other two languages.

These three points should be kept in mind as we discuss the other two issues.

3.1.2. The duration of Pauses

As we just mentioned, speech rate is a function of pauses. Therefore, the frequency and the duration of pauses become the determinant factors to account for the variation in speech rate (Goldman-Eisler 1956). In another study (1961b), Goldman-Eisler studied the duration of pauses produced in seven different situations, and the results are reproduced in the following table (p. 233).

Table 3. Mean Percentage Occurrence of Pauses of Different Durations

SS*	<0.5 sec.	1.0 sec.	2.0 sec.	3.0 sec.	3.0-8.0 sec.	>8.0 sec.
CDS	47.8	23.7	17.2	6.0	4.6	0.7
CSS	43.6	19.8	16.3	8.8	9.6	1.9
CDL	59.6	24.3	12.7	2.7	0.7	0.0
CSL	63.7	20.0	13.5	2.0	0.8	0.0
DA	49.9	37.1	12.0	1.0	0.0	0.0
DAD	41.4	41.1	16.0	1.3	0.1	0.0
PsyIn	16.4	33.9	28.6	10.8	9.6	0.6

* SS=speech situations; CDS=cartoon descriptions (spontaneous); CSS=cartoon summaries (spontaneous); CDL=cartoon descriptions (learned); CSL=cartoon summaries (learned); DA=discussions (adults); DAD=discussions (adolescents); PsyIn=psychiatric interviews

It is clear from the table that

(1) pauses less than 0.5 second are the most common (in all but the psychiatric

[2] The present writer knows that it is not very adequate to just average the means. But due to lack of access to the original data, this simplistic method is adopted.

interviews). These pauses are those between 0.25 and 0.5 second (for 0.25 is the cutoff point in this study).

(2) The pauses lasting for one second are the next frequent in all speech situations but the psychiatric interviews where they rank No. 1 (33.9%).

(3) The 2.0 sec. pauses rank No. 3 in all the speech situations but the psychiatric interviews where their rank is No. 2.

(4) Pauses longer than 3.0 seconds are found more in psychiatric interviews and spontaneous descriptions and summaries (a total of 21%, 11.3%, and 20.3% for each situation).

(5) The same trend can be found about very long pauses (those extend over 8 seconds). These long pauses can be as long as 20 or 30 seconds, but they were found in descriptions, summaries, and psychiatric interviews. However, the percentages are very small (p. 234).

(6) Discussions (both adults and adolescents) seem to be a unique case in that 99% of pauses are less than two seconds, and the rest never longer than three seconds.

(7) Psychiatric interviews seem to be another special case: they have a smaller percentage of short pauses (50% less than one second), a large proportion of middle-length pauses (2, 3, 3-8 seconds 28.69%, 10.8%, and 9.6% respectively)), and very few extremely long pauses.

(8) In both the well-learned situations, long pauses diminish, while the great majority of pauses are less than 0.5 second. From these results, Goldman-Eisler speculated,

> In fact, the psychiatric interview was the only speech situation investigated in which pause length maintained a central tendency; pause length for the other conditions of speech was distributed exponentially, the frequencies fast diminishing after a pause length of one second but having a significant tail of very long pauses when descriptions and particularly summaries of the meaning of a cartoon were formulated anew. (p. 235)

While on the on hand, i.e., in terms of pause length, psychiatric interviews are very much like original descriptions and summaries; on the other hand, i.e., in terms of the frequency of pauses or "phrase length" (defined as the number of words produced per pause), they are "nearly identical with the frequency distributions of the well-learned reproductions of the description" (pp. 235-236). To this, Goldman-Eisler concluded,

> In other words, interviews approach in their proportion of long pauses the

> most intellectual speech production, i.e., those requiring the highest level of verbal planning, while in frequency of pauses they are comparable to the most automatic speech production requiring least verbal planning. (p. 236)

Such incompatibility may be attributed to the "emotionally charged reminiscences" aroused by the particular speech situation (p. 236).

While the case of psychiatric interviews has helped shed light on the relationship between pauses and "the process of selection and planning" vs "automatic action" (p. 236), of special relevance here is the relation between pause length and frequency. Taking all the situations into consideration, Goldman-Eisler suggested,

> While the total pause time must be a function of the frequency and length of the individual pauses, the two components may be subject to variation, which would be in inverse direction. The function is hyperbolic (y=ab, where y is the total pause time and a and b are frequency and duration of individual pauses. (p. 235)

In other words, if one pauses often, then the pauses would be shorter in duration. On the other hand, if one pauses less, then the pauses would be longer in duration. This complementary effect is presumably a contributing factor to the constancy of articulation rate. From Goldman-Eisler's study, we come to understand while different speech situations might trigger pauses of different lengths, in normal situations, pauses of varying lengths distribute in a systematic manner, implying that different pause lengths serve different functions.[3]

To sum up, Goldman-Eisler's results suggest that while pause duration is a function

[3] Goldman-Eisler suggested that of the three kinds of decisions occurring during pauses (lexical, structural, and content), content decision is what matters in psychiatric interviews. "The decision confronting the interviewee will therefore be whether to utter or to suppress contents presenting themselves more or less automatically. Much of the pausing in psychiatric interviews should account for conflict of this kind and the success of a long pause by a long and fluent verbal sequence would indicate that the decision was one largely of whether to open the flood gates of surging material and to what extent to contain it rather than how to formulate it; on the other hand, when long pauses are followed by statements structured by short pauses into a series of short phrases, we would suspect the decisions involved to be largely lexical and structural. Thus the combined measure of pause and phrase length is necessary to appreciate more specifically the processes underlying speech utterance." (p. 237)

of speech situations, the great majority of pauses are less than one second irrespective of speech situations.

3.1.3. The Distribution of Pauses

From previous sections, we come to understand that the occurrence of pauses varies greatly from person to person and from situation to situation, that speech rate is a function of pauses, and that pauses of varying lengths (whose occurrence depends to a certain extent on speech situations) often signal different demands of verbal production. In this section, we shall review studies dealing with the subtleties of the relationship between pauses and speech, namely the distribution of pauses over speech.

First, we must understand the distinction made by Lounsbury (1954) between two types of pauses: juncture pauses and hesitation pauses. According to Lounsbury, the former are brief (0.1 sec) pauses at "high-order" constituent boundaries, and the latter are usually longer (1-3 sec) and appear at the beginning or end of speaker units (quoted from Rochester 1973, p. 53). It is the latter that attract attention because they signal the points of " highest statistic uncertainly in the sequencing of units" (Lounsbury 1954, p. 99). Juncture pauses are purely linguistic artifacts--they appear between phonetic segments (close juncture), or between syllables such as *I scream* or *ice cream* (open juncture), or between phrases (internal open juncture) or between clauses and sentences (terminal juncture). Juncture pauses, according to Goldman-Eisler (957), are for breathing, called breath pauses by Goldman-Eisler. Based on a reading sample, Goldman-Eisler found that breath pauses last about 0.5 -1 second and make about 2.5-25% of total speech time. The frequency and location of breath pauses appear to be a function of grammatical structures. This speculation was again verified by Henderson et al. (1965). Henderson et al. compared reading samples with spontaneous speech and discovered that 77.4% of pauses in reading are used for breathing but only 34.1% is used for the same purpose in spontaneous speech. Furthermore, the breath pauses (103 of them) in reading samples all occur at grammatical junctures, but in spontaneous speech, only 68.9% fall at grammatical junctures. So they concluded that breaths are taken exclusively at grammatical junctures, but in spontaneous speech other factors than grammatical ones come into play.

Although there is some difference between Lounsbury's and Goldman-Eisler's calculations of the duration of juncture pauses (breath pauses), in general such pauses are considered to be relatively short (Crystal 1969). Research has shown that they often escape detection (Boomer & Dittman 1962, Martin 1970, Butcher 1981, Duez

1985). For example, Butcher (1981) discovered that juncture pauses are identified (over 75% correct) only when they are above 220 ms, but hesitation pauses are identified when they are only 80 ms. Clearly, because juncture pauses are obligatory for the speaker (for taking breath) and for the listener (perhaps for integrating verbal information), they are taken for granted, therefore detection is difficult. And hesitation pauses are not linguistic pauses; therefore, they can easily be detected because they are not supposed to occur.

Based on Lounsbury's distinction, Boomer (1965) calculated the frequency of pauses longer than 200 ms at different locations within clauses of varying length (from two to ten words), and discovered that the average lengths of juncture and hesitation pauses are 1,027 ms and 747 ms respectively (p. 155).[4] Why was there such a big difference? The reason, as pointed out by Hawkins (1971), is that Boomer had confused hesitation pauses with junction pauses (p. 285). If we stick to Lounsbury's proposal, then many of Boomer's juncture pauses are more than merely junctures. As barik (1968) explained,

> An important question, however, applicable not only to Boomer's study but to any consideration of pauses, is whether the pause which occurs between two phonemic clauses always function only as a juncture pause for the preceding clause or whether, depending upon its duration, a component of it may not represent a hesitation pause associated with the ensuing clause. (p. 156)

So Barik suggested that long pauses (>700 ms) at major juncture locations may be a combination of both juncture and hesitation pauses, with the first part (500 ms) being juncture and the latter part (200 ms) being hesitations.[5] Barik's solution seems to be quite reasonable, for we shall shortly see that most pauses at major juncture locations are quite long.[6]

Although Boomer may have erred in distinguishing juncture from hesitation pauses, he has made two other important observations: (1) most pauses cluster at clause-initial

---

[4] Boomer's data is spontaneous speech of 16 American male adults who were asked to talk about such topics as hobbies, sports, summer vacations, etc. for three minutes.

[5] The issue of juncture pauses vs hesitation pauses is not really settled, for people have very different views about them. If the juncture is close or open, then it is usually very brief as described by Lounsbury. But if it is at phrase or clause or sentence boundary, then it is frequently longer than 0.1 second.

[6] Butcher (1981) gave an even better solution to the issue of breathing and pausing. He found that long pre-utterance pauses are composed of three portions: one preinhalation pause, one for inhalation (about 400 ms), and one post inhalation pause.

positions (8% before Word 1 and 41% before Word 2, 18% before Word 3, all three together making up 67% of the total number (749) of pauses, and if the number of juncture pauses were included, then the percentage rose to 82%); (2) and from Word 2 on the frequencies of pauses decline all the way down to the end of the clause. Such results are very revealing as far as the distribution of pauses is concerned, as we shall see shortly. Let's now focus on a related issue--the duration of pauses at major juncture locations.

Goldman-Eisler (1972) has found that in spontaneous speech 77.9% of sentences are separated by pauses longer than 0.5 sec., while 66.3% of clauses have either fluent transitions (i.e., without perceptible breaks) or are separated by pauses less than 0.5 sec. Furthermore, 50.3% of sentences are separated by pauses longer than one second, and 15% by pauses longer than two seconds.

Such a picture can not be found at clause boundaries where pauses longer than 0.5 sec make only 24.5% before relative clauses, and 43.1% in the case of coordinate clauses. The majority of pauses are shorter than 0.25 sec, 62.8% before relative clauses and 50.7% before subordinate clauses, but only 33.2% before coordinate clauses. As to pauses between words within clauses, 93.1% are fluent (<250 ms). In other words, different structures have different degrees of integration, the order from most loosely to most tightly being coordinate, subordinate, relative, and words.

