

PROSODIC FEATURES OF MANDARIN REPAIR IN CLASSROOM LECTURE SPEECH

Helen Kai-yun Chen, Wei-te Fang, Chiu-yu Tseng

Phonetics Lab, Institute of Linguistics, Academia Sinica, Taipei, Taiwan

{helenkychen, weitefang, cytling}@sinica.edu.tw

ABSTRACT

The current study examines some acoustic features of the *reparandum* (R1) and *reparatum* (R2) of Mandarin repair in university classroom lectures with respect to discourse structure and information weighting. By the comparison of *contrast degree* in acoustic features between R1/R2 of the repair and at boundaries of the prosodic phrase without repair, it is found that almost 60% of repair instances present a significant difference from non-repaired phrase boundaries in terms of acoustic cues including F0, duration and intensity. Further pairing of *contrast degree* in acoustic features with pause duration reveals that prosodically repair instances and boundaries of prosodic phrases without repair yield significant distinctions. The findings thus demonstrate that the speaker would orient to smaller acoustic degree of contrast while repairing, which in turn may suggest alternative planning and deploying for information highlights at higher level of discourse structure during repair in the continuous speech production.

Index Terms— Mandarin repair, speech prosody, discourse-prosodic boundary, discourse structure, information weighting

1. INTRODUCTION

Repair as a commonly occurring phenomenon in spontaneous speech functions to reflect speakers' online planning during speech production, i.e. to signal how speakers adjust, correct, qualify or elaborate what they have or are going to say. When speakers carry out these actions during speech production through repair, noticeable by listeners or not, most of the time the on-going speech signals would be disrupted. Such disruption can be reflected in the prosodic aspects of the speech signal including sudden stops, pauses, sound stretches, or even obvious changes in the speaking rate. While discussions of repair in previous studies have sometimes been included under the general theme of *disfluency*, the phenomena including pauses, fillers and repairs are deemed as 'noises' in speech signals and hence one of the factors that hinder processing of the speech. A commonly used method to solve the problem of repair/disfluency as noises is to eliminate the repair directly from signals in order to restore the fluency of the speech strings (cf. [1], [2], [3]).

One of the drawbacks of the repair eliminating method lies in that, most of the past studies on the prosodic features of repair have paid attention to the *reparandum* (hence R1) and *reparatum* (hence R2) themselves only [4], [5]. The general practice is to lift them directly out of speech flow as discrete speech units; their acoustic properties examined and measured at face value. It has been proposed, however, that filled pauses may carry information about larger discourse units and thus exhibit different prosodic features when locating at boundary breaks [6]. Since filled pauses are but one of the features specific to spontaneous speech, other features merit similar consideration. We are therefore interested in finding out whether positive correlations of repair in relation to discourse units and boundaries could be found in acoustic patterns. If so, the findings should help explain how spontaneous specific features are incorporated into the planning and production of continuous speech.

The focus of the current paper, therefore, is to examine the prosodic features of Mandarin repair by concentrating on the *relative* prosodic realization between R1/R2 of the repair in terms of the prosodic features at boundaries of prosodic units within a hierarchical discourse framework [10]. Examples of repair were annotated for 2-hour of spontaneous speech in form of classroom lectures. We first identify the corresponding prosodic units from the framework that bear instances of repair from the lecture speech. By the method of calculating relative *contrast degree* of acoustic realization between R1/R2 of the repair [7, 8] and comparing results against the acoustic contrasts at boundaries of corresponding prosodic units without repair, we aim at clarifying if the supra-segmental correlates *F0*, *duration*, *intensity*, and *pause* realized while repairing can further distinguish repair from boundaries of non-repaired prosodic units. We believe the examination of acoustic degree of contrasts of R1/R2 of repair in relation to breaks of non-repaired prosodic units could shed light on the prosodic characteristics in relation to speakers' planning of the on-going speech.

2. SPEECH DATA AND ANNOTATIONS

2.1. Speech data

The source of data is a National Taiwan University Lecture Corpus (NDLC), which consists of totally 45 lectures as a semester course taught by a male professor. Examples of Mandarin repair were manually tagged over almost two-

hour of the NDLC lecture data, totaling around 41,000 syllables [9].

2.2. Preprocessing and annotations

The aforementioned 2-hour lecture data was first preprocessed by using the HTK Toolkit to force-align preliminary segmentations. The processed data then underwent through manual spot-checking by trained annotators. The next step was to manually tag the data with discourse units by perceived prosodic boundaries as well as instances of repair.

