The Prosodic Status of Breaks in Running Speech: Examination and Evaluation

Chiu-yu Tseng

Institute of Linguistics (Preparatory Office), Academia Sinica
Taipei, Taiwan 115
cytling@sinica.edu.tw

Abstract

Recent investigations of prosody organization appeared to focus more on how best to determine operating prosodic units and how these units are put together. However, the prosodic role of breaks/pauses in running speech has not received due attention. This paper shows that breaks in running speech functions as a prosodic unit on their own, are perceptually significant in a systematic manner and therefore should be considered as a necessary feature and cue in the organization of speech prosody.

1. Introduction

By linguistic definition, prosody includes stress at the lexical level, intonation at the phrase or sentence level, and rhymes and beat at the poetic level. Acoustically, these phenomena involve duration, frequency height and amplitude. Compared with segments, the relative nature of the physical aspects has put them in a secondary position of existing theoretical frameworks, phonetic and phonological alike. In comparison, breaks in speech flow have received even less attention and as a result has not been considered as a possible operating speech unit. This came as no surprise since existing methodology employed in phonetic and phonological investigations tended to focus on discrete units such as segments, syllables, words, phrases and at most syntactic sentences rather than long stretches of running speech. In other words, though speech instead of written documents has been the main focus of linguistic investigation for more than half a century now, the way speech is approached has remained at the level of speech segments removed from the way speech is produced in the most natural sense. Furthermore, the very fact that breaks in speech flow involve no physical correlates may also be the main reason of their insignificance as a linguistic/phonetic unit. However, with the availability of speech database, the length of utterances under investigation is no more a concern, the envelope of units could increase without limitation, and the complexity of characterizing speech prosody and prosodic properties increased. Previous existing definition no longer sufficed. We see breaks in running speech as a main feature of speech prosody. Drawing evidence from digitized speech data and examining the signals in more detail, a revisit of prosody organization of running speech is only natural.

Following the ToBI framework to examine speech phenomena in layers, our examination of speech prosody included the following levels, namely, break/pause, emphasis, speech rate, volume, frequency height and frequency range. The present study focuses on how we examined in detail breaks in running speech and how they related to the overall flow of speech. We argue for their indispensable role in prosody organization and their lawful status in speech prosody in general.

We have collected three types of speech database of read Mandarin Chinese since 1994 for various purposes. The first speech database is a phonetically balanced corpus of 599 utterances ranging from 2 to 180 syllables/characters in length. We also controlled word frequency factor by drawing tokens from text corpus developed at CKIP, Institute of Information Science, Academia Sinica [1]. The second speech database is a corpus of 16545 prosody-oriented utterances ranging from 5 to 134 syllables/characters in length. Three syntactic sentence types were included, namely, declarative (805 total), exclamatory (303 total) and interrogative (546 total) utterances. The third speech database is a corpus of 161 stress-balanced utterances that were composed of high frequency words also. The control for this database is stress pattern. Words consisted of 2 to 5 syllables/characters with all possible stress patterns were selected to make up utterances ranging from 9 to 66 syllables/characters in length.

We noticed while analyzing our first phonetically balanced speech database that the flow of the collected natural speech was broken into perceptually identifiable fragments by pauses of various duration that did not always correspond to punctuation marks in the text. In fact, there were more such breaks in speech flow than punctuations in the corresponding text. We began to examine these phenomena on the bases of the breath-group theory [2], and subsequently designed a system to labeling levels of prosodic properties in speech flow [3, also Table 1]. The system is based on the physiological constraint of breathing during speech. Moreover, we also postulated that the largest possible speaking unit is a prosodic group that consists of at least one breath-group instead of a syntactic sentence. Such a prosodic group could consist of more than one breath-group also. [4, 5, 6, 7,]

We will present in this paper some detailed analyses of breaks in the 161 stress-balanced utterances and argue that breaks should be considered as one of the most significant features in speech prosody.

2. Methodology

The present investigation consists of break/pause analyses of 161 stress-balanced utterances ranging from 9 to 66 syllables/characters in utterance length. One male and one female native speaker of Mandarin Chinese, both in their mid-20's and college educated, read the text in sound-proof chambers through a microphone. The total recording time was 48 minutes. 244MB of digitized speech data was collected. The speech files were listened through headsets and manually labeled by 3 independent trained transcribers.