To put it in another way, there is a high degree of correspondence between utterance structure and length of pauses: higher order structures (e.g., sentences) are often separated by longer pauses, and lower order ones (e.g., words) by pauses of shorter durations.

This phenomenon has been found time and again. For example, using the Lund Corpus as data, Stenström (1986) obtained the following statistics for a monologue.

Table 4. Percentages of Pauses of Different Durations at Various Locations (p.210, p. 209*)

Pause types	Between sentences	Between clauses	Within clauses
Brief	32	79	80
Unit	44	21	18
Double	14	--	2
Treble	10	--	--

* The figures in column 1 are taken from Table 3 (p. 210), but those in the other two columns are calculated by the present writer on the basis of Stenström's Table 2 (p. 209).

Now we can see that Stenström's results are very much like Goldman-Eisler's on spontaneous speech in that most pauses between words and clauses are brief (80%), and those between sentences are usually longer (68% unit or longer).[7]

Furthermore, Stenström also studied the distribution and duration of pauses between larger units, namely, paragraphs in which discourse topics are developed. She found that treble pauses dominate at paragraph boundaries for both topic and subtopic paragraphs. The statistics are reproduced in Table 5.

Table 5. Silent Pauses (in frequencies) Separating topics and Subtopics

Type of paragraphs	Type of pauses			Total
	treble	double	unit	
Topic	15	7	9	31
Subtopic	10	6	5	21
Total	25	13	14	52

As we compare Stenström's results (Tables 4 & 5) with those of Goldman-Eisler's, we might come to a tentative conclusion: while the duration of pauses may be a relative matter conditioned in part by the type of data collected, pauses of varying durations do serve as demarcators of linguistic units of various sizes. Very short pauses are likely to appear between words, medium-length pauses at phrase or clause boundaries, and very long pauses at paragraph divisions. Such trends have been found time and again (Hawkins 1971, Grosjean & Deschamps 1975, Deese 1980, Chafe 1980).

It has been found that the distribution of pauses might be affected by different situations or tasks. For example Goldman-Eisler (1972) also investigated the distribution of pauses in a speech sample of reading, and some differences were found.

1. A higher percentage of short pauses (<0.25 sec) were found within clauses (between words) -- 98.3%. In other words, the speech of reading is more fluent than

---

[7] A brief pauses is here defined as "a silence perceivably shorter than unit," which in turn is defined as "the interval of an individual's rhythm cycle from one prominent syllable to the next," based on Crystal (1969, p. 171). This kind of description can be very confusing. For example, Deese (1980) also talked about "short" "medium" and "long" pauses. And such terms can mean quite different things to different researchers as shown in the following:

Garman (1990)		(Deese 1980)	
brief	100-400 ms	short	368-400 ms
unit	400-800 ms	medium	1388-1763 ms
double	800-1200 ms	long	3331-3856 ms
treble	1200 ms +		

spontaneous speech.

2. The same trend was found in terms of relative clauses: 65.5% were less than 0.25 sec, and only 16.8% less than 0.5 sec, 13.4% less than 0.75 sec.

3. But in terms of other subordinate clauses, the reverse was found: there was a decrease in the number of pauses less than 0.25 sec.

4. Such a reversal trend became more prominent in terms of coordinate clauses: there was a regression toward the mean "with the majority (72.5%) of transitions being pauses of less than 0.75 sec, the mode being less than 0.5 sec" (p. 110).

5. This regression toward the mean is even more pronounced with sentences, but the mode is 1.25 sec now.

In other words, reading seems to demand a different temporal pattern from spontaneous speech. "With readings, pauses within clauses disappear but between sentences they are highly concentrated about a mean value" (p. 11), but with spontaneous speech, fluent transitions are vary rare between sentences, but very common within clauses, and there is a wide range of pause length.[8]

Duez (1982) compared pause patterns of three speech styles, and found that the pause distribution in political speeches is quite different form those in the other styles (political interviews and casual interviews). Table 6 presents the mean duration and the percentage of pauses between clauses, phrases and within phrases.

Table 6. Mean Durations and Percentages of Pauses at Between-clause and Between-phrase and Within-phrase Positions in Three Speech Styles

	PS		PI		CI	
	MD	%	MD	%	MD	%
Betw-cl	930	43.2	797	45.0	802	57.8
Betw-ph	708	50.4	588	42.9	632	30.3
Within-ph	462	6.4	501	12.1	401	11.9

It is obvious that the duration of pauses is a function of linguistic units: between-clause pauses are longer than between phrase pauses which in turn are longer than within-phrase pauses in the three types of speech. It is also clear that the difference in the distribution of pause duration, in terms of both between-clause and between-phrase pauses, is larger between political speeches and the other two types than between

[8] This pattern has been found by Grosjean (1980) too.

the two other types. While in casual interviews, a greater percentage of pauses (57%) are located at clause boundaries, in political interviews such trend appears at phrase boundaries (50.4%). The small percentage of between-phrase pauses in casual interviews is probably an important factor contributing to the fast speech rate in such situations.

Duez further discovered that 55.7% of between-clause pauses and 39.2% of between-phrase pauses in political speeches are longer than 750 ms, but the corresponding averages are 38.3% and 21.3% in political interviews and 33.2% and 22.9% in casual interviews. Such facts suggest that pauses in political speeches might serve other functions than purely linguistic ones, "they help to emphasize ideas and arguments" or "to impress an audience" (p. 21), or more specifically to serve "stylistic function" (p. 26).

The last issue in this section is the distribution of pause duration within the smaller unit of either the clause or the phrase. Some of the studies we have reviewed have shown that these within-clause or within-phrase pauses are relatively short and occur not as frequently as in clause or sentence boundaries. But exactly in what manner do ther occur? A number of studies have addressed this issue.

The first study is Boomer's (1965) which we have already discussed a little. By counting the pauses at word boundaries over the phonemic clause, Boomer found that the total frequencies of pauses before Word 1 and onward until the last word are:

38, 310, 138, 93, 56, 46, 24, 11, 10, 3.

The individual frequencies for each clause of different length show the same tendency. Although Boomer's results are revealing in terms of the clustering of pauses at clause initial positions and the decreasing trend of frequencies at succeeding word boundaries, they tell us nothing about the nature of the words, nor the duration of pauses.

As to this question, Hawkins (1971) managed to provide some information. He asked children (aged 6.5-7) to tell a story and obtained the following results.

Table 7. Distribution and Duration of Pauses at Four Locations*

Locations	F	%	Total pause time	Mean duration
a	889	66	1416.6 (s)	1.59 (s)
b	137	10.2	126.9	0.92
c	185	13.7	209.4	1.13
d	136	10.1	104.1	0.76
Total	1347	100	1857	1.37

* (a) at clause boundaries (b) before the predicator (c) at group (phrase) boundaries within the clause (d) at word boundaries within the group (phrase)

From this table we find again that not only do the great majority of pauses occur at clause-initial positions, but also they are relatively longer in duration. Both (b) and (c) locations are phrase boundaries :(b) refers to verb phrase and (c) to noun and prepositional phrases), and (d) word boundaries. If (b) and (c) are combined, then the frequency of pauses at phrase boundaries is more than twice to that of pauses at word boundaries. What is more, the durations differ too (1.02 s for (b) and (c) combined, 0.76 for (d)).

Hawkins' results may be a little inflated (since his subjects were children), but they certainly agree well with our earlier conclusion: higher order structures are separated by longer pauses and lower order ones by short pauses.

In addition, Hawkins also pointed out that a great number of pauses (407, 29% of total) occur before a linking conjunction and 11.8% after a linking conjunction.

Such a trend is found by Stenström (1986) too, as shown in the following table.

Table 8. Pauses and Conjunctions at Sentence and Clause Boundaries (frequencies and percentages)

	Pause+Conj		Pause+Conj+Pause		Conj+Pause		Total
Bet S*	60	.82	11	.15	2	.03	73
Bet CC	28		1		--		29
Bet SC	13		--		3		16
subtotal	41	.94	1	.01	3	.05	45
Total	101	.86	12	.10	5	.02	118

*S=sentence, CC=coordinate clause, SC=subordinate clause

Stenström also studied other within-clause and within-phrase pauses than those around the conjunction. She found that most of these pauses are brief or unit and mostly "between elements of noun phrases and the prepositional phrases" as found by Hawkins too. The percentages of these pauses, along with those of other pauses are reproduced here too.

Table 9. Percentages of Pauses Within and Between TUs in the Syntactic Hierarchy (including Reformulations*)

Syntax	Between Tus	Within TUs
Bet sentences	0.40	0.00
Bet clauses	0.12	0.06
Bet cl ele	0.41	0.38
Bet words	0.04	0.51
Bet reform	0.03	0.05
Total	1.00	1.00

* Reformulations refer to repeats and false starts.

The table shows that tone unit boundaries often coincide with sentence boundaries (40%), phrase boundaries (41%) and clause boundaries (12%), but within tone units, pauses most often occur between words (51%) and between phrases (38%).

By now we shall have some ideas as to where pause might occur within the sentence or the clause, and how long these pauses might be.

Using the data of Crystal and Davy (1975), Garman (1990) obtained some statistics and presented in the clearest possible way as in the following figure.

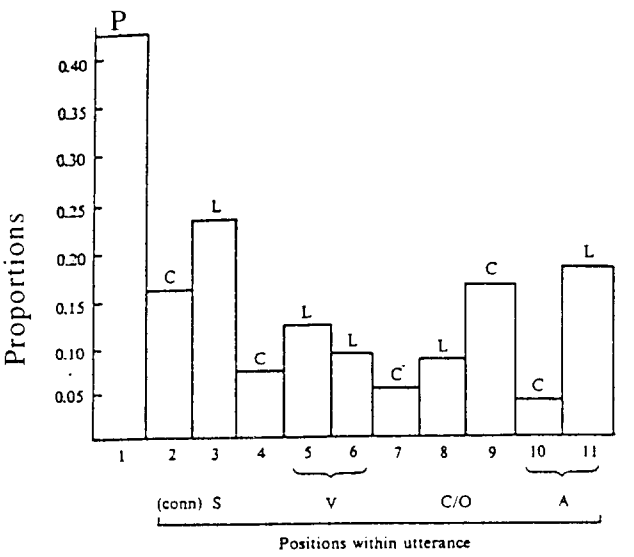

Figure 1. Proportions of Disfluencies at 11 Positions (p. 123)

The 11 positions are defined as follows:

1. pre-utterance;
2. after a connective (and, or, if, unless, etc.) occurring in utterance initial position
3. before a major lexical class word, an adjective or noun in the subject noun phrase, frequently occurring after a determiner;
4. before the verb phrase
5. within the verb phrase, between auxiliary verb and the following elements
6. some of the pauses in 4 and 5 standing directly in front of the main verb, and this position tallies with their combined incidence;
7. occurring between the verb phrase and the following phrasal constituent, either a complement or object;
8. before a major lexical-class word in complement or object constituents;
9. before an adverbial constituent, if there is one;
10. between a preposition and the following noun phrase;
11. before a major lexical-class word in an adjective phrase. (pp. 123-124)

Garman pointed out that 43% of clauses are preceded by a pause (position 1 in the figure), 16% have a pause between the connective and the next constituent (at Position 2), and 24% of pauses occur after a determiner, and that following Position 3 the percentages appear in gradual decline with the exception of Position 9 and Position 11 (if we compare the C columns and the L columns separately). The figure also indicates very clearly that pauses at different positions serve different functions--some for over-all content planning (P), some for constituent break (C), and some for lexical choice (L). Clearly, lexical choice involves mainly of the choice of nouns, adjectives, and verbs. Such analysis corresponds well with Goldman-Eisler's earlier speculations (1958, 1961b, 1968, 1972).

