PARENTHETICAL – A SPECIAL TYPE OF PROSODIC REDUCTION IN CONTINUOUS SPEECH

Chiu-yu Tseng, Helen Kai-yun Chen, Yen-Hsing Chen

Phonetics Lab, Institute of Linguistics, Academia Sinica, Taiwan
cytling@sinica.edu.tw

ABSTRACT

The current study investigates **parenthetical**, a type of prosodic reduction in multi-phrase speech paragraphs. Structurally a modifier of its antecedent to provide supplementary information, such reduction creates a lower level in the prosodic hierarchy nested within a discourse-prosodic unit. Perceptual annotation of parenthetical turned out to be consistent across listeners; their acoustic profiles distinctive. Further calculation of information density in relation to allocation of perceived emphasis also demonstrates that parenthetical triggered prosodic reductions are patterned and accountable. Therefore, in spite of low information standing, their existence in the prosodic hierarchy helps facilitate more precise information expression. In sum, current evidence illustrates how information planning is manifested via both emphases and reductions in global context prosody, why parenthetical caused reductions should be understood from a hierarchical perspective within speech context, and how prosodic reduction also plays a crucial role in contributing to comprehensive understanding toward context prosody.

Index Terms—discourse hierarchy, prosodic reduction, parenthetical, nesting, continuous speech and expressiveness, global context prosody

1. INTRODUCTION

In the explorations of speech expressiveness, it is often associated with the paralinguistic aspects of speech production and has been defined as 'added by speakers deliberately onto the linguistic information' [1]. One major facet of the paralinguistic features in contributing to speech expressiveness lies in prosody, especially how speakers deploy prosodic highlight, cues related to focus and emphasis, for listeners to orient to. While prosodic highlight functions to enhance part of the speech signals in order to foreground or to project the most salient part of the speech, what has been less addressed to in the relevant literature is how *prosodic reduction* may also contribute to speech expressiveness. Most of all, it is held that perceivable saliency is the result of not only the prosodic highlight (perceived prominences) but also co-constructed prosodic reduction to reach the most effective expressiveness in speech context. The objective of the present study thus addresses *prosodic reduction* that contributes to expressiveness in continuous speech. Specifically the current exploration concentrates on **parenthetical** as a particular type of prosodic reduction in speech context with illustration of its function in relation to information arrangement via global context prosody.

In relevant literature, parenthetical construction (also parenthesis) has been discussed from multiple approaches, including theoretical syntax, morpho-syntax, prosodic realization, meanings and interactional functions, as well as perspectives from information planning. Traditionally parenthetical has been treated as a construction that is 'linearly integrated in another linguistic structure' [2] because of its weak connection to the speech context [3:179]. Due to the absence of a single morpho-syntactic class in corresponding to the construction [4], no consensus on a clear definition for parenthetical has been yielded from previous discussions. Theoretically, parenthetical has been regarded as a sequence inserted to its host and is structurally independent and autonomous (e.g. [5]). Following the autonomous viewpoint, other studies focusing on the acoustic features of parentheticals have reported cues including lower F0 and/or compressed F0 range, weaker intensity, also faster speaking rate [2], [6], [7]. As for its meaning and function, it has been suggested that parentheses function to provide supplementary information or information of meta-communication. Recently, [8] examines parenthetical construction in continuous speech by taking it as a case of prosodic reduction. Neither insertion nor simply a linear integration, as demonstrated, parenthetical should be viewed as an integrated part of information planning from the whole speech context; eventually parenthesis as prosodic reduction interweaves with prosodic highlight to strengthen contrast degree and contribute to expressiveness in global context prosody [8].

The current analyses on parenthetical as a type of prosodic reduction aim at staging the construction in relation to the larger comprehensive context from continuous speech in Mandarin, with the goal to advance the claims made previously by [8]. In particular, we examine parenthetical with the assumption that the relationship between parentheticals and the speech context

can be accounted for as a nested status: in terms of F0 realization, for instance, it can be translated into the illustration in Fig. 1.

