Annotating a corpus of human interaction with prosodic profiles – focusing on Mandarin repair/disfluency

Helen Kai-yun Chen

The Hong Kong Polytechnic University
Laboratoire Parle et Langage, Aix-Marseille Université
E-mail: helenkychen@gmail.com

Abstract

This study describes the construction of a manually annotated speech corpus that focuses on the sound profiles of repair/disfluency in Mandarin conversational interaction. Specifically, the paper focuses on how the tag set of prosodic profiles of the *recycling repair* culled from both audio-tapped and video-tapped, face-to-face Mandarin interaction are decided. By the methodology of both acoustic records and impressionistic judgements, 260 instances of Mandarin recycling repair are annotated with sound profiles including: pitch, duration, loudness, silence, and other observable prosodic cues (i.e. sound stretch and cut-offs). The study further introduces some possible applications of the current corpus, such as the implementation of the annotated data for analyzing the correlation between sound profiles of Mandarin repair and the interactional function of the repair. The goal of constructing the corpus is to facilitate an interdisciplinary study that concentrates on broadening the interactional linguistic theory by simultaneously paying close attention to the sound profiles emerged from interaction.

Keywords: Mandarin disfluency, conversational prosody, human interaction

1. Introduction

Repair (also known as disfluency) is a commonly occurring phenomenon in human interaction. In the process of conversational exchanges, speakers often stop before the end of their speaking turns to make adjustments, i.e. to add, to correct, to elaborate, or to qualify what they have said. In the process of executing the same-turn self-repair, various linguistic resources, including grammatical elements, sound manifestations, even non-verbal gestures, all are involved in order to facilitate the production and comprehension of the on-going speech. Serving various functions, such as holding the speaking turn to search for the following lexical item or content of the conversational exchange, the action of repair or disfluency itself involves linguistic representations at not only syntactic or semantic levels, but also prosodic level, and mostly, interaction among the interlocutors.

This study provides a preliminary description of a manually annotated speech corpus that concentrates on the sound profiles of repair/disfluency in naturally occurred, face-to-face Mandarin interaction. Specifically, the current study examines the sound profiles of the particular example of *recycling*, or disfluent repetition, defined as "a brief, sometimes a longer repeat or re-saying of part of the utterance occurring in a conversational turn," following Schegloff (1987). One of such examples of recycling is presented in the following example (1):

(1) E: ta   na   shihou   **mai- [R1] mai[R2]**
         3sg   that time   buy   buy
         zhege   song   ipod
         this-classifier give   PN

    E: '(At) that time, it was **buying- buying** this one and getting one IPod free.'

Examples of Mandarin recyclings are identified and extracted from video-tapped and audio-tapped conversations, and each instance is tagged manually in terms of the prosodic realization, i.e. the length, pitch height, loudness of the R1 and R2 in carrying out the repair, as well as the silent pause around and in between R1/R2, and additional notations on whether there are sound cues such as cut-offs or sound stretch. All these acoustic measurements are taken by the combined approaches of both acoustic records and impressionistic judgments (cf. Benkenstein & Simpson, 2003). Eventually, a corpus of Mandarin repair/disfluency is constructed, with wealthy information on the prosodic profiles of each recycling instance annotated.

Here the paper focuses on how the tag set of sound profiles of each repair instance are decided by the aforementioned methodology. Some preliminary applications of the annotated speech interaction will be introduced, as well as possible further applications of the current corpus data. The eventual goal is the construction of a speech corpus of Mandarin disfluency annotated with prosodic information, while the tag set is emerged from actual interaction and better reflects behaviors and actions observable from human interaction.

The paper is organized as follows. Section 2 provides background information about the previous studies on speech disfluency. Section 3 describes the data and methodology incorporated in the present study, while section 4 explains in detail the annotation schema for profiling prosodic features of Mandarin recycling repair. Section 5 discusses some applications of the corpus and its significance toward analyses of interaction. Finally, section 6 is the conclusion and directions for future research.