To sum up, from the studies reviewed, the distribution of pause duration is partly conditioned by the size of linguistic units, and partly by the situation of speaking. While different speech situations might result in the variation of pauses at certain positions, by and large, larger linguistic units (e.g., paragraphs or sentences) are often preceded by longer pauses, and smaller units (e.g., phrases or words) by relatively shorter pauses. In addition, the locations of pauses within the clause seem to be predictable on the basis of syntactic structure too. This eventually leads us to the question of what makes up an utterance.

3.2 Pauses and the Unit of Speaking

As we have said in the very beginning, an utterance in this paper is defined temporally--namely, that portion of speech separated by pauses longer than 0.1 s. Previous studies have used other terms that refer to more or less the same thing. These include phonemic clause (Boomer 1965), idea unit (Chafe 1980, Beattie 1980a), tone group (Halliday 1967) or unit (Crystal 1969), performance unit (Stenström 1986), performance structure (Grosjean et al. 1979). Some of these units are defined syntactically, some semantically, and some phonologically. In the following, we shall try to have a close look at some such units.[9]

Based on the definition of Trager and Smith (1951), Boomer (1965) used the phonemic clause as unit of analysis.[10] He segmented his corpus into the phonemic clause, and then tabulated the frequencies of pauses occurring at various locations within and between clause boundaries. He discovered, much to the surprise of previous studies (Goldman-Eisler 1954, Lounsbury 1954), that the majority of pauses (both filled and unfilled) clustered at the beginning of the phonemic clause--the percentages for the first four positions being 13%, 41%, 17%, and 11%. It seemed that "the initial word in a phonemic clause sets certain constraints for the structure of what is to follow" (p. 156). This led Boomer to conclude that the phonemic clause is the unit of encoding.

Another study that looks upon the encoding unit from the same structural point of view is Goldman-Eisler's (1972). She analyzed pauses longer than 0.25 s and found that the majority of sentences in spontaneous speech are separated by pauses longer than 0.5 s (77.9%), while most clauses are divided by pauses less than 0.5 s (66.3%). Such great disparity led Goldman-Eisler to conclude that the sentence is the encoding unit. She said.

> the speaker...organizes his message in highly cohesive *sentence units* (underlined added) with a clear hierarchical structure whereby constituent clauses are temporally integrated into the same *sentence frame*, if by this we mean uttered with fluency, to a far greater extent than sentences are into the whole discourse ... Fluent transitions between sentences are

[9] The present author has written a short article on the unit of speaking (Yang 1996). But due to space limit, the paper gives only a very sketchy account of the speaking unit. Here, we would like to explore the issue to the best.

[10] A phonemic clause is a "phonologically marked macrosegment" which, according to Trager & Smith, contains one and only one primary stress and ends in one of terminal junctures /1, 11, #/" (p. 150).

> extremely scarce even in spontaneous speech, where the dynamics of improvisation may more frequently be expected to drive beyond the conventional point of arrest in continuous phonation. This scarcity of fluency between sentences as compared with clauses, not to speak of words within clauses, indicates that a basic property is involved, *marking sentences as distinct units of speech*, occupying in the general stream of the discourse a figure to ground position. The rarity of such fluent transitions in spontaneous speech would indicate that in most cases *a sentence presents the externalization of a thought unit.* We are reminded here of the great emphasis accorded to the sentence form by Wundt as a fundamental unit which becomes articulated and analyzed into the constituent elements of words and clauses as external linguistic expression is undertaken. (pp. 110-111,)

Although the correspondence between pause patterns and structural units (the clause and the sentence) is great enough not to be neglected, syntactic structure is not the only factor that causes pauses to occur, as is already implied in the definition of the phonemic clause. Other researchers have tried to look at the matter of the encoding unit from a perspective more relevant to the study of the spoken language, namely, prosody. Hence, the tone unit or intonation unit is proposed.

A tone unit, according to Crystal (1969), is "the most readily perceivable, recurrent, maximal functional unit to which linguistic meanings can be attached" (p. 204). It must have "one peak of prominence in the form of a nuclear pitch movement," and "after this nuclear tone there will be a tone-unit boundary which is indicated by two phonetic factors. Firstly, there will be a perceivable pitch change... . The second criterion is the presence of junctural features at the end of every tone-unit" (pp. 205-206). The pitch change can be either "stepping up or stepping down, depending on the direction of nuclear tone movement" (p. 205). And only when this pitch change can not be clearly specified must one turn to juncture pauses for aid.

Crystal discovered that the juncture pause is usually a slight pause and is often accompanied by "segmental phonetic modification (variations in length, aspiration, etc.)" (p.206). But "in the majority of cases the junctural pause co-occurs with a term from the pause system."[11]

Based on this idea of tone unit, Crystal collected natural spontaneous informal discussions and conversations from 30 British English speakers and obtained some

[11] "A term from the pause system" refers to a perceivable pause.

important results concerning the tone unit.

> 1. The average length of tone unit is 5 words; 80% fall within the range of 1 to 7 words.
> 2. Structurally, tone units most often coincide with the clause, but they may also coincide with the sentence (a two-clause one), or the phrase or even the word (adverbial, subject + predicator, predicator + complement, subject, complement, predicator, nominal group, vocative, postmodification within a nominal group), and it never happens that a tone unit will include two sentences (pp. 258, 260).

Crystal's data give us some idea as to what a tone unit looks like. Later authors managed to refine the definition of the unit (Cruttenden 1986, Du Bois et al. 1992, Chafe 1994). For example, Cruttenden (1986) defined a tone unit as composing of an external structure and an internal structure. There are four criteria for the external structure: pauses, anacrusis (the speeding up of speech at the beginning of utterances), final syllable lengthening, and resetting the pitch level of a new unit.

As to the internal structure, a tone unit (1) must have at least one stressed syllable, and (2) must involve pitch change before of after the stressed syllable.

Cruttenden's definition of the tone unit has been accepted by later researchers, e.g., Du Bois et al. (1992), Chafe (1980ab, 1987, 1994). Thus Du Bois et al. (1992) outlined the following five features as cues to the identification of the unit:

> 1. a coherent contour
> 2. reset
> 3. pauses
> 4. anacrusis
> 5. lengthening

Similarly, Chafe claimed that

> a coherent intonation unit is supported by a convergence of (a) the pauses preceding and following it, (b) the pattern of acceleration-deceleration, (c) the overall declination in pitch level, (d) the falling pitch contour at the end, and (e) the creaky voice at the end (p. 60).

Du Bois et al.'s criteria differ from Chafe's in just one aspect: the former include

reset but the latter creaky voice. However, both differ from Cruttenden's in that stress is not paid much attention to. For as Chafe pointed out, some utterances do not have accented stress and yet they are clearly identifiable as tone units. A typical intonation unit will acoustically look like the the following according to Chafe (1994):

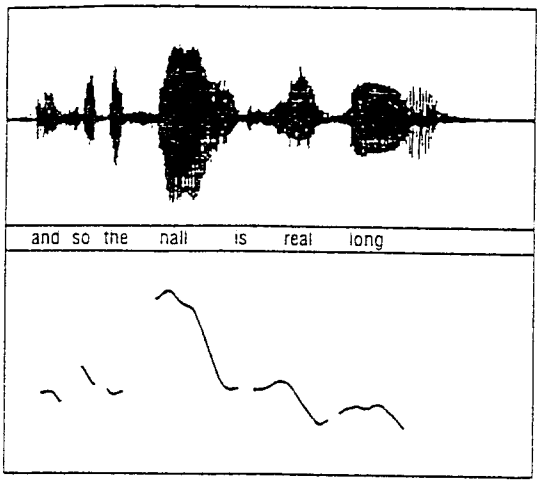

Figure 2. The Acoustic Properties of the Sentence "And so the hall is real long": upper – amplitude, lower – pitch (p. 58)

The figure has all of the cues prescribed just now, including a creaky voice at the end of the amplitude representation. What the figure shows is the form of an ideal tone unit. It is "prototypical" in that it exhibits all of the above-mentioned cues (Chafe 1987, Schuetz-Coburn et al. 1991). According to Chafe (1994), a typical tone unit often realizes itself in the form of a clause, called a clausal tone unit, which occurs about 60% of time in his data. The clausal tone unit is about 2 seconds in duration, and 5 words in length (1980).

Other authors also measured the duration and length of the tone unit. Butcher (1981) has found that a tone unit is 1.3 seconds in duration and 8 syllables in length, and Stenström (1986) has come up with 2.4 seconds and 4.1 words.

Chafe (1994) later classified complete tone units into two types: substantives are those convey "ideas of events, states, or referents"; and regulatories are those "regulating interaction or information flow." He obtained an average of 4.84 words for the former

and 1.36 words for the latter.

Although the tone unit is prosodically defined, there has been some psychologizing about it. For example, Chafe believed that the substantive type of tone unit is the typical type that expresses "a focus of consciousness." (1994, p. 65) Chafe (1987) speculated that when one intends to say something (event, state, or referent), he has to activate relevant concepts in the consciousness. And because of limited capacity in the memory span, the activated portion of consciousness (the focus) is often quite small and when realized it becomes the tone unit. As one concept is being activated, others are kept in the periphery (semi-active) and ready for activation. In this way, the stream of consciousness flows out and dresses itself in a sequence of tone units.

Grosjean (1980) also pointed out that the production of sentences is constrained by two factors: syntactic and performance. In terms of syntax, people tend to pause at grammatical junctures. However, "sentences were broken up into groups of words of more or less equal length" (p. 95) signaling a need "to balance the length of constituents in the output" (p. 102).

The above discussion may lead us to the conclusion that the most readily acceptable encoding unit is the tone unit, and that it often appears in the form of the syntactic clause. While this conclusion is valid to some extent, we have to remember that in real situations speech does not always emerge in clearly identifiable tone units. There are always atypical units that deviate from the prototypical to some degree. For example, sentence fragments (as in false starts) and backchannel expressions (such as OK, well) are often treated as tone units even if they do not have any prominence or nucleus or discernible pitch change (Schuetze-Coburn et al. 1991; Du Bois et al. 1992; Chafe 1994).

Thus, the identification of the tone unit can be very troublesome in actual practice. For instance, a stretch of speech may involve a coherent intonation contour and yet no pause follows, or it may be followed by a very long pause and yet there is no clear pitch change. As Brown et al. (1980) pointed out, Hallidaian analysis of the tone group is very problematic because it permits only one contour type. So they used pauses as delimiters of speech and were thus able to examine the various contour types within the unit on the basis of changes in fundamental frequencies. In their analysis of both reading from written texts and spontaneous speech, they found "the patterning of pause length corresponds to semantic structuring." (p. 56) Long pauses (over 1 second) are often used to mark topic divisions-called topic pauses by them. Contour-marking pauses (0.6-0.87 s) serve to separate complete contours most of which have two peaks of prominence rather than just one. The third kind of pause lasts 0.25-0.38 s and is used to "search"-hence called search pause. Such pauses often occur

at sequences with incomplete syntax. They are what others call as hesitation pauses.