2.2.1. Annotations by discourse units

Prior to the annotations of repair instances, the speech data has been tagged according to a prosodic framework of Hierarchy of Phrase Group (HPG, [10]). The HPG framework specifies discourse-prosodic boundaries at 5 levels, from lower level of syllable (SYL), prosodic word (PW), prosodic phrase (PPh), breath group (BG, a physio-linguistic unit constrained by breathing during speech production), to upper level of multiple phrase speech paragraph (PG). Each level corresponds to the prosodic boundary from B1 through B5. Precisely, by default the correspondences between prosodic units and prosodic boundaries from syllables upwards can be stated as: SYL/B1<PW/B2<PPh/B3<BG/B4<PG/B5. By definition the annotation makes same-level between-break units as sister prosodic units while specifies higher-level prosodic branching and governing relationship at the same time; and allows phrase units to be examined in relation to paragraph assigned discourse identity instead of unrelated intonation units.

2.2.2 Annotations of repair instances

The same speech data was further annotated manually by one of the authors with an additional layer for repair, which is defined as situations in which the speaker stops in the middle of utterance and repeats part of it, with or without modifications; also when the speaker replaces part of the utterance immediately with another lexical or syntactic constituents. When identifying instances of repair, the annotator also relied on acoustic cues including sound stretch and/or cut-offs at the point of repair initiation (cf. right after R1) to locate them. Each instance was categorized, based on methods of repairing, into one of the following types:

Recycling repair ([REC]) which is defined as a brief repeat or re-saying of part of the utterance.

- (1) tʃyɣl t'amən suo-uei tʃ tʃyɣkə[R1] tʃyɣkə[R2] tʃyɣkə
internal kən external tʃ evidence
This they so-called DE **this this** internal and external
DE evidence

'This is **the the** so-called internal and external evidence by them.'

Partial Recycling repair ([PartR]) is when the speaker initiates the repair in the middle of a word or phrase at R1, and completes the intended constituent throughout R2.

- (2) Suo-i nɣ tʃyɣkə ɣl [tʃyɣŋ-[R1] tʃyɣuən[R2]] limien tʃ
xəntuo k'uənnan tʃ uənt'i
So this be **Ch- Chinese** in of many difficult of question
'So, this is many of the difficult questions in **Ch- Chinese**.'

Addition repair ([Add]) is when the speaker recycles to repair while at the same time adding other elements to complete the repair.

- (3) Na mei-i-tʃɣ in-uei [t'atə[R1] t'a tɕiaŋtə[R2]] tuŋɕi pu
i-iəŋ tɕiəu pa t'a tɕ'ietʃ'ulai
then each entry because **its it talked of** thing Neg
same then BA it cut out
'Then since for each entry **its- it talks about** different things, it was cut out.'

Replacement ([RePl]) refers to instances when (part of) R2 is produced to replace R1.

- (4) zʊkuo [tʃyɣkə[R1] tʃyɣtuei[R2]] tʃl liən-ɕy tʃ'ueien tʃ xua
if **this these** character consecutively appear if
'if **this these** characters come up consecutively,'

Other ([Othr]) are instances that involve complex process of repairing and cannot be categorized.

- (5) Suo-i [t'atə OOV ɣl t'ɣ[R1] tʃyɣuən tʃ OOV tʃ rate ɣl
t'ɣpiɕ [R2]] kau tʃ'
So **its OOV be esp- Chinese of OOV of rate be especially high**
'So **it's OOV is esp- the OOV rate in Chinese is especially high**.'

Each repair was marked from the onset to the offset of its R1 and R2, with or without pauses in between R1/R2. In total there were 290 instances of repair identified from the speech data. Figure 1 summarizes the distribution of five types of repair:

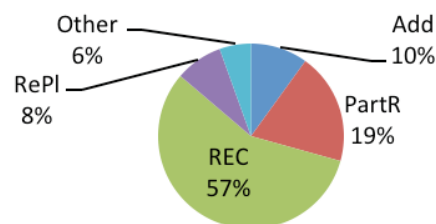


Figure 1: Distribution of types of repair.