2. Theoretical background and related research on repair/disfluency

2.1 Research on repair/disfluency

The phenomenon of same-turn self-repair has been the focus of studies in the relevant fields, including psycholinguistics (e.g. Levelt, 1989; Levelt & Cutler 1983), computational linguistics (e.g. on disfluency in speech: Shriberg, 1994; 1995; Tseng, 2003; 2006), general linguistics (e.g. Fox & Jasperson, 1995; Fox et al., 1996; Local, 1992; Fox et al., 2009), and also in conversation analysis (e.g. Jasperson, 1998; Schegloff, 1987; Schegloff et al., 1977)

Most of the earlier studies on repair/disfluency focused on the phenomenon mainly in English. It was not until Fox et al. (1996) that there had been discussion about repair in other languages such as Japanese. Some previous studies that focused on repetition repair in other languages include: Benkenstein and Simpson's study on the phonetic correlates of self-repair involving word repetitions in German speech (2003); Henry and Pallaud discussing word segments and repeats in French Speech (2003); and Tseng discussing repetitions in spontaneous Mandarin (2003). It is worthy mentioning that Fox et al. reported a typological study of the site of initiation in same-turn self-repair across 7 languages (2009).

As mentioned, the current study concentrates on the specific type of *recycling* repair. There are several reasons that this particular method of carrying out same-turn self-repair has been chosen as the main focus: first of all, as shown in some past quantitative studies on Mandarin repair (cf. Tseng, 2003; 2006), this type of repetition repair is the most frequent type of repair in Mandarin conversation. Moreover, Fox et al. also suggests that Mandarin speakers consistently initiate repair after the word is recognizably completed (2009). Most of all, the preference of initiating Mandarin repair after recognizable completion provides a sound justification to compare the sound realization of repeated words or phrases in doing the repair: since the recycling would be a complete repetition of the same word or phrase, it actually allows for a simple comparison of the sound realizations of R1 and R2.

2.2 On conversational phonology

Some of the past studies on repair within the filed of interactional linguistics have paid attention to the discussion of the relationship between prosody and interaction in conversation (cf. Schegloff, 1979; 1987). However, it is not until recently that interactional linguists have started paying attention ot the organization of phonetic and phonological details in natural conversation. The approach "conversational phonology" (cf. Local & Local, 1989) or "interactional prosody" (cf. Couper-Kuhlen & Selting, 1996) suggests incorporating an approach to the study of the sound system with the following theoretical points:

- The material considered derives entirely from naturally occurring face-to-face conversational interaction
- The analysis attempts to prejudge as little as possible the salience of phonetic features
- The analysis seeks explicitly to motivate and warrant the functional categories employed by reference to the observable behaviour of the conversational participants (Kelly & Local, 1989: 263)

In other words, "no invented or hypothesized material" or pre-determined phonetic or phonological categories are assumed or "allowed to contribute to the analysis" (Kelly & Local, 1989: 263). Instead, the conversational prosody approach advocates an "impressionistic" analysis by closely listening to the production of real speech and notating phonetic details which a trained ear could perceive, including properties such as pitch, loudness, tempo, and others (Kelly & Local, 1989).

One of the studies that bear implications for the methodology adopted in the current study is Benkenstein and Simpson's research on the phonetic correlates of self-repair involving word repetitions in German speech (2003). In their study, they provide detailed phonetic descriptions of the sound realization in self-repair sequences involving repetitions of words in German speech by incorporating the approaches of both impressionistic auditory and acoustic records. Following Berkenstein and Simpson (2003), in the current study, both the precise acoustic records and impressionistic judgments are incorporated as the main methodology to analyze and code the sound production of the R1 and R2 of the disfluent repetition.

3. Data and methodology

3.1 Data

The data of the present study consists of 7 video- and/or audio-tapings of naturally occurring, face-to-face conversational interaction in Mandarin Chinese. Each segment of recording lasted from 15 minutes to about 1 hour. The total recording time added up to about 334 minutes. Each conversation was lead by two or multi-party participants, and there were a total of 10 female participants recorded. All speakers except for one are native speakers of Taiwanese Mandarin [1]. Furthermore, all participants were aware of the recording for the purpose of a linguistic research, though they may or may not be aware of the nature of the study. The recordings were casual and spontaneous; there were no prescribed topic set for each recording[2].