Brown et al.'s study is largely qualitative, arguing for theoretical adequacy of their approach. They only pointed out that some of their pause-defined units are also tone units (in the Hallidaian sense), and some others contain two tone units. They did not present any quantitative information about the size and the structure of the units.

As pauses occur more than 80% of the time with other cues for the identification of the tone unit (Schuetze-Coburn et a. 1991), we want to analyze our speech samples by pauses only. We want to know how far we can go by using pause-defined units in uncovering the process of speech production.[12]

4. The Present Study

The present study aims to uncover the pause pattern of Mandarin speech. Specifically, we want to address the following issues:

1. pauses and speech rate and articulation rate in Mandarin speech,
2. the distribution and duration of pauses in Mandarin speech,
3. pauses and the unit of speaking in Mandarin speech.

In addressing these issues, as based on the literature reviewed, we hypothesize that while Mandarin is a language different very much from Western languages, in terms of the processing mechanism within the human mind, there will be some similarities found between the pause pattern of Mandarin speech and that of other languages (e.g., speech rate and articulation rate), but there may also be some differences (e.g., in the distribution of pauses). Specifically, we predict that

1. articulation rate will remain constant, but speech rate will fluctuate depending largely on the frequency and duration of pauses;
2. the duration of pauses will vary according to speaker/situation differences on the one hand, and to the syntactic structure on the other;
3. based on the constancy of articulation rate, the size and duration of the utterance, the encoding unit, will also remain quite constant.

[12] Lehiste (1979) also used pauses to segment speech. But again, he was more interested in exploring the idea of paratones than laying foundations for the theoretical status of pauses.

4.1. The Data

The speech data for analysis come from four abult male Mandarin speakers. They are pseudonamed as A, B, C, and D. All four of them are native speakers of Mandarin. A is about 60 years old, and his Mandarin is slightly accented because he comes from An-hwei, a province along the Yangtze River in the eastern part of Mainland China. He is a distinguished Buddhist monk and has been preaching Buddhism to the general public for about 35 years. B is about 45 years old, and speaks Mandarin with a little bit of Taiwanese accent. He is an instructor in a university in Taipei; he teaches philosophy courses in the school. C is again a teacher in a university in Taipei. He is a linguist, and comes from the central part of Taiwan. He speaks Mandarin without any accent at all. He is around the age of 40. D comes from Chekiang, a province in eastern China. He is a distinguished political figure, and speaks accentless Mandarin. He is about 55. All four can be said to be very good speakers in that they have had a lot of public speaking experience and are all well liked.

The speech of the four subjects was recorded while they were all speaking to groups of people. A was preaching Buddhism to people in a big lecture hall (with an audience size of about 200 people), B telling a story on a historical figure to his class of audience (about 40), C speaking to a group of graduate students (about 20) on the topic of neurolinguistics, and D narrating his own experience in practicing Buddhism to an audience of more than 100 people. As they spoke, they did not read from any written manusscript. They all knew their subject matters very well, and they all spoke in the most spontaneous way.

The speech thus collected is monologue in rather formal situations. The lengths of the four speech samples are 7 m 37.735 s for A, 7 m 35.596 s for B, 7 m 16.982 s for C, and 8 m 4.504 s for D.

4.2. Data Analysis

The collected speech was first transcribed by two linguistic majors, and then checked by the present writer. Then the speech was fed into the SoundPro Edit program of Macintosh for pausal analysis. In carrying out the analysis, a number of measures were taken:

1. Silent pauses lasting less than 0.1 second were neglected on the basis of Hieke et al's suggestions. Most pausal studies used the cutoff point of 0.25 second (as proposed by Goldman-Eisler 1956), but Hieke et al. discovered that pauses as short as 0.13

second were something other than merely articulatory pauses. We hope, by using the 0.1 second as the cutoff point, to capture as much as possible the significance of pauses in relation to speech production.

2. In segmenting the silence in the sound waves on the screen from the SoundPro, we have tried our best to exclude the silence resulting from syllabic lengthening from real silence. We feel that syllabic lengthening is a unique property of spoken Mandarin and deserves a paper of its own. So, we do not want to confound it with the silent pauses we are studying in the present paper.

3. Our next step is to use pauses as the only delimiters of speech. We use the utterance (as defined earlier) instead of the tone/intonation unit as our unit of analysis because we want to know what pauses do to speech. We have come, from previous studies (Brown et al. 980), to notice that the identification of tone units can be very troublesome. We believe that our method will be more objective although we are aware that in so doing we might run into other problems too.[13]

So, we have, on the basis of our pausal analysis, retyped the transcription in such a way that each "utterance" (as defined earlier) occupies one line on our transcription sheet. Then, we tried to decide the constituent structure of each utterance and code each of them for further statistical analysis. Thus doing, we were able to understand the close relationship between pausing and speech.

Our final step was to put all the figures from pausal analysis and constituent analysis into the computer for statistical analysis.

4.3. Results

The results of our analysis will be reported in four sections: 4.3.1 on the rate of speaking, 4.3.2 on pause durations, 4.3.3 on the distribution of pauses, and 4.3.4 on the unit of speaking. But before we proceed, let us present some information about the total time of pause and articulation and the total numbers of utterances and syllables in the four samples, as shown in the following table.

[13] Indeed we have encountered the problem of categorizing many speech segments. The place where one might pause is not always accountable by structural analysis.

Table 10. The Total Amount of Pause Time, Articulation Time, Speech Time, and the Total Numbers of Utterances and Syllables in the Four Samples

	PT (sec)	AT (sec)	ST (sec)	NU*	NS
A	201.194	256.541	457.735	143	783
B	131.137	319.275	455.596	181	1607
C	112.841	324.141	436.982	214	1943
D	192.851	291.653	484.504	186	1579

* The numbers of utterances are also the numbers of pauses.

The table tells us at once how much time each of the four speakers spent in pausing and in articulation. And within about equal amount of speaking time (from 436.982 sec to 484.504 sec), they produced different amount of speech (from 783 syllables to 1943 syllables). A's pausing time makes up 43.95% of total speaking time, B's 29.92%, C's 25.82%, and D's 39.8%. Both A and D paused a lot more than B and C. Such differences can further be found in Table 11 where the means, the standard deviations, and the variation coefficients are given (when calculated utterance by utterance).

Table 11. Profile of Pause Time, Articulation Time, and Utterance Length: Mean, Standard Deviation, and Variation Coefficient

	PT (ms)			AT (ms)			UL (syl)		
	M	SD	V	M	SD	V	M	SD	V
A	1407	1453	103.26	1794	962	53.65	5.48	2.75	50.27
B	726	624	86.06	1764	1021	57.87	8.88	5.54	62.45
C	527	393	74.59	1514	865	57.15	9.10	5.83	64.10
D	1036	813	78.47	1556	750	48.22	8.49	4.70	55.31

Here, we see that A's mean pause time is almost twice as much as B's and three times as much as C's, while D's about twice as much as C's There is also great variation within each speaker in pause time, as can be seen in the variation coefficients. Here, all the four have variation over 50%, and A and B show more variation than C and D.

As to mean articulation time, not much difference can be found from speaker

to speaker (from 1514 ms to 1794 ms), and within each speaker, the variation coefficient is also relatively small compared to that of pause time. In other words, mean articulation time (as far as the average is concerned) seems to remain quite constant.

Table 11 also shows the mean utterance length: A has a much shorter length than the other three and also the lowest fluctuation. This we shall return as we deal with the speaking unit--the utterance. Now, let us proceed to the matter of the rate of speaking.

4.3.1. Pauses and the Rate of Speaking

Table 12 presents the results on the rates of speaking--speech rate and articulation rate or absolute speech rate of the four speakers as well as the percentages of pauses in each speech sample.

Table 12. Mean, Standard Deviation, and Variation Coefficient of Speech Rate, Articulation Rate, and Percentage of Pauses of the Four Speakers

	SR			PP			AR		
	M	SD	V	M	SD	V	M	SD	V
A	1.97	0.96	48.59	38.96	21.94	56.30	3.28	1.21	37.04
B	3.51	1.25	35.23	29.66	17.80	60.02	5.01	1.23	24.53
C	4.28	1.61	37.58	27.85	17.20	61.76	5.82	1.48	25.43
D	3.39	1.25	36.88	37.18	17.75	48.73	5.36	1.05	19.49

The fact the both A and D pause a lot more than B and C is again found in this table (in the PP column), and it appears that B and C fluctuate more than A and D in this aspect (as indicated by the variation coefficients). The mean PP here is somewhat smaller than what we have obtained in total pause time and speaking time, because here calculation is done utterance by utterance.

As we look at the speech rate and the articulation rate, we find a greater range between the four speakers in speech rate than in articulation rate. Variation within individual speaker is also greater in speech rate than in articulation rate. Speaker A is a very special case: he has very slow rate in both SR and AR, although his PP is not much higher than Speaker D (38.96% and 37.18%). Also, he showed much greater variation than the other three in both AR and SR (as indicated by the variation coefficients).

As we conducted an ANOVA on SR, AR and PP of the four speech samples, we discovered that individual differences are significant at the 0.0001 level. After conducting Tukey's Studentized Range Test, the direction of difference in AR can be represented as C>D>B>A. The same test done on SR results in C>D=B>A. That is, while both D's and B's SR are greater than A's, C's is still greater than both D's and B's. There is no difference between D and B. In other words, in terms of SR, C and A are extreme cases. But in terms of AR, each of the four contributed to the variance in a systematic way: C's is greater than D's which in turn is greater than B's, which in turn is greater than A's. In the case of PP, the result is A=D>B=C; that is, A and D are very similar and both pause more than B and C who resemble each other very much in this aspect.

Despite the great difference among the four speakers in all the three variables (AR, SP, PP), there is some correlation between the three variables within each individual speaker. Table 13 presents the correlation coefficients matrix.

Table 13. Correlation Coefficients among AR, SR, PP in Each Sample

	SR	PP
(A)		
AR	0.62*	0.11
SR		-0.67*
(B)		
AR	0.67*	-0.04
SR		-0.75*
(C)		
AR	0.82*	-0.33*
SR		-0.79*
(D)		
AR	0.62*	-0.12
SR		-0.84*

* Significant at 0.0001

As the table shows, in all the four speech samples, SR and AR are positively correlated, and SR and PP are negatively correlated. While the close relation between AR and SR is not too surprising, the negative correlation between SR and PP is enlightening. In other words, the idea that speech rate is a function of pauses is again supported

by the present set of data even though individual difference is very great.

It is also necessary to point out that AR and PP are significantly correlated (negative) in C, a phenomenon seldom found in the literature. It is probably attributable to the small percentage of pausing (only 27.85%) found in this sample.

4.3.2. The Duration of Pauses

We have found from Table 11 that the total pause time for each of them is 201.194 seconds (A), 136.321 s (B), 112.841 s (C), and 192.851 s (D). Let us now take a good look at the frequencies and percentages of pauses of different durations in the four samples, as presented in the following figure.