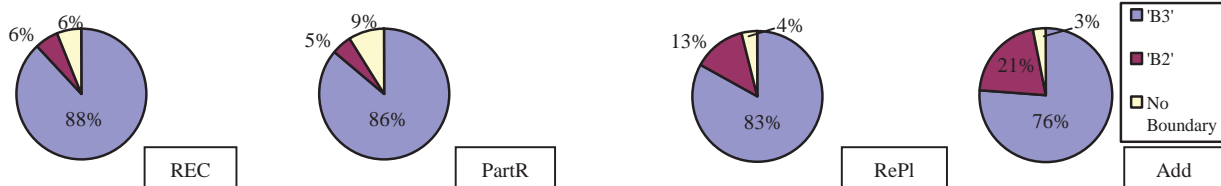


Figure 2: Repair at DU boundaries.

As can be seen, the type of *recycling* repair as repetitions of words or phrases is the most commonly used method of repair. In addition to the *other* type, the type of *replacement* is another less frequently used method to carry out repair. Since the type of *other* repair is the least frequently used, we will exclude instances of this type from the following analyses.

3. METHODOLOGY

Since the main focus of the study is to compare the relative prosodic realizations of repair in terms of discourse-prosodic units of spontaneous speech, the first step is to identify the discourse-prosodic unit (DU) from the HPG framework that serves as the carrying unit for Mandarin repair. This is done by simply counting the most frequently occurred boundary breaks immediately after the repair initiation point, namely after R1 of the repair.

For the prosodic profiles of R1/R2 of the repair, in the current study we started by focusing on features including: *F0*, *duration* and *intensity* of R1/R2. The measurements of *F0* and intensity were automatically extracted by averaging the *F0* and intensity (dB) values from the onset to the offset of R1 and R2. As for duration, we calculated the relative tempo/speech rate by extracting duration values of R1/R2 that were subject to Z-score normalization. Then the duration value was further refined to remove the effect from inherent segmental duration as well as in-repair pause (in between R1/R2). Eventually the duration factor was presented as averaged duration of syllables in the unit of *prosodic word* (PW) from HPG; i.e. syllables of the PW were averaged to derive the mean duration, defined as:

$$(6) PW_Dur = \sum_i^M syl_dur_i$$

$$PW_AveDur = PW_dur / M$$

While *M* donates the number of syllable in PW, *i* stands for the index of the syllable.

To facilitate relative comparison of measurement results, we adopt the concept of *contrast degree* (CD) from [7] and [8]. CD is calculated by simply subtracting the value of one of the acoustic features in R1 from the corresponding R2 of the same repair. Currently CD is calculated for *F0*, duration, intensity (cf. Sec 4.2), then paired with pause

duration when applicable (cf. Sec. 4.3). Thus in case of negative values derived, it implies R1 is relatively higher in *F0*/longer in duration/ louder in intensity. The results are presented as follows.

4. ANALYSES OF CONTRAST DEGREE AND DISCUSSIONS

4.1. Repair at discourse boundary

We first identify the underlying DU incorporated while repairing, and the result is presented in Figure 2 above. The result in the figure has been arranged according to types of repair. As can be seen, the most frequently occurred discourse-prosodic boundary in between R1/R2 is the prosodic phrase (PPh/B3). Since the result is quite consistent across the four major repair types, it implies that the default DU carrying repair should be PPh.

4.2. Analysis 1: contrast degree of prosodic features in repair

Since we have identified B3 as the most frequently occurred boundary break in between R1/R2, the B3 unit of PPh, a sub-unit to the HPG specified multi-phrase speech paragraphs, is defined as the bearing prosodic unit of repair. Our first analysis focuses on the *contrast degree* (CD) of prosodic features including **F0**, **duration**, and **intensity** of R1/R2 of the repair. At the same time, we calculate the CD value of the same prosodic features at the PPh boundary that contains no repair, also by the unit of *prosodic word* (PW). The CD values from repaired PPh and non-repaired PPh boundaries would undergo further T-test to determine if the CD value reflected from repair can be distinguished from that at ordinary PPh boundaries. Note here in order to avoid the influence from boundary effects of larger discourse units, we calculate the CD from boundaries of only the medial PPhs that carry no repair instance.

4.2.1. Contrast degree in F0

In the following Figure 3 (left panel), it summarizes the results of CD in *F0* from R1/R2 by types of repair, as well as from the PW at non-repaired PPh boundaries. The results in numbers are otherwise presented in Table 1.

