Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop.
10-12 September 2005, Aix-en-Provence, France, pp. N-N.

A preliminary study of Mandarin filled pauses

Yuan Zhao & Dan Jurafsky
Stanford University, California, U.S.A.

Abstract

The paper reports preliminary results on Mandarin filled pauses (FPs), based on a large speech corpus of Mandarin telephone conversation. We find that Mandarin intensively uses both demonstratives (*zhege* 'this', *nage* 'that') and *uh/ mm* as FPs. Demonstratives are more frequent FPs and are more likely to be surrounded by other types of disfluency phenomena than *uh/mm*, as well as occurring more often in nominal environments. We also find durational differences: FP demonstratives are longer than non-FP demonstratives, and *mm* is longer than *uh*. The study also revealed dialectal influence on the use of FPs. Our results agree with earlier work which shows that a language may divide conversational labor among different FPs. Our work also extends this research in suggesting that different languages may assign conversational functions to FPs in different ways.

1. Introduction

Filled Pauses (FPs) are prevalent in Mandarin spontaneous speech and pose a major problem in Mandarin speech recognition. Although much work related to English and other languages has been done [1, 3, 5, 6], little empirical work on Mandarin FPs in spontaneous speech has been carried out. Previous work on Mandarin spontaneous speech, such as Tseng [7], mainly focuses on repairs and repetitions. The lack of research on FPs directly leads to the confusion of FP judgments and random use of characters in transcribing Mandarin speech. It is therefore necessary to identify the distinctive forms of FPs in Mandarin and to investigate their acoustic features and discourse functions.

The goal of the present work is to carry out a descriptive study of Mandarin FPs. More specifically, the study aims at identifying FPs and investigating their basic acoustic properties and distribution across syntactic units. In addition, it also examines the sociolinguistic variables that might influence speakers' use of FPs.

2. Method

The research was mainly based on the data drawn from LDC 98-HUB5 Mandarin corpus of telephone conversations, in which FPs such as *uh* and *mm* are hand-labeled by LDC. There are 37 conversations in the corpus, which have complete speaker information. Only the callers' information can be identified, which is shown in Table 1.

Table 1: Speaker information of 98-HUB5

Gender	Female	17
	Male	20
Dialect	Northern	21
	Southern	16
Age	Range (yr)	20-39
	Average (yr)	28.6
Education	Range (yr)	10-22
	Average (yr)	17.6

In addition, in order to see how FPs vary across corpora, another two LDC telephone conversation corpora, 2003-HUB5 and 96-CallHome, were also used for identifying distinctive forms of FPs in Mandarin. Demonstrative FPs were hand-labeled by the first author.

3. Results and discussion

3.1 How many FPs are there in Mandarin?

Four FPs are systematically used in three corpora, including *uh, mm,* and the demonstratives *zhege* (literally "this") and *nage* (literally "that"), as shown in Table 2.

Table 2: Occurrence of FPs in 98 HUB-5 per 1000 words

Demonstrative	*nage*	4.51
	zhege	2.17
	Total	6.68
Reduced Vowel	*uh*	2.55
	mm	1.46
	Total	4.01

In general, the total frequency of FPs (10.69 per 1000 words) seems to be smaller than in English. In the CallHome English corpus, for example, *um* occurs 7.15 times per 1000 words, and *uh* occurs 7.10 times per 1000 words.

Like Japanese and Spanish, besides reduced vowel FPs, Mandarin intensively employs demonstratives as FPs [4, 9]. In fact, demonstratives form the largest FP category in Mandarin. We also investigated other potential FPs, but found only these four (*nage/zhege/uh/mm*) seemed to act as true FPs. For

Table 3: Acoustic properties of FPs and comparison between FP and non-FP demonstratives

		uh	*mm*	FP *nage*	Word *nage*	FP *zhege*	Word *zhege*
Duration	Range (sec)	0.068-0.944	0.144-1.376	0.18-0.547	0.14-0.252	0.194-0.691	0.13-0.263
	Average(sec)	0.277	0.483	0.383	0.195	0.382	0.177
Pitch	Range (Hz)	97.6-224.1	90.2-198.6	98.8-187.0	101.2-269.0	101.0-319.8	79.7-261.2
	Average (Hz)	135.5	133.9	139.4	162.9	158.1	154.1

example, we examined 200 instances of discourse particles transcribed as "oh"; we found that in each case particles with rounded vowels do not act as FPs, but as backchannels suggesting the addressee's attentiveness to the ongoing conversation.