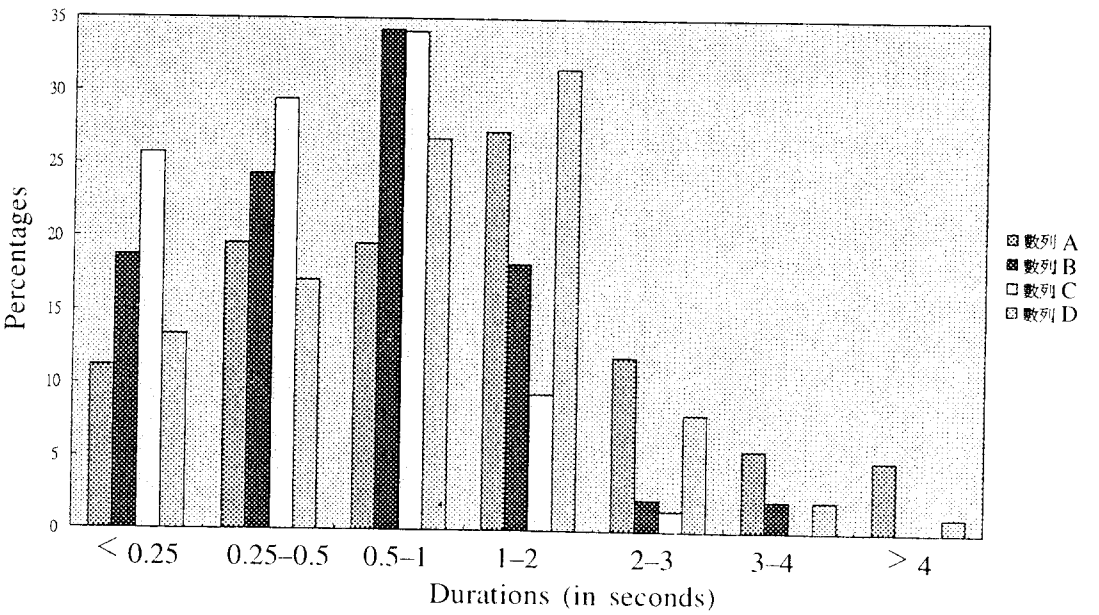

Figure 3. Percentages of Pauses of Different Durations in the Four Speech Samples

From the figure, we can see that for the four speakers the great majority of pauses are below 2.0 seconds. The greatest percentages for A and D are pauses lasting from 1.0 to 2.0 seconds, but for B and C from 0.5 to 1.0 second. A and D have a much larger number of pauses of longer durations (49.65% for A and 42.78% for D on

pauses over 1 second), while the other two tend to have more pauses less than 1 second--76.77% for B, and 89.25% for C. Clearly, both A and D showed a preference for longer pauses, while B and C veered to the other direction. The preference for shorter pauses is especially clear in C's speech, a total of 55.24% less than 0.5 sec. It seems that each individual has his own style of pausing.

If we lump together the last three categories into one (i.e., pauses longer than 2 seconds), then A seemed to have a even distribution of pauses of different lengths, with a slightly lower percentage for pauses less than 0.25 sec. B's pauses are mostly between 0.25 sec and 1 sec, with pauses less than 0.25 sec and 1-2 sec soming next. And C's majority of pauses are less than 1 sec. Finally, D's pauses cluster between 0.5 sec and 2 sec.

In order to see whether such distribution is significant, we conducted a Chi-square, and the results are significant: A at 0.05 level, and the rest at 0.001 level.

Such individualistic style of pausing, as reflected in pause durations, certainly affect the rates of speaking, as we have already seen. And yet, what is more interesting is where are all the pauses located? Do their occurrences have anything to do with the stretches of speech that follow them? These two issues we shall now turn to.

4.3.3. The Distribution of Pauses

Since we have used pauses as delimiters of utterances, there is always a pause before each utterance. So in order to understand the distribution of pauses, we have simply to decide he structure of the utterances. As we examined our samples closely, we came up with the following scheme for analysis: we clasified the utterances into sentences, clauses, phrases, and words. Several types of clauses are identified: simple, compound, complex, and compound-complex. We also noticed some fragments of clauses. As to phrases, we have identified six of them: noun phrase, verb phrase, prepositional phrase, adverbial phrase, adjective phrase and incomplete phrase. Finally, six categories of word classes are identified: noun, verb, adverb, adjective, preposition, conjunction, and interjection.

It is necessary to point out that the identification of form classes is very difficult in Chinese because Chinese is a loosely inflected language. However, we try our best to categorize the utterances purely on formal grounds. For example, the sequence 為甚麼 (*for what*) is designated as a prepositional phrase not as an adverb because it is clearly composed of a preposition and a noun, while 怎麼樣 (*how*) is treated as an adverb because the three morphemes are bound as a whole and make up an adverb.

Table 14 gives the frequencies and percentages of pauses and their mean durations before various types of constituent structures.

Table 14. The Frequencies and Percebtages if Various Types of Constituent Structures and the Mean Durations (in milliseconds) of Pauses Preceding them in the Four Samples

Type	Sample A		Sample B		Sample C		Sample D	
	F(%)	MD	F(%)	MD	F(%)	MD	F(%)	MD
Sentences								
Compo	0(0)	0	16(8.8)	773	6(2.8)	644	18(9.7)	1015
Compl	0(0)	0	13(7.2)	527	10(4.7)	460	13(7.0)	1189
C-C	0(0)	0	13(7.2)	994	2(0.9)	391	5(2.7)	1021
Frag	1(0.7)	1702	3(1.7)	644	6(2.8)	613	6(3.2)	957
Subtotal	1(0.7)	1702	45(24.9)	757	24(11.2)	536	41(22.0)	1064
Clauses								
Simple	37(25.9)	1171	47(26.0)	700	29(13.6)	625	61(32.8)	1004
Coord	1(0.7)	414	4(2.2)	782	4(1.9)	391	3(1.6)	1411
Subord	7(4.9)	986	9(5.0)	838	8(3.7)	334	9(4.8)	1437
Frag Sp	8(5.6)	903	8(4.4)	633	9(4.2)	664	7(3.8)	1025
Frag Co	2(1.4)	3681	2(1.1)	1188	7(3.3)	664	3(1.6)	473
Frag Sb	1(0.7)	782	0(0)	0	8(3.7)	604	1(0.5)	1380
Subtotal	56(39.2)	1179	70(38.7)	730	65(30.4)	582	84(45.2)	1088
Phrases								
Noun Ph	18(12.6)	2319	14(7.7)	697	38(17.8)	575	13(7.0)	1087
Verb Ph	36(25.2)	1213	42(23.2)	547	42(19.6)	323	26(14.0)	787
Prep Ph	3(2.1)	1150	1(0.6)	322	18(8.4)	606	5(2.7)	1408
Adv Ph	5(3.5)	1159	0(0)	0	4(1.9)	472	3(1.6)	675
Adj Ph	2(1.4)	2048	0(0)	0	1(0.5)	230	0(0)	0
Frag	0(0)	0	0(0)	0	0(0)	0	5(2.7)	524
Subtotal	64(44.8)	1543	57(31.5)	654	103(48.0)	470	52(28.0)	855
Words								
Noun	10(7.0)	1109	1(0.6)	644	4(1.9)	299	2(1.1)	989
Verb	4(2.8)	989	0(0)	0	2(0.9)	414	3(1.6)	966
Adv	3(2.1)	2423	3(1.7)	1779	4(1.9)	598	0(0)	0
Prep	0(0)	0	0(0)	0	5(2.3)	460	0(0)	0
Conj	1(0.7)	184	0(0)	0	4(1.9)	690	1(0.5)	1610
Adj	0(0)	0	0(0)	0	2(0.9)	805	1(0.5)	4187
Interj	4(2.8)	3060	5(2.8)	598	1(0.5)	2577	2(1.1)	414
Subtotal	22(15.4)	1312	9(5.0)	1495	22(10.3)	448	9(4.8)	975

Perhaps the most striking fact about the results, and one that is not unexpected, is that a great majority of pauses are used to segment clauses (especially simple clauses) and phrases (especially verb and noun phrases). The three categories combined make 63.7% for A, 56.9% for B, 51% for C, and 53.8% for D. While B, C, and D have comparatively larger percentages of sentences separated by pauses (24.9%, 11.2%, and 22% respectively), A has produced only one such instance. As to the pauses before single words, the four speakers seem to diverge a lot from one another. While one may prefer to pause more before nouns and verbs, another may do so in front of adverbs and prepositions.

On the basis of the subtotals, the mean durations of pauses in each major category differ very much from sample to sample. But within the same sample, the mean durations of pauses appear to be dependent upon the syntactic categories. For Speakers C and D, pauses before phrases and works are shorter than those before sentences and clauses. And for Speaker B, the same trend appears except for the Word category where pauses are rather long (1495 ms). As for Speaker A, pauses are longest before sentences, but shortest before clauses. In fact, he has quite long pauses, all exceeding 1 second with all the four categories.

As we examine the pauses within the sentence catgory, there does not seem to be any consistency in terms of sentence types: compound-complex sentences do not necessarily begin with longer pauses. This is true of all four samples.

But as we look within the clause category, pauses before subordinate clauses are longer than those before coordinate ones which in turn are longer than those before simple clauses in Samples B, C and D.

In the phrase category, the pauses before NP are longest in Sample A and B, but those before PP and NP are longest in C and D. In all the four samples, the pauses before VP and PP are in general relatively short.

Finally, in the word category, there is really not much systematicity across the four samples.

It is necessary to point out that pauses occurring before fragments (either sentence framents or clause fragments or phrase fragments) and those before conjunctions and interjections behave very much like those before complete sentences or clauses--as junctural. This is why sometimes they can be quite long (as in A's and C's interjections and D's conjunctions).[14]

[14] Being long does not necessarily signal major constituent boundaries. Some pauses before single words may be real hesitant pauses.

4.3.4. Pause-Deined Units

In Table 14 we have presented the frequencies and percentages of utterances that are categorized in terms of syntactic structure and the mean durations of pauses before each specific category. We know that the great majority of utterances are in the form of clauses and phrases in all the four samples: in A 39.2% and 44.8% (84%), in B 38.7% and 31.5% (70.2%), in C 30.4% and 48% (78.4%), and in D 45.2% and 28% (73.2%). Clearly, A and C have larger presentages for phrases and B and D for clauses. And the most prevalent type of clause is the simple clause (24.9%, 26%, 13.6%, 32.8%), and that of phrase is the verb phrase (25.2%, 23.2%, 19.6%, 14%) and the noun phrase (12.6%, 7.7%, 17.8%, 7%). Our pause-defined units that come next to clauses and phrases in proportion are sentences and words, having a perecntage of 0.7%, 15.4%, 24.9%, 5%, 11.2%, 10.3%, 22%, 4.8% respectively for the four speakers. All these figures give structural information about the units. What about other characteristics? Let us look at the matter of size first. Table 16 gives a profile of the size of utterances in the four samples.

Table 15. A Profile of the Size of Utterances in the Four Samples: Mean Length (in syllables), Standard Deviation, Variation Coefficient, Range, Mode, and Median.