---

[1] Note that of all the 9 speakers recorded, the data produced by one other speaker was not considered in the present study either, as she's from another county in Taiwan that's different from all the other speakers.

[2] For the ethnographic information about each participant, please refer to Chen (2011) that incorporates the same conversational data for analysis.

After recordings, part of the conversations was transcribed based on *turn constructional unit* from conversation analysis approach as the basic unit. Then the next step was to identify instances of Mandarin disfluent repetitions from the transcription. To extract instances of recyclings (i.e. the repetitions of the verb **mai** *to buy* in (1)) from the current conversational data is a fairly straightforward task in simply searching for repetitions of words or phrases from the transcription/recordings. This does not entail, however, that all examples of repetitions are considered examples of recycling for repair. At least repetitions for the purpose of emphasizing were excluded.

In all, there were approximately 260 cases of recyclings identified, excluding examples that occurred in overlapped turns or produced with background noises. This collection of recycling repair examples will be referred to as "the recycling corpus" in the following discussion.

3.2 Methodology

To analyze the sound production, various acoustic measurements are made for the R1 and R2 of each token of repair. As mentioned, the methodology for acoustic measurements is similar to Benkenstein and Simpson (2003) in that the approaches of both acoustic records and impressionistic auditory are incorporated. The acoustic measurements are carried out by using the computer software Praat (© Boersma & Weenink (2007)). Additional judgments would be made based on the analyst's impressionistic interpretation of most of these auditory cues, following the impressionistic approach from conversational phonology (Kelly & Local, 1989). Detailed description of the annotation procedures for the prosodic profiles is provided in the next section.

4. Annotations for prosodic profiles

The prosodic profiles for the R1 and R2 of each recycling include: **pitch**, **loudness**, **silence**, and **duration**. Additional perceptible prosodic cues such as **cut-off** and **lengthening** are otherwise noted as well.

4.1 Pitch

The pitch height of the onset of the R1 and R2 of each recycling is measured and then compared. The pitch height refers to the fundamental frequency (F0) and is recorded in Hertz (Hz). If the onsets of the R1 and R2 don't yield any measurable pitch contour (such as an onset with a fricative [s]), the pitch height of the initial vowel would be measured instead.

Sometimes when the F0 difference between R1 and R2 is too small to be considered as hearable difference, the measurement of *semitone* would be incorporated to help determine if the R1 and R2 might be perceived as realized at the same pitch height.

4.2 Loudness

Loudness refers to the relative amplitude between the R1 and R2 of the repair. The measurement is taken by locating the point of highest intensity in R1, compared with the intensity of the same point in the segment of R2. The measurement taken by using Praat is recorded in decibels (dB). After the measurement of intensity is recorded, the result is checked against the researcher's impressionistic judgment. If the measurement reported in decibels is not consistent with what the analyser has perceived as the louder segment of the repair, the exact measurement would not be reported.

4.3 Silence

The profile of silence records, in seconds, any audible pause located before the R1, after the R2, or in between the R1 and R2 of the repair. In the current study a cut-off point at 0.2-second has been applied, following Jasperson's study on focused English repair (1998). Any silent pause under 0.2 seconds is considered as part of the articulatory process and corresponded to a "beat of silence", which is followed by the immediate resumption of talk (Jasperson, 1998). It is thus treated as having no significant impact on the processing of the repair. On the other hand, silent pauses longer than 0.2 seconds are taken as serving possible interactional functions, and are reported with exact measurements.

4.4 Duration

Duration refers to the length of the R1 and R2 of the repair, reported in milliseconds. The measurement of duration is taken starting from the onset to the ending vowels or nasal consonants of the word or phrase in both the R1 and R2.