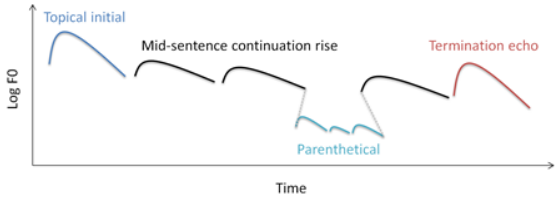

Figure 1: *Schematic diagram of speech production.*

It is argued that, although perceptually distinguishable from its context in creating a discontinuous 'down-stepping' gap corresponding to lower information loading, parenthetical nested within discourse-level prosodic units nevertheless plays a unique role similar to post focus compression in bringing out information arrangements more precisely when co-constructing with prosodic highlight. Consequently greater degree of contrast due to more prosodic ups and downs in prosodic output turns out to be another source of expressiveness. Yet it is held these variations are not just randomized allocations and alternations between prominent prosodic highlight and reduction in the speech flow. Crucially, we will demonstrate that even as a specific type of reduction, parenthetical could be patterned prosodically in such a way that it can be realized as a *replicate* of the overall global context prosody from higher-level discourse units. Through calculation of emphasis tokens distribution and their pattern allocation, as will be shown, our current analyses offer substantiation to the nested status from parentheticals while showcasing that their relationship to the entire speech context would be yielded only when taking into account the overall features and constrains from global prosody. Neither constructed autonomously nor incorporated independently, parenthetical as a type of prosodic reduction is in fact an integral part of overall speech planning with its unique status in contributing to the construction of and comprehensive understanding toward global context prosody.

2. SPEECH DATA AND ANNOTATION

2.1. Speech data and preprocessing

The present data consist of a selection from university classroom lecture (SpnL) [9], delivered by a male professor. The content of the lecture was delivered in Taiwanese Mandarin and the total time of the selected data is up to 2.5 hours, equaling to approximately 33980 syllables. During the preprocessing stage, the data was first force-aligned into preliminary segmentations by using HTK Toolkit and the output was then manually checked by trained transcribers. Afterwards the data have undergone labor-intensive annotations in separate layers for prosody-related and perception-based information. Three layers have been labelled independently including: discourse-prosodic units/boundaries, perceived prosodic highlight and instances of parentheticals, which are described as follows.

2.2. Data annotations

2.2.1. Annotations for discourse-prosodic unit (DPU)

Discourse-prosodic units (DPU) in five levels were annotated for the current data according to the hierarchical prosodic phrase grouping (HPG) framework (i.e. [10], [11] and [12]). The boundaries of five levels are marked from B1 to B5, corresponding respectively to syllable (SYL), prosodic word (PW), prosodic phrase (PPh), breath group (BG, a physio-linguistic unit constrained by change of breath while speaking continuously) and multiple phrase speech paragraph (PG). By definition, the boundary breaks, prosodic units and their relationship within the HPG framework could be accounted for by:

SYL(B1)< PW(B2)< PPh(B3)< BG(B4)< PG(B5) [12].

2.2.2. Annotations for perceived prosodic highlight

The same speech data have undergone the labelling process by trained annotators into a string of perceived emphasis/non-emphasis tokens (ETs). The annotation of ETs has been carried out in a separated and independent layer and the decision of breaking up the speech stream into ETs/non-ETs was not constrained by any syntactic-based unit nor pre-defined constituents but based on four relative degrees of prominence:

- E0 – reduced pitch, lowered volume, and/or contracted segments
- E1 – normal pitch, normal volume and clearly produced segments
- E2 – raised pitch, louder volume and irrespective of the speaker's tone of voice
- E3 – higher raised pitch, louder volume and with the speaker's change of tone of voice

With this annotation scheme, it is noted in particular that degrees of prominences can be consistently perceived only by limited levels of contrastiveness.

2.2.3. Identification of parenthetical

Parenthetical construction (**PAR**) was annotated in an additional layer. The identification for **PAR** is *perception*-based, defined as a construction that is disruptive to the current speech production and is perceived distinctively by discernible acoustic features from the context. While syntactic discontinuity may occur prior to its pre-boundary, the content of parenthetical is still related to its immediate speech context. Most of all, the projected content from parenthetical as a complete construction functions to

facilitate the understanding towards the on-going speech in planning. For the current data, 81 tokens of **PAR** have been identified. The following are two selected examples of **PAR** (marked by the square bracket) within larger speech context:

- 所以語音辨識有一堆 error [deletion 這個 substitution, insertion 的各種 error] 然後有 OOV 嘛

  'So there are lots of errors in speech recognition, [types of errors such as deletion, substitution, and insertion], and then there is OOV'

- 那中文的問題是說你[它是一串字]，你不曉得詞在哪裡

  'Then the problem with Mandarin Chinese is that you, [it's a string of words] you don't know where the boundaries of lexical items locate.'