	ML	SD	V	Range	Mode	Median
Sample A	5.48	2.75	50	1-14	4	5
Sample B	8.88	5.54	62	1-32	6	7
Sample C	9.10	5.83	64	1-36	3	8
Sample D	8.49	4.70	55	1-26	7	7

Compared to B, C, and D, A's utterances are smaller in size (range=1-14, and M=5.48). B, C, D are closer to one another in both aspects. Based on the variation coefficients, we can see that the size of the pause-defined units has a greater variance in Samples B and C than in A and D; B's and C's utterances can be quite long. It is very interesting to find that each speaker has his favorite size of unit (mode): from 3 to 7, and yet their medians seem to cluster toward 5-8, the size closest to the means. This tendency is best illustrated in the following figure.

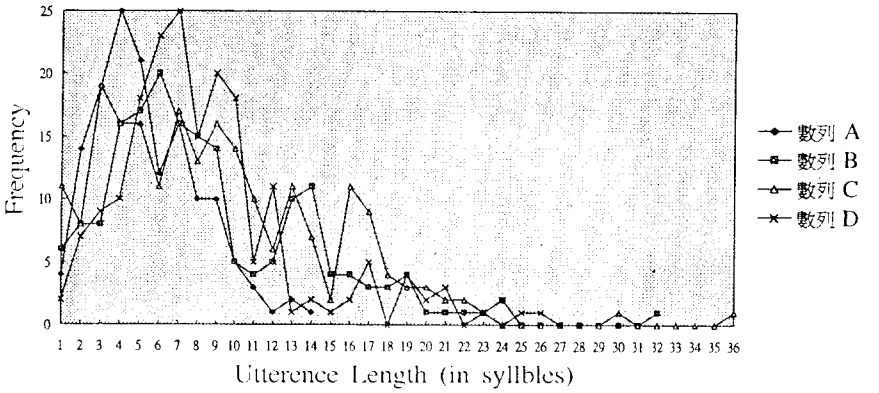

Figure 4. Frequencies of Utterences of Varying Lengths

The figure presents the frequencies of utterances of varying lengths. The four speakers overlap in some utterance size and divert in others. But on the whole, the four speakers all prefer to cluster toward the utterance length of 3-8.

Now, let us look at the temporal aspect of the units. In Table 11, we have seen that the average articulation time is 1794 ms for A, 1764 ms for B, 1514 ms for C and 1556 ms for D. These figures are also the durations of the average utterance for each of the four speakers.

To tell the story a little simpler, Figure 5 presents the frequencies and percentages of articulation time of different durations.

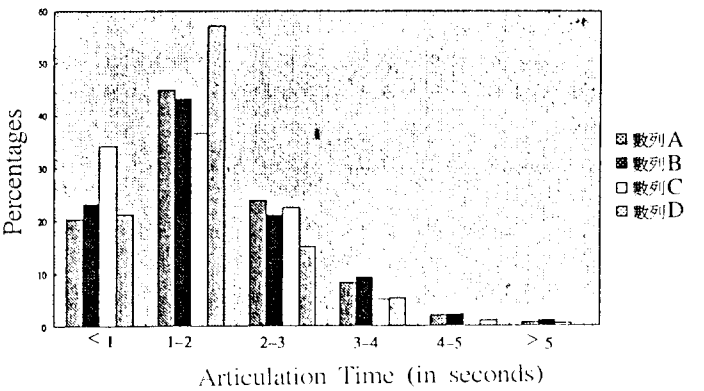

Figure 5. Percentages of Articulation Time of Different Durations

The figure shows that for the four speakers, the most frequent length of articulation time is between 1 and 2 seconds, a fact that has been represented by the average articulation time we have obtained.

As we have shown already, the range of utterance length is quite large for all but Speaker A, so we need to know the average amount of time spent in uttering speech of varying lengths by the four speakers. Table 16 presents the results of our calculations.

Table 16. Mean Durations, Standard Deviations, Variation Coefficients of Utterances of Varying Lengths

UL	Sample A			Sample B			Sample C			Sample D		
	MD	SD	V	MD	SD	V	MD	SD	V	MD	SD	V
1	357	69	19	422	94	22	381	71	19	322	--	--
2	766	232	8	466	168	36	460	135	29	598	228	38
3	1058	360	34	713	202	29	632	147	23	700	76	11
4	1563	465	30	946	212	22	788	143	18	879	157	18
5	1665	529	32	1017	173	17	906	136	15	976	146	15
6	1873	685	37	1376	413	30	1016	220	22	1208	305	25
7	2082	627	30	1481	332	22	1150	204	18	1389	168	12
8	2213	770	35	1693	279	16	1430	352	24	1386	227	16
9	2922	531	18	1604	378	24	1541	253	16	1682	267	16
10	2881	616	21	1997	564	28	1686	319	19	1779	283	16
11	3804	369	10	2002	302	15	1891	344	18	2052	285	14
12	3175	--	--	2448	324	13	2063	280	14	1987	349	18
13	4809	749	16	2562	432	17	2087	275	13	2347	--	--
14	4509	--	--	2654	483	18	2117	475	22	2577	130	5
15	--	--	--	2669	399	15	2508	553	22	2209	--	--
16	--	--	--	2935	586	20	2447	385	16	2784	423	15
17	--	--	--	3464	254	7	2510	303	12	2733	297	11
18	--	--	--	2945	138	5	2704	304	11	--	--	--
19	--	--	--	3716	608	16	3083	159	5	3141	279	9
20	--	--	--	3405	--	--	3114	270	9	3313	455	14
21	--	--	--	4785	--	--	3221	130	4	3512	116	3
22	--	--	--	4003	--	--	3635	325	9	--	--	--
23	--	--	--	3221	--	--	3727	--	--	4187	--	--
24	--	--	--	3980	488	12	--	--	--	--	--	--
25	--	--	--	5107	--	--	--	--	--	4187	--	--
26	--	--	--	--	--	--	--	--	--	3727	--	--
30	--	--	--	3773	--	--	--	--	--	--	--	--
32	--	--	--	5797	--	--	--	--	--	--	--	--
36	--	--	--	--	--	--	5199	--	--	--	--	--

As can be seen from the table, there is nothing surprising about the increasing of mean durations along the increase of utterance length. However, it is surprising that the four speakers used varying amounts of time producing a unit of similar length and vice versa. By far, A spent a lot more time than the other three among whom B's mean durations are in general longer than D's which in turn are longer than C's.

Another important thing about the table is that the variation of speech time within each utterance length is mostly small: cases in which variation coefficients are over 30 are few in number, suggesting that there is a limit as to how much time is spent on uttering a certain sequence of syllables. As the four speakers are compared, B fluctuates more than the others in this respect. So, it would be interesting to know exactly what causes the big variation (This issue will be discussed soon).

As we said before, the differences between the four means (total) are not very great, the greatest being 280 ms, the smallest only 42 ms. But, here the differences, especially those between A and the rest, are quite great, and these differences imply that articulation time is a variable subject to change from speaker to speaker. Each individual speaker has his own rate of speaking, some faster and some slower.

Taking the two seemingly contradictory results together--that the average articulation time remains rather constant from speaker to speaker (but the variation within the individual is still great) and that articulation time for utterances of different lengths varies individually, we begin to understand that while there is some constraint on the duration of the average utterance, within that duration the speaker can decide on how many syllables to utter. For example, Speaker A spoke five syllables within the duration of 1665 ms, but Speaker B did so within only 1017 ms, while C only 906 ms and D only 976 ms. Similarly, if we keep the duration constant, for example, 1500 ms, Speaker A could utter 4 syllables only, but 7, C and D each 9. This factor seems to be strong enough for Speaker A to have a much lower rate of articulation (1.97 syllables/second).

5. Discussions

We have just presented the results of our analysis, and have seen some consistency and some variations across the four samples in terms of our research questions. Let us now compare our results with those found for other languages.

5.1. Pauses and Speech Rate

As we totaled the four samples as a whole, our results for the speech rate, the articulaion rate and the percentage of pauses come closest to Goldman-Eisler's (1956), as shown in the following:

	SR	PP	AR
Goldman-Eisler (1956)	3.3	32.5	4.95
This Study	3.4	32.9	4.99

If we refer back to Table 2 in which the results of three studies are presented, we find that our results are closest to those of Duez's political speeches (SR=3.26, PP=37.3, AR=5.2).

What this comparison leads us is that on the one hand speech rate is a function of pauses and that there is some invariance in speech rate, articulation rate and the percenage of pauses across languages. But on the other hand, there are also perceptible differences, as can be clear in the following discussion.

5.2. The Duration of Pauses

In order to compare our results on the duration of pauses with those found by Goldman-Eisler (1961b), we need to reactegorize our durations in such a way as shown in the following table (we keep only the percentages).

Table 17. Percentages of Pause Durations Recategorized on the Basis of the Statistics in Figure 3

	<0.5 s	1.0 s	2.0 s	3.0 s	3-8 s
A	30.77	19.58	27.27	22.38	0
B	43.09	34.25	18.23	4.42	0
C	55.14	34.11	9.35	1.40	0
D	30.65	26.74	31.55	11.5	0

As we check the figures in this table with those in Table 3, we find that on the whole our four samples have more pauses longer than 2 seconds, but they do not have pauses longer than 3 seconds, probably constrained by the fact that the speakers are very

much experienced and may all know their subject matters very well. Our Samples B and C behave very much like Goldman-Eisler's Adult Discussions in that about half of all the pauses are below 0.5 second, and one-third between 0.5 and 1 second. Pauses over 2 seconds are of the minority. As we have already dicussed, Goldman-Eisler's Psychiatric Interview is a very special case in that the majority of pauses are between 1 and 2 seconds (a total of 62.5%), but it also has a larger percentage of pauses longer than 2 seconds (21%). Our Sample A and D do not have an exact similarity; however, Sample A does have a relatively large proportion of pauses over 2 seconds which is a result of the specific situation--a public preaching (Clemmer et al. 1979). Pauses in this very sample are perhaps rhetorical. However, there are other possible factors. For one thing, Speaker A happens to be the oldest of the four speakers. Age might contribute to the style of slow speech. Audience size could be another, for Speaker A was speaking to a larger size of audience than the other three.[15] Research has shown that the attendance of an audience can make a difference in eliciting more pauses of longer durations (Deese 1980). Still another factor might be the content of the speech. Sample A is about a typical topic in Buddhism--the life cycle. And honestly speaking, it is not a concept easy to explicate. Although both speaker B and D were also talking about Buddhism, one was doing it by telling a story about a historical well-known personae, and the other giving an account of his own experience in learning to meditate.

So the differences found in the durations of pauses in the four samples seem to be attributable to such factors as situation, age, audience size, and content. These same factors my account for the differences between this study and Goldman-Eisler's. Perhaps the language involved is a fastor too.[16]

5.3. The Distribution of Pauses

Let us first rearrange our results in Table 14 and put them side by side with those found by Duez (Table 6)

---

[15] The larger the audience size, the greater the formality, B&C were delivered in a less formal situation than A&D, thus resulting in more shorter pauses.

[16] Linguistic factors might include syntax, discourse, or the size of utterances. But we are unable to give any account for them at the present time.

Table 18. Summary Table of the Results on the Distribution of Pauses of Two Studies

		Between clauses		Between phrases		Within phrases	
		%	MD	%	MD	%	MD
Duez	PS	43.2	930	50.4	708	6.4	462
	PI	45.0	797	42.9	588	12.1	501
	CI	57.8	802	30.3	632	11.9	401
This	A	39.2	1179	44.8	1543	15.4	1312
Study	B	38.7	730	31.5	654	5.0	1495
	C	30.4	582	48.0	470	10.3	448
	D	45.2	1018	28.0	855	4.8	975*

* Our totals for percentages will not reach 100 because we are not including the percentages of pauses preceding sentences.