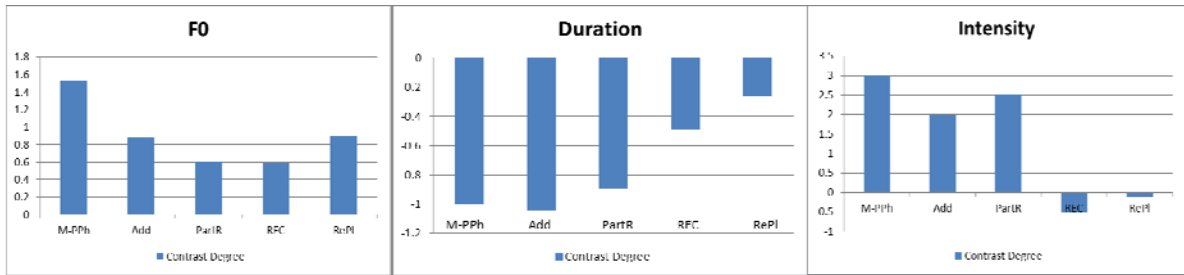


Figure 3: Contrast degree in F0 (1st panel), duration (2nd panel) and intensity (3rd panel)

Pitch	Contrast Degree	ttest: h	ttest: p
M-PPh	1.5242	-	-
Add	0.8832	0	0.3741
PartR	0.6071	0	0.1126
REC	0.5808	1	0.0047
RePI	0.9041	0	0.4594

Table 1: Results of contrast degree in F0.

From Figure 3 (1st panel) and Table 1, we can see that the medial-PPh without repair shows the largest CD value in F0 at boundary breaks. Although the 4 major repair types demonstrate a consistent pattern of positive CD value, which implies R2 is realized in relatively higher pitch, the CD values between R1/R2 are always smaller. Most of all, the T-test results suggest that the CD in F0 between R1/R2 of only the *recycling* type is significantly different from the CD at the non-repaired PPh boundaries.

4.2.2. Contrast degree in duration

The results of CD value in duration are presented in Figure 3 (2nd panel) as well as in the following Table 2, with numbers in detail.

Tempo	Contrast Degree	ttest: h	ttest: p
M-PPh	-0.9995	-	-
Add	-1.0436	0	0.9047
PartR	-0.9009	0	0.7153
REC	-0.4951	1	0.0017
RePI	-0.2641	0	0.0692

Table 2: Results of contrast degree in duration.

As shown above, the result of CD in duration of R1/R2 is also consistent across the 4 repair types in that R2 is in faster speech rate. When comparing to non-repaired medial-PPh, on the other hand, only the *recycling* type displays a significant difference. In other words, the R1/R2 of the *recycling repair* demonstrates a relatively smaller CD difference in tempo when comparing to ordinary medial-PPh.

4.2.3. Contrast degree in intensity

Intensity	Contrast Degree	ttest: h	ttest: p
M-PPh	3.0127	-	-
Add	1.9764	0	0.2544
PartR	2.514	0	0.4499
REC	-0.4951	1	0.0017
RePI	-0.1144	1	0.0018

Table 3: Results of contrast degree in intensity.

Based on results from Figure 3 (3rd panel) and Table 3, we find the CD value in intensity realized in R1/R2 is not as consistent across the 4 repair types. But in general the CD value of intensity in repair is still smaller than that from the non-repaired medial PPh boundaries. Finally, the *recycling* and *replacement* types reflect rather small contrast degree in intensity that results in statistical distinctions from non-repaired medial-PPh.

4.2.4. Discussion

For analysis 1, we examined the acoustic degree of contrast from R1/R2 of Mandarin repair, in comparison with the CD values from non-repaired PPh boundary breaks. Based on the results, it is suggested that of the 4 types of repair, the *recycling* demonstrates obviously smaller CD value acoustically comparing to the non-repaired PPh. In other words, over half of the repair instances from our lecture data exhibit relatively smaller CD in F0, duration and intensity that can be significantly distinguished from the CD values at boundaries of PPh without repair.

Still, from the above results, there are about 27% of the repair instances exhibiting CD differences indistinguishable from boundary breaks at non-repaired PPh. One place with regard to the acoustic aspects of the repair that we haven't addressed is the *pause duration* in between R1 and R2. Thus in a follow-up analysis, we take the factor of **pause duration** into consideration by a modal of analyzing prosodic features in speech data proposed in [7, 8].