Note that there's no restriction in any specific morpho-syntactic category and the sizes of **PAR** can range differently from a prosodic phrase corresponding to a clause, to several prosodic phrases (that are consisting of more than one NP).

3. METHODOLOGY

To validate the nesting relationship between parentheticals and their context, we carry out the following calculations: overall acoustic realization, information density, and also emphasis distribution/allocation.

3.1. Acoustic realizations

In order to identify the acoustic cues discriminating **PAR** construction from its context perceptually, mean values of major acoustic features were computed, including: F0, F0 range, duration and intensity. We incorporated the DPU *prosodic phrase* (PPh) as the unit of calculation.

First of all, we used SAP toolbox [13] to extract raw F0 value from each PPh within **PAR** tokens. In addition to the mean value calculation, raw F0 was used to obtain F0 range by simply subtracting minimum from the maximum. As for duration, we normalized the length of every phoneme to remove the intrinsic duration differences, and then averaged the phoneme duration to obtain speaking rate. Finally intensity (dB) was extracted also by using SAP toolbox and by PPh unit.

3.2. Information density

Information density is an ad-hoc estimation based on levels of prosodic highlight perceived and annotated for the current data. It is assumed that such estimation directly reflected the loading of information content throughout speech production, especially the distribution of information by the concept of density. Following similar rational from [14], we arbitrarily assigned weighting scores to the emphasis degree tags: [-1 0 1 2] for [E0 E1 E2 E3] respectively. To demonstrate the comprehensive differences in information density distribution, we average scores from each PPh constituting the parenthetical and also PPhs in the speech context.

3.3. Emphasis distribution/allocation

Following the method in [15], emphasis distribution is examined by simply counting the percentage of four emphasis/non-emphasis tokens based on the PPh unit. Through the distribution it provides further evidences regarding whether the perceived emphasis allocated in parenthetical pattern differently from whole speech content.

On the other hand, patterns of emphasis token allocation can be derived by each PPh. Then we further merged patterns of the same ETs in sequence into a unique type and the respective frequency by each type was computed. Cumulative frequency distribution (CDF) is adopted and defined below.

$$\mathrm{F}a(X) = P(a \le X) \tag{1}$$

where the right side of the equation denotes the probability that the pattern a takes on a value less than or equal to X.

4. RESULTS AND DISCUSSIONS

4.1. Acoustic features

To identify the perception cues discriminating **PAR** construction from its *context* (defined as part of speech signal that does not contain **PAR**), we first calculate the mean values of major acoustic cues. As explicated in 3.1 here we incorporate prosodic phrases (PPhs) as the unit of analyses. The results of mean values from each acoustic cue with further normalization are summarized in Fig 2 and Table 1.

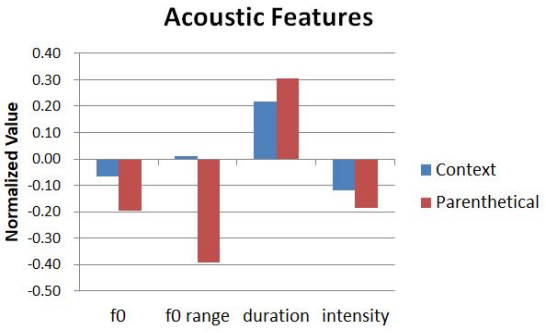

Figure 2: *Normalized value of acoustic features from parenthetical vs. context* (a negative value after normalization reflects lower F0/ narrower F0 range/ shorter duration/ weaker intensity).

Table 1: *Values of acoustic features in numbers.*

	Context	PAR	h	p
f0	-0.07	-0.20	0	0.083
f0 range	**0.01**	**-0.39**	**1**	**<0.0005**
duration	0.22	0.31	0	0.240
intensity	-0.12	-0.19	0	0.255

4.1.1. Discussions

Fig. 2 features parenthetical in terms of acoustic realizations including: *lower* f0, *narrower* f0 range, *longer* duration and *weaker* intensity, comparing to those PPhs in the speech *context* but without **PAR**s. In other words, the reduced nature of parenthetical construction is more prominently realized in cues of F0 and intensity. Such results are in general consistent with what's been reported previously ([2], [6], [7]). Here further substantiation of the results from acoustic measurements is provided by statistical test and Table 1 shows that the most and the only significant feature is f0 range compression. This further implies that to distinguish the **PAR** construction from its context, reduced F0 range would be the most pronounced cue for listener to follow from within the speech context.