Thus, our results look very much like Duez's in that most pauses occur between clauses and between phrases, and those within phrases are of the minority. But Duez's mean durations are more homogeneous and reflect the influence of syntactic structure on them more systematically. None of our samples show such systematicity in the mean durations.

From this result, our prediction that before major constituents (sentences and clauses) pauses will be longer and at other places (phrases and words) they will be shorter is not very strongly supported. However, it is not difficult to understand the rather long durations of pauses preceding words, many of which are perhaps genuine hesitant pauses, unlike those preceding phrases or clauses which are more junctural in nature. And in the case of Speaker A, long pauses are not necessarily junctural or hesitant; in fact, in that very speech situation--a public preaching situation, they are more rhetorical than otherwise.

As we can see in Table 14, there are not many sentences separated by pauses in each sample. In other words, it seldom happens that one produces long sequences without pausing. The pauses that occur before sentences that are produced in one breath are not necessarily longer than those occurring elsewhere probably because these sentences are themselves easy to produce in this way. Not only do they not demand pausing in the middle of the production process, but also they do not need much pausing in the beginning either. Thus, if we want to look at the relationship between constituent structure and the length of pauses, we need to look at those sentences that run across

more than one utterance. Table 19 presents the number of complete sentences (five types)[17] and the mean durations of pauses preceding the utterances that begin sentences (These pauses are sentence-initial pauses). Along with them are the mean durations of pauses that precede utterances that do not begin a sentence (These pauses are the sentence-internal pauses).

Table 19. Frequencies of Different Types of Sentences and Mean Durations of Pauses Preceding Them

	Sample A		Sample B		Sample C		Sample D	
	F	MD	F	MD	F	MD	F	MD
Simple	27	1619	43	738	23	954	28	1547
Compound	5	2834	18	1034	8	696	16	1380
Complex	17	1204	23	832	24	851	24	1533
Comp-Compl	2	1702	12	613	6	652	15	1472
Frag	8	3169	5	920	5	865	3	1150
NSI*	84	1120	80	623	148	383	100	650

* NSI refers to those utterances that do not begin sentences.

From this table, it becomes very clear that all the pauses that precede sentence-initial utterances are longer (very much longer in some cases) than those preceding non-sentence-initial utterances in all the four samples. However, there is little consistency in the relationship between pauses and sentence types. It is very puzzling that compound-complex sentences do not necessarily demand longer pausing time before they are produced.[18]

The relationship between pausing and syntactic structure is even more impressive as we take a closer look at the sequences of speech sample. Table 20 presents one sequence from each sample, each sequence being a complete sentence.

[17] The number is also the number of pauses that precede the utterances.

[18] This could be due to the loose structural constraints in the Chinese language. The marking of coordination and subordination is usually not explicit. So, in identifying such structures, we use semantic rather than syntactic cues. This may account for our unexpected results.

Table 20. Pause Distribution Preceding the Utterances that Make Up A Complete Sentence

Sample A		Sample B		Sample C		Sample D	
*Ut No	MD	Ut No	MD	Ut No	MD	Ut No	MD
44	1748	57	782	28	920	3	1702
45	782	58	276	29	552	4	1426
46	276	59	276	30	138	5	690
47	322	60	368	31	460	6	138
48	966	61	552	32	552		
62	138	33	276				

*Ut No=utterance number; MD=mean pause duration

Patterns of this kind are very common, especially in Samples B, C, and D. It is as if the speaker has to pause more in the beginning of the sentence than at the middle or the end. The durations of pauses appear to be steadfastly decreasing (though straightforth decreasing is lacking) as the sentence is being produced.

This kind of declination for the durations of pauses can also be found in larger sequences than the sentence, as shown in the following table.

Table 21. The Pattern of Pausing Over Succeeding Sentences in Samples B, C, and D[19]

Sample B		Sample C		Sample D	
Ut No	PD	Ut No	PD	Ut No	PD
*27	3080	*116	782	*13	3037
28	644	117	230	14	1702
29	966	118	736	*15	874
*30	460	119	506	16	552
31	184	120	184	*17	644
32	276	*121	506	18	138
	122	138			
	123	230			

"*" indicates that this is the beginning of a sentence.

[19] Sample A is excluded because, as we have pointed out, it tends to have long pauses everywhere.

As to the pauses before or after conjunctions, we also counted the frequencies of four conditions, and the results are in the following table.

Table 22. Frequencies and Percentages of Pauses Before or After Conjunctions

	P + Conj		P + Conj + P		Conj + P		Conj (alone)	
	F	%	F	%	F	%	F	%
A	8	88.9	1	11.1	0	0	0	0
B	14	82.4	1	5.9	0	0	2	11.8
C	21	75	4	14.3	1	3.6	2	7.1
D	8	88.9	0	0	0	0	1	11.8

In other words, our results are pretty much like those of Stentröm's in that most conjunctions are preceded by a pause.

5.4. Pause-Defined Units

Our previous results on the size of pause-defined units are 5.48 syllables for A, 8.88 for B, 9.10 for C, and 8.49 for D. When the four samples are combined, we obtain an average utterance length of 8.17 syllables, and the average time spent on producing each utterance is 1.65 seconds. Table 23 lists our results along with those of other studies for other languages.

Table 23. Mean Length of Utterance and Mean Articulation Time of Four Studies

	Utterance Length	Duration
Chafe 1987 (English)	5 w	2 s
Stenström 1986 (Eng)	8 syl	1.3 s
Butcher 1981 (German)	4.1 w	2.4 s
This Study (Chinese)	8.17 syl	1.65 s

Clearly, the size of an utterance, whether defined pausally or intonationally (that is what other studies did), is pretty fixed, although we have already pointed out that

there is much variance between our four speakers and within each of them. Our longest utterances can be as long as 36 syllables although such long utterances are rare.

A big difference between such pause-defined units and the commonly adopted tone/intonation units is that the latter would never cover two sentences within one unit (Crystal 1969). In our four samples, we have found a total of 111 such instances. Such utterances are usually quite long because they are either compound or complex or compound-complex. As we did a simple calculation, we obtained an average of 10, 15.2, 17.2 and 14.6 syllables for each of the four samples, their range being 10-14, 6-32, 9-36, abd 9-26. Although they are not the only contributors to account for the long utterances found in the four samples (some clauses and phrases can be as long as 15 syllables), they are the major ones.

As we have said earlier, these long utterances are not necessarily preceded by long pauses. Even when they are clearly composed of more than one intonation unit, they do not need any pausing at the unit boundary. There are several possible reasons to account for such a phenomenon. For one thing, they are ready for production: no extra time for planning is needed. Secondly, the components within them are so tightly related that they prefer to be produced in one breath. Finally, the speaker, when producing such long utterances, is perhaps under some kind of pressure (psychological or situational) that he needs to say them without stop.[20]

Although our distribution of pauses of varying durations is very much like Duez's (1982), many of our words are the points of hesitancy if defined intonationally. This is why we often have very long pauses before words.

In Section 4.3.4, we have mentioned that there is some variation in speech time within the same utterance length and that a certain speaker may utter a larger sequence of syllables within the same amount of time than others. It is now time to see what causes such variations within and between speakers.

Let us try to pick out from our four samples the utterance size of 6 syllables where the variation coefficients are greatest. The reaults of mean durations and ranges are as follows: 1873 ms 1012-2991 ms for A, 1376 ms, 828-2301 ms for B, 1016 ms, 644-1380 ns for C, 1208 ms, 874-2025 ms for D. In order to see the contrast in durations, we present three longest utterances (in terms of durations) and three shortest ones from each sample in the following.

[20] It is also possible that the morpholopical structure of Chinese makes it easier to compress words tightly together.

Sample A:

1. 見與煩惱是因 (2761)
   jian yu fan-nau shi yin
5. 關係最重要的 (2991)
   kuan-xi zui zhong-yau de
77. 死了就不得了 (2807)
   si le jiu bu de liau
30. 看出看得出來 (1150)
   kan chu kan de chu-lai
107. 解決這個問題 (1104)
   jie-jue zhe-ge wen-ti
108. 啊這個他來了 (1012)
   a zhe-ge ta lai le

Sample B:

38. 嘿嘿一片白雪 (1979)
   hei hei yi pian bai xue
76. 貪字貧之殼啊 (2301)
   tan zi pin zhi ke a
158. 衣食僅能自充 (1887)
   yi shi jin neng zi chong
15. 那個心歷路程 (920)
   na ge xin lu li cheng
73. 如果再怎麼樣 (1012)
   ruguo zai zemeyang
129. 這麼好的風水 (920)
   zheme hao de fongshui

Sample C:

29. 我們科學發展 (1104)
   wo-men ke-xue fa-zhan
176. 當太陽在這個 (1380)
   dang tai-yang zai zhe-ge
206. 因為呢講很多 (1334)
   yin wei ne jiang hen duo
106. 沒有甚麼道理 (782)
   mei you she mo dao li
169. 一百二十公尺 (644)
   yi bai er shi gong chi
171. 提二十八公尺 (828)
   ti er shi ba gong chi

Sample D:

38. 仁慈悲心發功 (1795)
   ren ci bei fagorg
87. 大概十幾分鐘 (1656)
   dagai shiji fenzhong
131. 你的情緒煩惱 (2025)
   ni de qingxu fannao
71. 你這種如果你 (828)
   ni zhejong ruguo ni
85. 那個力量給它 (874)
   zhege liliang geita
143. 那大家不知道 (874)
   na dajia bu zhidao

The number preceding the utterance is the sequence number in the sample, and the number after the utterance is the duration (in ms). As we examine these utterances, we find that syntax, semantics, and discourse structure all play a role in deciding their durations. If an utterance involves syntactically important categories, such as nouns or verbs, or if it is semantically important, or if it expresses some information that

is to be paid attention to, then it is usually uttered very clearly and slowly. On the other hand, in a situation when the utterance does not carry much information, some parts of the utterance, especially function words, are often spoken very weakly and quickly. For example, although C's 176 includes the function word 這個 (this) and it is a subordinate clause, because the whole utterance is in the sentence initial position, it becomes an attention caller, therefore produced quite slowly. Numbers are usually quite redundant (Taylor & Taylor 1990), so C's 169 and 171 are spoken quite fast. Also, when citation is involved as in B's 76 and 158 (both are classical Chinese), then the utterance is produced slowly. Although D's 143 does not have too many function words, because it serves as a concluding remark, it is uttered more quickly than it should be.

In a word, while the length or the duration of an utterance may be an easy thing to calculate, the variation found in actual articulation time can best accounted for by syntax, semantics and discourse, as reflected in the pause pattern.

6. Summary and Conclusion

Our present study focuses on the pause pattern of Mandarin Chinese by using four samples of spontaneous monologues produced in public situations by four male speakers aged between 40 and 60. Our major findings include the following.

First, on the whole, pauses and articulation rate are closely related; more pauses mean lower articulation rate. In other words, articulation rate is a function of pauses.