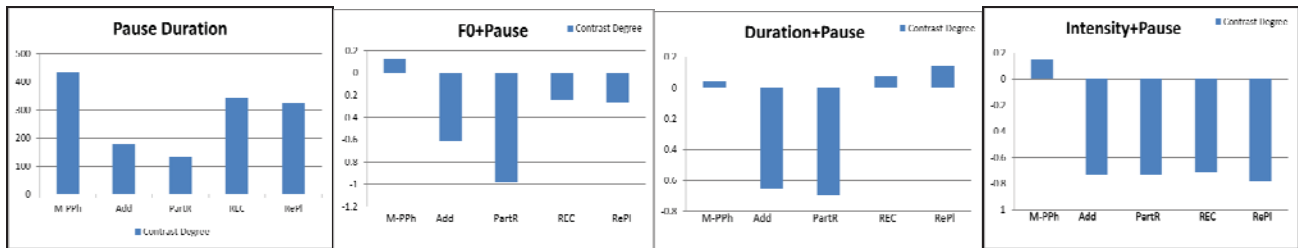


Figure 4: Results of normalized pause duration (1st panel), and the results of normalized contrast degree in F0, duration and intensity plus the normalized pause duration (2nd, 3rd, 4th panel respectively).

4.3. Analysis 2: contrast degree of prosodic features in repair by taking pause duration into consideration

Since the HPG-specified PPh boundary properties are positively correlated with between-PPh contrast degrees [7] and contrast degrees have shown to be positively enhanced by pairing CD with pause duration [8], we hypothesize that the same pairing of factors from *contrast degree* with **pause duration** (PD) would result in better discrimination of repair from non-repaired PPh. In the following analysis, *contrast degree* and *pause duration* are used to generate paired-combinations as variables. The 3 CD values (in F0, duration, and intensity) and values of PD are subjected to Z-score normalization respectively. Normalized CD and PD are then added up to obtain results that would undergo further T-test in order to substantiate the proposed hypothesis.

4.3.1. Normalized pause duration

First of all, results of normalized pause duration have been presented in the first panel from Figure 4, and the detailed results in numbers are presented in the following Table 4.

Pause	Normalized Duration	ttest: h	ttest: p
M-PPh	431.36	-	-
Add	177.4277284	1	0.000248
PartR	134.6608261	1	3.09E-09
REC	341.3064394	1	0.0041
RePl	324.3656796	0	0.1592

Table 4: Results of normalized pause duration.

From the results, we can see that the pause duration in between boundary breaks of medial PPh without repair has the largest value. As for the PD in between R1/R2 of repair, they are significantly smaller except for the type of *replacement* repair.

4.3.2. CD in F0 plus PD

The 2nd panel of Fig. 4 and the following Table 5 summarize results from normalized CD in F0 plus the result of normalized PD.

F0 +Pause	CD + Pause	ttest: h	ttest: p
M-PPh	0.1216	-	-
Add	-0.6191	1	0.05
PartR	-0.9806	1	0.000339
REC	-0.2417	1	0.05
RePl	-0.2679	0	0.3826

Table 5: Results of normalized CD in F0 plus PD

As have shown, by adding up the normalized CD of F0 and PD, the majority instances of repair could be distinguished from non-repaired PPh boundary breaks. The only exception is the *replacement* type which presents indistinguishable result from the non-repaired medial PPh.

4.3.3. CD in duration plus PD

Turning to CD in duration plus pause duration, the results in numbers are presented in the following Table 6:

Tempo+Pause	CD + Pause	ttest: h	ttest: p
M-PPh	0.043	-	-
Add	-0.6567	1	0.0072
PartR	-0.6922	1	0.000115
REC	0.0704	0	0.8115
RePl	0.1393	0	0.7362

Table 6: Results of normalized CD in tempo plus PD

Here as well as from the 3rd panel of Fig. 4, it is demonstrated that when taking the combined factors of *duration* and *pause* values into consideration, the prosodic realization in *addition* and *partial* repair can be distinguished from non-repaired medial PPh boundaries. One place to note is that, for the *recycling* type, although the T-test result was not significant, it could be due to the cancelation from adding up the positive and negative

duration values. After taking the absolute duration value to re-calculate, we obtained a significant result for the *recycling* repair (0.5093, P=0.0156). Thus the only exception remains is the *replacement* repair which presents no significant difference.