3.2 Acoustic features of FPs

The basic acoustic features of the FPs, such as duration and pitch movement were measured from a sample of 100 FPs produced by male speakers, which are summarized in Table 3.

It was found that the duration of *mm* is over 1.5 times longer than that of *uh* and the difference is highly significant (t=3.306, p<.002, MD=.2053). In English, *uh* and *um* contrast mainly in the delays they initiate [1]. However, in the 100 samples of Mandarin *uh/mm* FP, less than 5 cases have pauses around them, which might indicate that Mandarin speakers tend to prolong the vowel or the nasal to fill an entire pause. This suggests that Mandarin FPs divide their labor according to the length of pause they need to fill: *mm* tends to fill a longer pause and *uh* tends to fill a shorter pause.

We also compared demonstrative FPs with their non-FP uses. It was found that the duration of both *zhege* and *nage* is significantly longer than their non-FP uses respectively (t=5.531, p<.001, MD=.2052) (t=6.278, p<.004, MD=.1882). In this case, FP demonstratives are around two times longer than their non-FP counterparts. No significant difference between the duration of *zhege* and *nage* was found. In addition, the statistics shows that the pitch of FP *nage* is significantly higher than that of non-FP *nage* (t=1.356, p<.024, MD=23.52).

3.3 Location of FPs

The second study examined the locations where FPs occur. FPs were found to occur mainly before syntactic constituents, such as before clauses, as shown in (a); before predicate VPs as shown in (b) and before complements as shown in (c). Another location where FPs frequently occur is within an NP constituent, such as between a modifier and a noun head, as shown in (d):

a. **nage** [$_S$ wo mingtian you shijian].
that I tomorrow have time
'mm, I have time tomorrow.'

b. wo **nage** [$_{VP}$ du-guo liang bian].
I that read-Asp two CL
"I, uh, read it twice'

c. [$_{VP}$ huainian **nage** [$_{NP}$ daxue shenghuo]].
miss that college life
"(I) miss, uh, the life in college.'

d. [$_{NP}$ gongzuo de **nage** [$_N$ mafan]]
works NOM that trouble
'The work's, uh, trouble'

The distribution of FPs occurring in different syntactic contexts was examined, as summarized in Table 4 and Figure 1.

Table 4: Distribution of FPs among syntactic units

	Before Clause	Before Predicate	Before Complement		Within NP
			S	NP	
nage	26.4%	15.1%	6%	34.0%	18.4%
zhege	29.4%	14.7%	3.9%	28.4%	23.5%
uh	77.7%	12.6%	4.7%	1.5%	3.5%
mm	86.3%	7.6%	3.5%	0.9%	1.7%

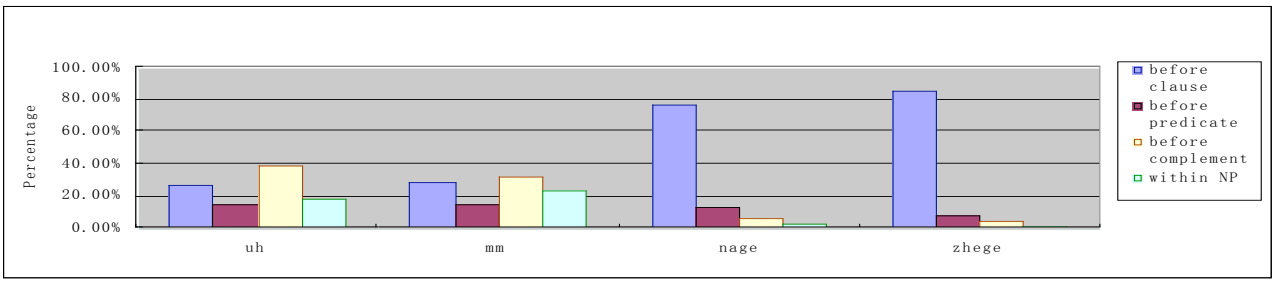

Figure 1: Distribution of filled pauses

Figure 1 suggests that most *uh* and *mm* are used clause initially. In addition, the occurrences of *uh* and *mm* decrease as the syntactic units become smaller and simpler. Grammatical weight theory predicts that speaker would encounter more planning problems before larger syntactic constituents [8]. Our result on FPs of *uh/mm* is in accordance with what grammatical weight theory predicts. In this case, the larger and more complex a constituent is, the more likely an *uh/mm* FP is to occur before it.