4.2. Information density

The next analysis involves explorations on the differences of information loading between **PAR** and its *context*. Here the corresponding information density is computed based on levels of prosodic highlight annotations. Following the methodology in 3.2, we average the weighting scores derived from each PPh within parentheticals and also from its *context*. The results are summarized in Table 2.

Table 2: *Average weighting score of parenthetical vs. context (E0=-1; E1=0; E2=1; E3=2).*

Context	PAR
0.12(0.33)*	0.07 (0.30)*

*Std in parenthesis

4.2.1. Discussions

From Table 2, it is clearly that the average weighting score is lower in **PAR** than in its *context*. This is within our expectation as parenthetical functions to offer supplementary information and/or information for meta-communication, they may contain less amount of new information. In other words, the major distinction between parenthetical and the speech *context* is also reflected in how the speaker manipulates the deployment of prominence levels for information planning and distributed information loading for the purpose of communication.

4.3 Emphasis distribution/allocation

After identifying the acoustic features distinguishing **PAR** and calculating information density scores, in the following we turn to the distribution of emphasis tokens and their allocations within **PAR**. To elaborate further the findings from [8] that **PAR** is neither an insertion nor linear integration, the current analyses demonstrate that parentheticals, as a part of continuous information planning, is related to its *context* by the same organization in terms of emphasis token (ET) distribution and allocation, thus substantiating its nesting relation within higher level context prosody.

Following the methodology from 3.3, the results of ET distribution and pattern allocations are summarized in Fig 3 through Fig 5.

4.3.1. Discussions

Fig. 3 presents results of emphasis level distribution by both **PAR** and its *context*. Surprisingly, there is not much difference found in the distribution of ET tokens with actual emphases e.g. E2/E3 (24.1% in PAR and 25.9% in its context). Assuming emphasis levels E0 and E1 both carry no perceivable emphasis and can be further merged, then we would arrive at no difference from the distribution of non-emphasized tokens between **PAR** and its *context* either (75.9% vs. 74.1%). Instead the proportion of E0 and E1 tokens are only slightly different in that there's about 4% more of E0 tokens in **PAR**, which otherwise reinforces the reduced nature of parenthetical construction. In other words, although the information density is lower in **PAR** (as discussed in section 4.2.1) and the construction is perceptually more reduced, the emphasis (information) distribution is highly similar, especially in terms of the ratio between no-emphasis/emphasis distinctions.

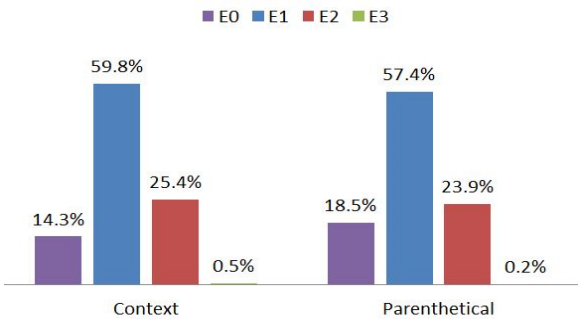

Figure 3: *Emphasis token distribution in parenthetical vs. context.*

Considering the relative nature among emphasis levels, we wonder if the same contrastiveness of acoustic realization would still be held among each emphasis level in both parenthetical and its *context*. So we further calculate the mean value of major features with additional normalization. As shown in Fig. 4, the contrast degrees between E0/E1/E2&E3 in terms of acoustic realization