2.1. Perception Based Manual Labeling of Breaks in Running Speech

2.1.1. The Labeling System

3 transcribers listened to speech files and manually labeled the breaks/pauses they heard. During this task, the transcribers were asked to listen and label only the perceived breaks in speech flow. Decision of breaks was made on their perception rather on the wave files shown on the monitor. Positioning the curses at zero crossing of the sound files to denote the exact duration of heard breaks was required. Table 1 shows the system of 6 levels of prosodic labeling.

Table 1: Labeling system used to categorize heard breaks in running speech.

Level / Tags	0	1	2	3	4	5
BREAK	reduced syllabic boundary	normal syllabic boundary	minor-phrase boundary	major-phrase boundary	breath group boundary	prosodic group boundary
EMP	reduced	normal	moderate	strong		
RATE	very slow	slow	normal	quick	very quick	
VOLUME	very low	low	normal	high	very high	
PITCH	very low	low	normal	high	very high	
RANGE	very small	small	normal	large	very large	

2.1.2. Transcription Consistency

The aims of the manual transcription were to test if two kinds of transcription consistency could be achieved. One is inter-transcriber consistency, another intra-transcriber consistency. To achieve intra-transcriber consistency, a two-week training period was necessary for each transcriber. During this period, a transcriber would first learn the transcription system and then transcribe about 40 to 50 utterances. Each transcriber was then asked to re-label all the labeled utterances from the beginning to inspect if s/he agreed with the first version of transcription. By this time each transcriber has mastered the labeling system as well as the labeling task and would make revisions of their previous transcriptions. To achieve intra-transcriber consistency, all three transcribers were asked to label the entire corpus of 161 utterances. Their labeling was then compared on weekly basis. Table 2 shows the comparison of labeling results between transcribers.

Table 2.1. Comparison of consistency between two transcribers' labeling results of the speech produced by the male speaker

chi\lit	B1	B2	B3	B4	B5
B1	2878	93	4	1	0
B2	97	1651	19	0	0
B3	2	22	306	2	0
B4	1	0	5	46	0
B5	0	1	0	0	155
chi\lit	B1	B2	B3	B4	B5
B1	97%	3%	0%	0%	0%
B2	6%	93%	1%	0%	0%
B3	1%	7%	92%	1%	0%
B4	2%	0%	10%	89%	0%
B5	0%	1%	0%	0%	99%

Table 2.2. Comparison of consistency between two transcribers' labeling results of the speech produced by the male speaker.

chi\rub	B1	B2	B3	B4	B5
B1	2830	108	2	0	0
B2	144	1584	24	0	0
B3	1	26	299	3	0
B4	0	0	6	44	0
B5	0	0	0	0	155
chi\rub	B1	B2	B3	B4	B5
B1	96%	4%	0%	0%	0%
B2	8%	90%	1%	0%	0%
B3	0%	8%	91%	1%	0%
B4	0%	0%	12%	88%	0%
B5	0%	0%	0%	0%	100%

Table 2.3. Comparison of consistency between two transcribers' labeling results of the speech produced by the male speaker.

lit\rub	B1	B2	B3	B4	B5
B1	2867	74	1	1	0
B2	105	1621	23	0	1
B3	2	23	303	4	0
B4	1	0	4	42	0
B5	0	0	0	0	154
lit\rub	B1	B2	B3	B4	B5
B1	97%	3%	0%	0%	0%
B2	6%	93%	1%	0%	0%
B3	1%	7%	91%	1%	0%
B4	2%	0%	9%	89%	0%
B5	0%	0%	0%	0%	100%

2.2. Results

Various analyses of the labeled breaks were performed. The results obtained are presented below.

2.2.1. Mean Length and Range of Utterance

A mean analysis was calculated of the 161 utterances. Table 3 shows the mean length of utterances in both syllable numbers and in duration in ms.

Table 3.Mean length and range across utterance in syllable numbers and in duration

	Syllable # / ms
M of utterance	33.7 / 8300
SD	12.8 / 3200

Range	66 ~ 9 / 16255~2214

Note that the mean length of utterance is 34 syllables or 8300 ms. However, note also that an utterance of 9 syllables is only 3200 ms whereas an utterance of 66 syllables 16255 mesc.

2.2.2. Mean Length and Range of Breaks

A mean analysis of length of each kind of break in our system was also calculated. Table 4 shows the results.

Table 4. Mean length and range of labeled breaks in the corpus.

	M / SD (ms)	longest/ shortest (ms)
B2	12 / 16	91 / 9
B3	346 / 200	810 / 10
B4	607 / 112	890 / 397

Note the difference in duration from B2 to B3 is 334 ms, and from B3 to B4 261 ms.