4.5 Other prosodic cues

4.5.1. Cut-off

Here cut-off is defined as an articulatory closure that interrupts the air stream, and it typically involves glottal or other stop closures (Jasperson, 1998; 2002). To determine if there is a cut-off occurring, the analyzer follows the impressionistic description of ways in which the cut-off is articulated, as proposed by Jasperson (1998; 2002). At least two types of cut-offs are distinguished: "glottalized" cut-offs, which have salient interruption glottalization, and "soft" cut-offs that have either unnoticeable or no interrupted glottalization (cf. Jasperson 1998; 2002). The glottalized cut-offs are indicated by a percent sign "%" while the soft ones are marked by a dash "-".

4.5.2. Sound stretch (lengthening)

The prosodic cue of sound stretch (or lengthening) records any perceptible prolongation on any syllable of the R1 and R2 of the repair. To determine if there is perceivable sound stretches, an impressionistic judgment is made and the result is marked on the transcription of the interaction when any lengthening is observed. It should be noted that for the identification and annotation of both cut-offs and sound stretch, they are made mostly based on the impressionistic judgment from the

researcher or the transcribers of the data.

5. Applications of the corpus and significance: toward analyses of interaction

The construction of the corpus of Mandarin repair serves the function as a rich speech resource that provides detailed acoustic and prosodic information of naturally occurred interaction. As one of the applications of the current corpus, Chen (2011) reports 6 sound patterns that consist of the various acoustic cues discussed above, and explores the correlation between each of the sound patterns and the possible interactional functions of the recycling as a specific type of Mandarin disfluency. It is suggested, for instance, that the sound pattern [longer R1 with cut-off] + [significant silence] + [R2 at higher pitch height] is used in Mandarin recyclings for the function of *word* or *content* search, while the pattern [longer R1 with cut-off] + [R2 at the same pitch height] has been identified to associate with the function of projecting a *continuation* of further explanation of the current discourse topic (Chen, 2011). Another prosodic pattern [longer, higher R1 initiated by lengthening or cut-off] has been identified with the function of *restarting* new information after substantial production problems or in second pair part (Chen, 2011). The findings from the study thus highlight the *interaction-specific*, *sequence-specific*, and *function-specific* examples of Mandarin repair in relation to the use of the prosodic patterns as combinations of various prosodic cues (Chen, 2011).

In addition to a 'synthetic' analysis on how the various prosodic cues form sound patterns in correlating with specific interactional function, the corpus provides a collection of data on how each individual sound cue may reflect aspects of human interaction. For instance, as it has been suggested in the literature of interactional inguistics that cut-offs are one of the signals for indicating the initiation of repair in conversation (cf. Schegloff, 1979), further study can be carried out by focusing on cut-off/ sound stretch as a cue to initiate repair in the language. Finally, but not the least, the corpus reported in the current study may also provide the data for exploring Mandarin repair/disfluency in relation to correction of tones in Mandarin interaction.

The significance of the manually tagged corpus of Mandarin repair in terms of prosodic manifestations lies mainly in that the tag set for the prosodic profiles of each repair is not decided based on pre-determined phonetic or phonological categories. Instead the annotations utilized in the current speech corpus reflect the sound representations emerged from actual human interaction. Thus further applications of the current corpus annotated with prosodic information have the advantage of better demonstrating that "a close consideration of the content and structure of conversational talk can contribute to the identification and elaboration of relevant functional categories" in interaction, as suggested by Kelly and Local (1989).

6. Conclusion and directions for future studies

The paper describes the construction of Mandarin speech corpus with focus on the prosodic profiles of repair/disfluency extracted in conversational interaction. As repair is such an interaction-significant device, the current project presents a language recourse that highlights actions in human interaction, such as repair, in correlation with sound profiles of the language production. The paper provides in details how the annotations of each disfluent repetition from the current corpus are decided by the methodology of both acoustic measurements and impressionistic judgments. Some possible further applications of the corpus are introduced.

As a next step for the current research, the annotation could be further validated by double annotations and further evaluations. The eventual goal is to extend the current resource based on face-to-face interaction into a *multimodal* resource. It is suggested that the multimodality model of annotations on various linguisitc levels (including prosody, syntax, pragmatics, gestures, etc.) proposed by Blache et al. (2008) and Blache et al. (2010) can be implemented for the building of the multimodality resource for Mandarin Chinese.