Second, although the articulation rates we have found for the four samples as a whole and individually are mostly within the range as found for other languages, our sample A has very low articulation and speech rates, a fact that illustrates the particularity of this speaker. Not only does he pause a lot, but also he pauses longer than the other three. Frequent and long pauses do not necessarily mean more hesitation. In fact, Speaker A's long and frequent pauses have been speculated to be largely rhetorical and attributable to situation, audience size, his seniority, and the content of his speech. This very instance prompts us to say that while articulation rate is a function of pauses, both speech rate and articulation rate are also affected by the way the speaker utters the syllables.[21]

In terms of pausing time, while the overall percentages of pauses are not greater

[21] What is more, he lengthend many of his syllables a lot, a phenomenon commonly found in reading classical Chinese. But we did not look into this matter in the present paper.

than those for other languages, our four speakers seem to have a longer pause duration than the subjects studied by others. Our results seem to suggest that the style of pausing is a relative matter. A pause of certain duration may be quite long for one speaker and yet quite short for another. Besides, it might function as a juncture pause at one location and as a hesitation at another. Therefore, it is wrong to think of the duration and the function of pauses as fixed. What lies beneath this apparent fact is not clear at the moment. Is this a trait of the speakers or of the language -- Chinese? This is something we can brood over in the future.

Although there is much variation in articulation time and pause time, there seems to be some universal principles that constrain the amount of time for each utterance. This applies to all four speakers even when Speaker A spoke far more slowly (with fewer syllables in one utterance).

Though pauses of varying lengths have been found to be dependent upon constituent structures, our results do not lead to such a strong claim. Yet on the whole, we find longer pauses when a new topic is initiated or when a point is stressed or when a sentence begins, and shorter pauses before phrases and clauses. And yet there is no consistency between pauses and subordination/coordination strcutures (shorter pauses are to be found for subordination and longer ones for coordination). The reason could lie in the language itself: Chinese is not highly marked. Therefore, our categorization in terms of sub-/co-ordination may not be adequate enough.

Since we used pauses to segment our sampled speech (in order to avoid the difficulty and subjectivity involved in identifying intonation units), we are not too surprised to find that sometimes an utterance can be as long as 36 syllables involving more than one clause or sentence or intonation unit. It this way, we can more easily see when speech is fluent and when it is not. And yet our units do not differ very much in size and duration from the usual intonation units studied by other researchers.

So, on the whole, most of the findings of the present study have borne out our initial assumptions about the matter of speech rate, the distribution, and frequency, and the duration of pauses and the structure and process of the utterance. However, we also have some unexpected results due to our particular samples (or the language we study). Although by using pauses as single delimiters of speech, we run into the problem of categorization, we are still very happy that our pauses have revealed much about the structuring of speech in the Chinese language. We might conclude: because of pauses, we find speech an interesting object to study.

REFERENCES

Barik, H. C. 1968. On defining juncture pauses: a note on Boomer's "Hesitation and grammatical encoding." *Language and Speech,* 11, 156-159.

Beattie, G. 1980a. The role of language production processes in the organization of behavior in face-to-face interaction. In Butterworth, B. ed. *Language Production,* 69-107. New York: Academic Press.

Beattie, G. 1980b. Encoding units in spontaneous speech: some implications for the dynamics of conversation. In Dechert & Raupauch, eds, pp. 131-143.

Boomer, D. S. 1965. Hesitation and grammatical encoding. *Language and Speech,* 8, 148-158.

Boomer, D. S. 1966. Review article--rapport critique, *Lingua,* 25, 152-164.

Boomer, D. S. & Dittman, A. T. 1962. Hesitation pauses and juncture in speech. *Language and Speech,* 5, 215-220.

Brown, G., K. L. Currie, & J. Kenworthy. 1980. *Questions of Intonation.* London: Croom Helm.

Butcher, A. 1980. Pause and syntactic structure. In Dechert and Raupauch, eds, pp. 85-90.

Butcher, A. 1981. Aspects of speech pause: phonetic correlates and communicative functions. Institut für Phonetik, Universität Kiel, Arbeitsberichte Nr. 15.

Butterworth, B. 1975. Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research,* 1, 75-87.

Butterworth, B. ed. 1980. *Language Production. Vol. 1. Speech and Talk.* London: Academic Press.

Chafe, W. 1980. Some reasons for hesitating. In Dechert & Raupach, eds. pp. 169-180.

Chafe, W., ed. 1984. *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production.* Norwood, N. J.: Ablex.

Chafe, W. 1987. Cognitive constraints on information flow. In R. S. Tomlin, ed. *Coherence and Grounding in Discourse,* pp. 21-51. John Benjamins.

Chafe, W. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing.* Chaicago: The University of Chicago Press.

Clemmer, E. J., D. C. O'Connell, & W. Loui. 1979. Rhetorical pauses in oral reading. *Language and speech,* 22, 397-405.

Cruttenden, A. 1986. *Intonation.* Cambridge: Cambridge: Cambridge University Press.

Crystal, D. 1969. *Prosodic Systems and Intonation in English.* London: Cambridge University Press.

Crystal, D. & D. Davy. 1975. *Advanced Conversational English.* London: Longman.

Dechert, H. W. & M. Raupach, eds. 1980. *Temporal Variables in Speech: Studies in Honor of Frieda Goldman-Eisler.* The Hague: Mouton.

Deese, J. 1980. Pauses, prosody and the demands of production in language. In Dechert & Raupauch, eds, pp.69-84.

Du Bois, J. W., S. Schuetze-Coburn, D. Paolino, & S. Cumming. 1992. *Santa Barbara Papers in Linguistics, vol. 4. Discourse Transcription.* Dept of Linguistics, University of California at Santa Barbara.

Duez, D. 1982. Silent and non-silent pauses in three speech styles. *Language and Speech,* 25, 11-28.

Duez, D. 1985. Perception of silent pauses in continuous speech. *Language and Speech,* 28:4, 377-389.

Garman, M. 1992. *Psycholinguistics.* Cambridge: Cambridge University Press.

Goldman-Eisler, F. 1954, On the variability of the speed of talking and on its relation to the length of utterances in conversations. *British Journal of Psychology* (General Section), 45, 94-107.

Goldman-Eisler, F. 1956. The determinants of the rate of speech output and their mutual relations. *Journal of Psychosomatic Research,* 1, 137-143.

Goldman-Eisler, F. 1961a. The significance of changes in the rate of articulation. *Language and Speech,* 4, 171-174.

Goldman-Eisler, F. 1961b. The distribution of pause durations in speech, *Language and Speech,* 4, 232-237.

Goldman-Eisler, F. 1967. Sequential temporal patterns and cognitive processes in speech," *Language and Speech,* 10, 122-132.

Goldman-Eisler, F. 1968. *Psycholinguistics: Experiments in Spontaneous Speech.* London: Academic Press.

Goldman-Eisler, F. 1972. Pauses, clauses, sentences. *Language and Speech,* 15, 103-113.

Grosjean, F 1980. Linguistic structures and performance structures: studies in pause distribution. In Decher & Raupauch, eds, pp. 106.

Grosjean, F. & A. Deschamps. 1975. Analyse contrastive des variables temporelles de l'anglais et du francais: vitesse de parole et variables componantes, phenomenes de'hesitation. *Phonetica,* 31, 144-184.

Grosjean, F. & Lane, H. 1977. Pauses and syntax in American sign language. *Cognition,* 5, 101-117.

Grosjean, F. & Collins, M. 1979. Breathing, pausing and reading. *Phonetica,* 36, 98-114.

Grosjean, F., Grosjean, L. & Lane, H. 1979. The patterns of silence: performance structures in sentence production. *Cognitive Psychology,* 11, 58-81.

Halliday, M. A. K. 1967. *Intonation and Grammar in British English.* The Hague: Mouton.

Hawkins, P. R. 1971. The syntactic location of hesitation pauses. *Language and Speech,* 14. 277-287.

Henderson, A., F. Goldman-Eisler, & A. Skarbek. 1965a. The common value of pausing time in spontaneous speech. *Journal of Experimental Psychology,* 17, 343-345.

Henderson, A., F. Goldman-eisler, & A. Skarbek. 1965b. Temporal patterns of cognitive activity and breath control in speech. *Language and Speech,* 8, 236-242.

Henderson, A., et al. 1966. Sequential temporal patterns in spontaneous speech. *Language and Speech,* 9, 207-216.

Hieke, A. E., S. Kowal, & D. C. O'Connell. 1983. The trouble with articulatory pauses. *Language and Speech,* 26:3, 203-214.

James, W. 1980. *The Principles of Psychology.* 2 vols. New York: Henry Holt. Preprinted 1950 by Dover Publications, New York.

Lehiste, I. 1979. Perception of sentence and paragraph boundaries, in *Frontiers of Speech Communication Research,* pp. 191-201. London: Academis Press.

Lounsbury, F. G. 1954. Transitional probability, linguistic structure, and system of habit--family hierarchies. In C. E. Osgood & T. A. Sebeok, eds. *Psycholinguistics: A Survey of Theory and Research Problems,* pp. 93-101. Baltimore: Waverly Press.

Maclay, H. & C. E. Osgood. 1959. Hesitation phenomena in spontaneous English. *Word,* 15, 19-44.

Martin, J. G. 1970. On judging pauses in spontaneous speech, *Journal of Verbal Learning*

and Verbal Behavior, 9, 75-78.

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

Martin, J. G. & Strange, W. 1968a. Determinants of hesitations in spontaneous speech. *J. Exp. Psychol,* 76, 474-479.

Martin, J. G. & Strange, W. 1968b. The perception of hesitation in spontaneous speech. *Perception and Psychophysics,* 3, 427-438.

O'Connell, D. C. 1977. One of many units: the sentence, in S. Rosenberg, ed. *Sentence Production: Development in Reseearch and Theory,* pp. 307-313. Hillsdale, N.J.: Eribaum.

O'Connell, D. C., S. H. Kowal, & H. Hörmann. 1969. Semantic determinants of pauses. *Psychologische Forschung,* 33, 50-67.

O'Connell, D. C. & S. Kowal. 1983. Pausology. In W. A. Sedelow & S. Y. Sedelow, eds. *Computers in Language Research,* pp. 221-299. Berlin: Walter de Gruyter.

Osser, H. & F. Peng. 1964. A cross cultural study of speech rate. *Language and Speech,* 7, 120-125.

Rochester, S. R. 1973. The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research,* 2, 51-81.

Schuetze-Coburn, S., M. Shapley, & E. G. Weber. 1991. Units of intonation in discourse: a comparison of acoustic and auditory analyses. *Language and Speech,* 34:3, 207-234.

Stenström, A. 1986. A study of pauses as demarcators in discourse and syntax. In Willem Aarts-Jan & Meijs, ed. *Corpus Linguistics: New Studies in the Analysis and Exploration of Computer Corpora,* pp. 201-219. Amsterdam: Rodopi.

Taylor, I. & M. M. Taylor. 1990. *Psycholinguistics: Learning and Using Language.* Englewood Cliffs, N. J.: Prentice-Hall.

Trager, G. L. & H. L smith. 1951. *An Outline of English Structure.* Studies in Linguistics, Occasional papers, 3. Washington: American Council of Learned Societies: Norman, Okla: Battenburg Press.

Yang, I-li. 1996. Silence is gold: an introduction to pausology. *NCCU Research Bulletin,* 6, 43-77.

Yule, G. 1980. Speaker's topics and major paratones. *Lingua,* 52, 33-47.