2.2.3. Forward Mean Length of BreakX to the Next

BreakX

In an attempt to see whether these breaks in running speech is in any way related to prosodic units that a speaker may plan ahead in accordance with breathing, we performed forward calculation of the length in ms for each break to the next same break. Namely, the duration from each breakX to the next breakX is measured. Table 5 shows the results.

Table 5. Forward mean length of breakX to breakX.

	M (Syllable # / ms)
B2	2.08 / 514
B3	7.89 / 1941
B4	13.15 / 3237

2.2.4. Backward Mean Length of BreakX to the Next BreakX

We also performed in the linear sense backward calculation the labeled breaks in the same manner we did in forward measurements. Table 6 shows the results.

Table 6. Backward mean length of breakX to breakX.

	M (Syllable # / ms)
B2	1.99 / 491
B3	3.26 / 803
B4	4.29 / 1056

2.2.5. Measurements of Inclusiveness for Each Break

Because our data showed that each utterance as a speaking unit could consist of more than one breath-group, we made an inclusive measurement of each break labeled. For example, a B4 is expected to include one or more B3 within itself, a B3 one or more B2 within itself, etc. Table 7 shows the analysis.

Table 7. Mean number of breaks found within each break.

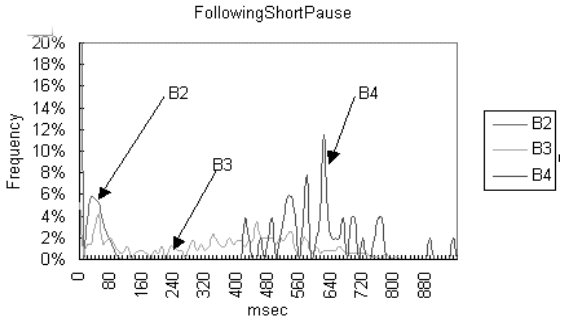
	M
# of B4 in B5	1.33
# of B3 in B4	1.82
# of B2 in B3	4.03

Note that the mean number of B2 within each B3 is 4. That means a B3 could be seen as made up of 4 B2's; a B4 of 2 B3's, and a B5 of almost 1.5 B4's.

2.2.6. Distribution of Labeled Breaks by Duration

We also plotted the respective distribution of the labeled breaks. The results are shown in Figure 1

Figure 1 Distribution of break duration



3. Discussion

We would like to point out some features of our prosody investigation that we consider worthy of attention before we discuss the results above. Firstly, the breaks under discussion are breaks occurred in speech flow of relatively long stretches of running speech. This allowed us to study speech prosody in units larger than syntactic sentences. Secondly, note that though the envelope of our investigation usually exceeds syntactic sentences and covers a short discourse, we were looking to see if a discourse is broken into more than one prosodic unit. In other words, we assume that prosody is an independent linguistic level during speech. The units involved may or may not correspond to other linguistic units. Furthermore, speech flow and running speech are synonymous in this report and were used interchangeably.

Our study shows that by analyzing breaks in running speech as a prosodic feature alone, the following results are found 1.Our transcription based on perceived breaks demonstrates that consistency within and across listeners can be achieved without difficulty. 2. The consistency also shows that breaks are perceived in a systematic manner. Furthermore, our analyses of breaks show that 3.the mean length of utterance is 38 syllables/characters (or 8300 ms) within the range of utterances from 9 to 66 syllables. The mean length clearly exceeds the range of a normal syntactic sentence and is proof that speaking units are usually larger than syntactic sentences. 4. Each break denotes the boundary of a corresponding prosodic unit in speech flow and is an inseparable as well indispensable unit in speech prosody. However, other prosody systems [8, 9] focus on prosodic units 5. Breaks in speech are hierarchical. The bigger the break number is, the more inclusive it is of breaks of smaller number.

6. These breaks also correspond to various prosodic units. For example, the B2 in our system (termed minor phrase boundary) corresponds to the boundary of prosodic word in other systems [8]; B3 (major phrase boundary) to prosodic phrase; B4 (breath-group) to utterance; and finally B5 (prosodic group) which some systems also took up [9] or sometimes called utterance group. In short, breaks in running speech are a salient feature; patterns can be found by analyzing breaks in running speech along. [10]

Our results also showed (see Table 4) that in spite of the wide range of duration within each break in speech flow, the mean of these breaks are of great difference and are definitely significant in perception. Note that the difference between the mean of B2 (12 ms) and B3 (346 ms) is 334 ms, and 261 ms between B3 (346 ms) and B4 (607 ms). Moreover, the difference between the mean of the shortest break (B2 12 ms) and the longest (B4 607 ms) is 595 ms. The distribution analysis of these breaks in Figure 1 further shows that in terms of duration, there is no overlap between the range of B2 and that of B4. In other words, B2 and B4 could be distinguished by duration factor alone. Note also that although B3 demonstrated the widest range of duration, indicating that duration may not be the single most salient factor to characterize B3, the consistency shown in Tables 2 proved that it is a definitely an identifiable cue in perception. In other words, B3 is not at all confusing in the perceptual sense in the sense that regardless of its duration variation, it was always perceived consistently as a cue. Rather, B3 can be seen as a strong evidence for the existence of breaks as linguistic cues of systematic nature in speech prosody.