7. Acknowledgements

8. References

Benkenstein, R. and Simpson, A. (2003). Phonetic correlates of self-repair involving word repetition in German spontaneous speech. In *DiSS '03*, pp.81--84.

Blache, P., Bertrand, R., Ferré, G. (2008). Creating and exploiting multimodal annotated corpora. In *LREC 08*, pp. 1773--1777.

Blache, P., Bertrand, R., Guardiola, M., Guénot, M.-L., Meunier, C., Nesterenko, I., Pallaud, B., Prévot, L., Priego-Valverde, B., Rauzy, S. (2010). The OTIM formal annotation mode: a preliminary step before annotation scheme. In *LREC 10*, pp. 3262--3267.

Boersma, P. and Weenink, D. (2007). Praat: Doing phonetics by computer. www.praat.org.

Chen, K. (2011). Sound patterns in Mandarin recycling repair. PhD Dissertation. University of Colorado.

Couper-Kuhlen, E. and Selting, M. (1996). Towards an interactional perspective on prosody and a prosodic perspective on interaction. In E. Couper-Kuhlen, and M. Selting (Eds.), *Prosody in conversation: Interactional studies*, Cambridge: Cambridge University Press, pp.11--56.

Fox, B., Hayashi, M., Jasperson, R. (1996). Resources and repair: A cross-linguistic study of the syntactic organization of repair. In E. Ochs, E. Schegloff, & S. Thompson (Eds.), *Interaction and grammar*,

Cambridge: Cambridge University Press, pp. 185--237.

Fox, B. and Jasperson, R. (1995). A syntactic exploration of repair in English conversation. In P. Davis (Ed.), *Alternative linguistics: Descriptive and theoretical modes*, Amsterdam/Philadelphia: John Benjamins, pp. 77--134.

Fox, B., Wouk, F., Hayashi, M., Fincke, S., Tao, L., Sorjonen, M.-L., Laakso, M., Hernandez, W. (2009). A cross-linguistic investigation of the site of initiation in same-turn self-repair. In J. Sidnell (Ed.) *Conversation analysis: Comparative perspectives*, Cambridge: Cambridge University Press, pp. 61--103.

Henry, S. and Pallaud, B. (2003). Word fragments and repeats in spontaneous spoken French. In *DiSS '03*, pp.77--80.

Jasperson, R. (1998). *Repair after cut-off: Explorations in the grammar of focused repair of the turn-constructional unit-so-far*. PhD Dissertation. University of Colorado.

Jasperson, R. (2002). Some linguistic aspects of closure cut-off. In C. Ford, B. Fox, & S. Thompson (Eds.), *The language of turn and sequence*, Oxford: Oxford University Press, pp. 257--286.

Kelly, J., Local, J. (1989). *Doing phonology: Observing, recoding, interpreting*. Manchester: Manchester University Press.

Local, J. (1992). Continuing and restarting. In P. Auer, A. di Luzio (Eds.), *The contextualization of language*, Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 273--296.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.

Levelt, W. and Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics* 2 (2), pp. 205--217.

Schegloff, E. (1979). The relevance of repair to syntax-for-conversation. In T. Givón (Ed.), *Discourse and syntax*, London: Academic Press, pp. 261--286.

Schegloff, E. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organisation. In G. Button & J. Lee (Eds.), *Talk and social organisation*. Clevedon, Avon: Multilingual Matters, pp. 70--85.

Schegloff, E., Jefferson, G., Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language* 53 (2), pp. 361--382.

Shriberg, E. (1994). Preliminaries to a theory of speech disfluencies. PhD Dissertation. University of California.

Shriberg, E. (1995). Acoustic properties of disfluent repetitions. In *ICPhS*. Stockholm, Sweden, pp. 384--387.

Tseng, S.-C. (2003). Repairs and repetitions in spontaneous Mandarin. In *DiSS '03*, pp. 73--76.

Tseng, S.-C. (2006). Repairs in Mandarin conversation. *Journal of Chinese Linguistics* 34 (1), pp. 80--120.