4.3.4. CD in intensity plus PD

Intensity+ Pause	CD + Pause	ttest: h	ttest: p
M-PPh	0.1526	-	-
Add	-0.7362	1	0.0037
PartR	-0.7366	1	6.01E-05
REC	-0.7169	1	1.09E-10
RePI	-0.7863	1	0.0052

Table 7: Results of normalized CD in intensity plus PD.

Finally, the result from the 4th panel in Figure 4 and in Table 7 above explicate that, when taking into consideration the CD in intensity plus PD values, there is a clear distinction between R1/R2 of the repair and boundary breaks at non-repaired medial PPhs. The T-test results are significant across all four types.

4.3.5. Discussion

Based on the results presented, we confirmed our hypothesis that the pairing of CD values with the value of PD provides better discriminations between R1/R2 of Mandarin repair from the boundary breaks of medial PPh that contains no repair. In other words, our findings suggest that while repairing during speech production, various prosodic features including pause duration have been combined to create the contrast degree that should be differentiated from boundary breaks of ordinary prosodic phrases. Obviously, the speaker in our data doesn't create as dramatic acoustic contrasts while repairing. This in turn implies that, at the higher discourse level, the speaker could be aiming at rather different planning distance for the information following the repair in order to maintain the on-going speech production.

5. GENERAL DISCUSSION AND FUTURE RESEARCH

In the present study we started out by identifying the prosodic phrase (PPh) as the bearing unit from HPG framework for our Mandarin repair instances. We carried out analyses to further clarify if the relative acoustic degree of contrasts between R1/R2 of the repair could be distinguished from breaks at non-repaired medial PPh boundaries. The result from our analysis 2 demonstrated that, by pairing pause duration with results of CD from analysis 1 it yields the best distinctions between repair and non-repaired PPh.

The current results thus reflect that, when carrying out repair, the speaker actually tones down the acoustic differences between R1/R2. Since the speaker doesn't create as much prosodic highlight during repair as other ordinary phrasal breaks, this implies that repair itself possibly carries lighter information load. A crucial step to follow-up with the current analyses, then, is to explore further how speakers adjust and deploy the acoustic contrast degree between repaired constituents and the following content in relation to the allocation of information highlight in spontaneous speech.

Since the current data of spontaneous lecture speech is produced by one speaker in the format of monologue, to test if the findings are genre-specific and generalizable across speakers, future works in the same vein will extend to at least spontaneous dialogues, as well as more speakers. Finally, but not the least, as speakers' planning of higher-level discourse structures should share similar features cross-linguistically, additional future works would also include testing the current findings in speech data other than Mandarin Chinese.

6. REFERENCES

- [1] A. Stolcke & E. Shriberg, "Statistical language modeling for speech disfluencies," In: *Proc. ICASSP-96*, Atlanta, Georgia, pp. 405-08, 1996.
- [2] J. Duchateau et al., "Handling disfluencies in spontaneous language models," *Language and Computers*, pp. 39-50, 2003.
- [3] J. Yeh et al., "Disfluency correction of spontaneous speech using conditional random fields with variable length features," In: *Proc. ROCLING*, Hsinju, Taiwan, 2006.
- [4] E. Shriberg, *Preliminaries to a theory of speech disfluencies*, PhD thesis, Stanford University, 1994
- [5] S.-C. Tseng, "Repairs in Mandarin conversation," *Journal of Chinese Linguistics*, pp. 80-120, 2006.
- [6] M. Swerts, "Filled pauses as markers of discourse structure," *Journal of Pragmatics*, pp. 485-96, 1998.
- [7] C.-Y. Tseng, & C.-H. Chang, "Pause or no pause? – Prosodic phrase boundaries revisited," *Tsinghua Science and Technology*, pp. 500-09, 2008.
- [8] C.-Y. Tseng, & Z.-Y. Su, "Boundary and lengthening—On relative phonetic information." *The 8th Phonetics Conference of China and the International Symposium on Phonetic Frontiers*. Beijing, China. 2008.
- [9] C.-Y. Tseng, & Z.-Y. Su, "Spontaneous Mandarin speech prosody—the NTU DSP lecture corpus," In: *Proc. Oriental COCOSA 2008*, pp. 171-74, Kyoto, Japan, 2008.
- [10] C.-Y. Tseng, *Prosody analysis. Advances in Chinese Spoken Language Processing*, pp. 57-76, World Scientific Publishing, Singapore, 2006.