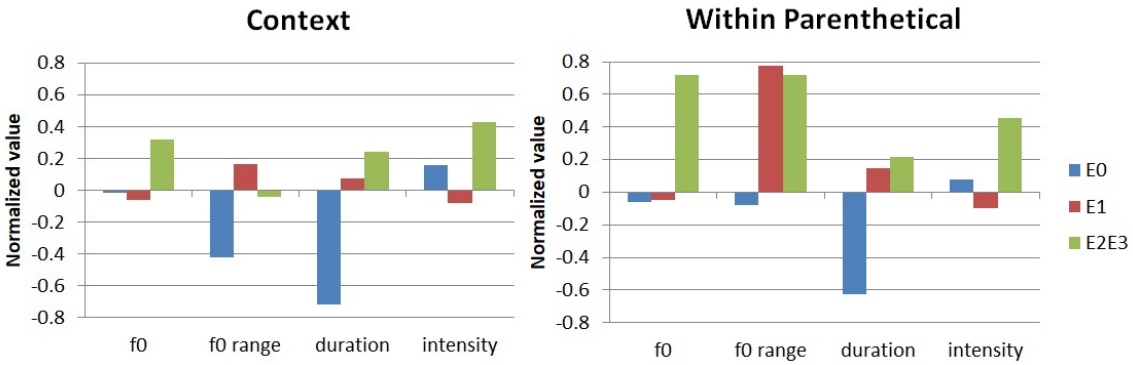

Figure 4: *Contrast degree of perceived emphasis levels in parenthetical vs. context.*

derived from **PAR** and its *context* form rather similar pattern, except for F0 range. With respect to the whole *context*, the contrast between E0/E1 is most distinct in F0 range and duration; whereas E2&E3 are realized in higher F0, longer duration and louder intensity. Similar patterns of acoustic realization can be identified repeatedly within parenthetical as well. It is thus suggested that although acoustic signals may be reduced in **PAR** (as shown in 4.1), the recipient still incorporate and maintain similar degree of contrastiveness as result of no-emphasis/emphasis token placements.

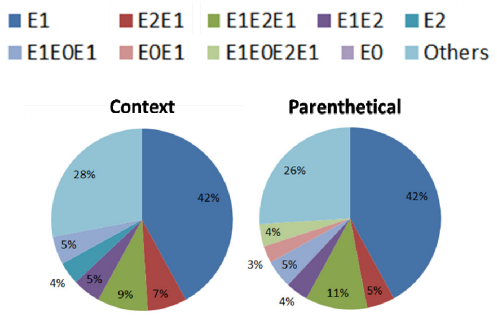

Figure 5: *Emphasis allocation in parentheticals vs. context.*

Finally turning to the sequential pattern of emphasis token allocation, we present in Fig. 5 further evidence to the claim regarding nested status by **PAR**. Most of all, the present results shows that excluding the 'others' category, some of the major prosodic highlight patterns such as 'E1,' 'E1E2E1,' 'E2E1' are shared between parenthetical and its context. This further implies that speakers may incorporate similar emphasis-correlated information planning mechanism during the on-line speech production in global context, as well as throughout the planning of parenthetical. The planning of information allocations within parenthetical, therefore, turns out to be a *replication* of how speakers usually plan for information allocations from the speech context. This is one of the evidences supporting the claim that parenthetical can be held as in a nested relationship within the whole *context*. Last but not the least, it should be mentioned that we still observe slightly different pattern distribution from PAR in that there are more patterns consisting of reduction E0 (e.g. 12% in PAR while only 5% in the context). This again foregrounds the reduced natural of **PAR** in terms of prominence annotations.

5. GENERAL DISCUSSIONS AND FUTURE WORKS

The current study focuses on the role of prosodic reduction in contributing to expressiveness in continuous Mandarin speech, with particular concentration on parenthetical construction as a type of prosodic reduction. It is assumed that the relationship between parenthetical and the context they occurred in can be accounted for as a nested status. Although prosodically distinctive from its bearing speech context and perceptually compressed in reflecting lower information loading, parentheticals are not to be taken as random insertion nor are they autonomous as defined traditionally. By staging the construction in higher-level discourse prosodic context, parenthetical as a type of reduction adds to more precise information expression in speech via co-constructing with prosodic highlight perceived prominently in strengthening gaps among prosodic contrasts. While this serves as a source of expressiveness and results in surface variations from prosody, we otherwise offer solid evidences to argue for how the strengthened contrastiveness forms patterns that are regular and accountable. Through similar emphasis pattern distribution and maintaining same contrast degree between emphasis/non-emphasis tokens within parenthetical, the current analyses showcase the significance of the nested status for the construction: their compressed nature does not allude to an additional level beneath the current discourse-prosodic units; instead parenthetical is integrated in and hence constrained by higher-level discourse unit in the same prosodic hierarchy. Had we lifted the construction from its context or examined it by isolation, the identifiable prosodic patterns would never merge. In the end, through further exploration of this nested status from parenthetical it contributes to a more comprehensive justification towards

incorporation of prosodic reduction in the global context prosody.