4. Conclusions

Breathing imposes the most natural physiological constraint while speaking [2]. But this constraint does not necessarily correlate with syntactic sentence as the largest unit while speaking. Breaking running speech into various prosodic units by inserting pauses here and there is the single most important feature of perceived naturalness in the flow of speech. And no doubt a feature of speech prosody in its own right. Yet, breaks in running speech have always been taken for granted. Simply imagine how speech would flow without breaks .Recent studies in speech synthesis made clear that break insertion is necessary to make speech output more natural. However, how speakers adjust, plan and program their breathing while speaking has received very little attention. The proposed break labeling system is in fact a study of how we as speakers breathe while speaking, where we breathe, and how we plan our breathing in accordance with speaking. In short, we all speak in prosodic units identifiable by breaks in the flow of speech. However, one breathing cycle may not always allow enough time for a speaker to express a topic or theme. Therefore it's only natural that a speaker breathes in a complete cycle, i.e. a breath-group, when running out of air, and takes longer breath after completing a thought unit and before he begins another unit, i.e. a prosodic group. The end of a prosodic group is always a breath group signaling an end of some kind of expression. The end of a breath group may or may not always be a prosodic group. A prosodic group is a larger speaking unit in running speech and such units are necessary. In short, the organization of speech prosody should take into consideration the breaks that are necessary in speech flow because they are the keys to the tone of voice and rhythm of speech; they are the key to the naturalness of speech flow. And that is what speech prosody is about. Without these breaks, there will be no prosody in natural running speech..

5. References

[1] Corpus-Based Frequency Count of Characters, 1993. *Corpus-Based Research Series No. 1*, Chinese Knowledge Information Processing Group (CKIP). Institute of Information Science, Academia Sinica, Taipei, Taiwan. (`http://godel.iis.sinica.edu.tw/CKIP/`)

[2] Lieberman, P., 1976. Intonation, Perception and Language. *Cambridge University Press*. Cambridge, UK.

[3] Tseng, C., 1997. Prosodic group: Suprasegmental characteristics of Mandarin Connected speech from a speech database. *The Sixth International Conference on Chinese Linguistics (ICCL-6).* Leiden, the Netherlands.

[4] Chou, F.; Tseng, C.; Lee, L., 1998. Automatic segmental and prosodic labeling of Mandarin speech. *International Conference on Spoken Language Processing.* Sydney, Australia.

[5] Tseng, C., 1999. Investigating Mandarin Chinese prosody through speech database. *Oriental COCOSDA workshop*, (May 12-14, 1999), Academia Sinica, Taipei, Taiwan, R.O.C. 65-68.

[6] Tseng, C.; Chou, F., 1999. Machine readable phonetic transcription for Chinese dialects spoken in Taiwan. *The Journal of Acoustical Society of Japan (E).* Vol. 20, No. 3: 215-223.

[7] Tseng, C.; Chou, F., 1999. A prosodic labeling system for Mandarin Chinese speech database. *Proceedings of the XIV International Congress of Phonetic Science.* San Francisco, USA. 2379-2382.

[8] Chu, M., 2001. Prosody investigation and naturalness in speech synthesis. *Proceedings of the 5th National Conference on Modern Phonetics*. Tsing Hua University, Beijing, China. (`http://www.tup.tsinghua.edu.cn`) 295-301 (in Chinese).

[9] Wang, R.; Hu, Y.; Li, W.; Ling, Z., 2001. A large vocabulary Mandarin speech synthesis system based on decision trees. *Proceedings of the 6th National Conference on Man-Machine Speech Communication (NCMMSC-6, Nov. 19-24).* Shenzhen, China. 183-187 (in Chinese).

[10] Tseng, C., 2001. Prosodic cues and features in speech flow. *Proceedings of the 6th National Conference on Man-Machine Speech Communication (NCMMSC-6, Nov. 19-24).* Shenzhen, China. 169-172 (in Chinese).