One final note to add is that, as introduced, initially we made a comparison between parenthetical construction in continuous speech and the phenomenon of post focus compression (PFC) in the prosodic realization of speech. The reduced realization by both PFC and parenthetical in speech prosody shares the similarity in that both may co-construct with emphasis perceived more prominently to strengthen the contrastive degree between prosodic highlight and reduction, which eventually results in further expressiveness from speech production. Now with detailed analyses of parenthetical as a type of prosodic reduction, we have also arrived at a better understanding towards the mechanism and status behind the incorporation of parenthetical in continuous speech. Most of all, unlike PFC as an acoustic feature concomitant with focus or prosodic highlight in continuous speech, parenthetical is a construction belonging to the continuous information planning and projection by higher-level discourse units. Eventually speakers and hearers may be able to deduce out of alternations between prosodic highlight and reduction those regular underlying patterns to help facilitate and achieve communicative goals in the process of speech delivery and exchanges.

For future studies, we plan to further investigate prosodic reduction by, first of all, clarifying other types and instances of reduction that can be consistently perceived and marked similarly across speech signal. We are interested in if the same contrast degree maintained in between prosodic highlight and reduction from current speech data is applicable across languages and contributes similarly and/or differently to speech expressions. In addition, we also plan to explore reduction in continuous speeches of different genres. As for parenthetical as a type of prosodic reduction, further analyses can address in particular its interaction with perceived prosodic highlight and boundary effects. We believe that these analyses are the essential additions to further our understanding toward and deconstruct global context prosody in continuous speech.

6. REFERENCES

[1] H. Fujisaki, ‘‘Prosody, information, and modeling—With emphasis on tonal features of speech,’’ in Speech Prosody 2004 Proceedings, Mar 23-26, Nara, Japan, pp. 1-10, 2004.

[2] N. Dehé. *Parentheticals in spoken English: The syntax-prosody relation*. Cambridge University Press: London, 2014.

[3] N. Burton-Roberts. "Parentheticals." *Encyclopedia of Language and Linguistics*, E. Brown Ed. Elsevier Science, pp.179-182. 2006.

[4] L. Grenoble. "Parentheticals in Russian," *Journal of Pragmatics*, vol. 36, no. 11, pp. 1953-1974. 2004.

[5] H. Bussmann, *Routledge Dictionary of Language and Linguistics*, Routledge: London. 1996.

[6] M. Payá, “Prosody and pragmatics in parenthetical insertion in Catalan” *Catalan Journal of Linguistics*, vol. 2, pp. 207-227. 2003.

[7] H. Mazeland. "Parenthetical sequences," *Journal of Pragmatics*, vol. 39, no. 10, pp. 1816-1869. 2007.

[8] H. Chen, Y, Chen and C. Tseng, "All for a reason—prosodic reduction in continuous speech." Oriental COCOSDA 2017 Proceedings, Nov 1-3, Seoul, Korea, pp.256-261, 2017.

[9] C. Tseng and Z. Su, "Spontaneous Mandarin Speech Prosody—the NTU DSP Lecture Corpus." Oriental COCOSDA 2008 Proceedings, Kyoto, Japan, pp.171-174, 2008.

[10] C. Tseng, S. Pin, Y. Lee, H. Wang, and Y. Chen, "Fluent speech prosody: Framework and modeling," *Speech Communication*, Vol. 46, no. 3-4, pp. 284-309, 2008.

[11] C. Tseng and C. Su, "Discourse prosody and context – Global F0 and tempo modulations," in INTERSPEECH 2008 – 9th Annual Conference of the International Speech Communication Association Proceedings, Brisbane, Australia, , 2008, pp.1200-1203.

[12] C. Tseng. "An F0 analysis of discourse construction and global information in realized narrative prosody," *Language and Linguistics*, vol. 11, no. 2, pp. 183-218, 2010.

[13] J.-S. Jang, "Speech and Audio Processing (SAP) Toolbox", retrieved from http://mirlab.org/jang/matlab/toolbox/sap.

[14] H. Chen, W, Fang and C. Tseng, "Information Content, Weighting and Distribution in Continuous Speech Prosody – A Cross-Genre Comparison," Oriental COCOSDA 2015 Proceedings, Shanghai, China, pp.75-80, 2015.

[15] C. Tseng and C. Su, "Information Allocation and Prosodic Expressiveness in Continuous Speech: A Mandarin Cross-genre Analysis," The 8th International Symposium on Chinese Spoken Language Processing, ISCSLP Proceedings, Hong Kong, pp. 243-246, 2012.