Contrasting with *uh/mm* FPs, demonstrative FPs are most likely to occur before complements, especially NP complements. In addition, unlike *uh* and *mm,* which are rarely used between a modifier and a noun head, demonstratives also occur frequently in this nominal environment. As shown in Table 4, 34% of *nage* and 28.4% of *zhege* occur before NP complement and 18.4% of *nage* and 23.5% of *zhege* occur before noun head. Demonstratives *nage* and *zhege* occur in nominal environments 52.4% and 51.9% of the time, respectively. Demonstrative FPs therefore contrast greatly with *uh/mm* in occurring in nominal-related disfluency environments. This suggests that demonstrative FPs may play a role in nominal search in Mandarin. We can formalize this suggestion as the "Nominal-Search Hypothesis": speakers are more likely to choose demonstratives as FPs when they encounter nominal-search problems in Mandarin

3.4 Disfluencies around FPs

Research such as [1, 2, 8] has shown that some FPs signal larger planning problems than others. One way to investigate which FPs represent bigger planning problems is to examine neighboring disfluencies occurring around the FPs. The distribution of disfluencies before and after the four types of FPs is summarized in Table 5.

Table 5: Disfluencies around filled pauses

	before	after	repetition	Total
zhege	1.96%	23.5%	8.8%	34.3%
nage	0.99%	20.3%	8.0%	29.3%
uh	7%	7.8%	0	14.8%
mm	5.8%	4.9%	0	10.7%

The result in Table 5 suggests that demonstrative FPs are more likely to be accompanied by other types of disfluency phenomenon than *uh/mm*. For example, demonstrative FPs are often repeated several times and followed by other disfluencies. In comparison, *uh* and *mm* are not that likely to be accompanied by other type of disfluencies. The fact that demonstrative FPs are more likely than *uh* and *mm* to be followed by other disfluencies, and far more likely to be repeated, may suggest that demonstratives are more indicative of more serious planning problems. On the other hand, speakers may simply be using other strategies to indicate severe planning problems with *uh* and *mm*. For example, *uh* and *mm* could be prolonged to indicate severity, as suggested in [1]. This issue clearly calls for further investigation.

3.5 Social variables and filled pauses

Finally, the study examined speakers' demographic factors and their influence on the use of FPs. These factors include gender, age, education and place of growing up.

Figure 2 shows the distribution of FPs among individual southern and northern speakers. A t-test shows that place of growing up has a significant effect on the number of FPs being used. Speakers growing up in the south of China use significantly more FPs than those grow up in the north (t=3.431, p<.002). The main distinction lies in the total amount of *uh*. Southerner use significantly more *uh* as FPs than northerners (t=2.888, p<0.002). We found no significant effect in the use of nasal *mm* between northern and southern speakers

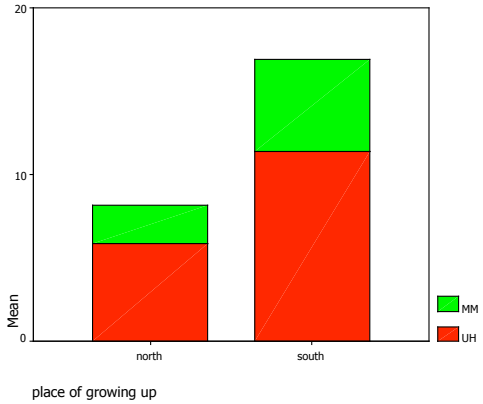

Figure 2: *uh/mm* used among southern speakers and northern speakers

Sex, age and education do not have a significant effect on the use of FPs. Previous work on English FPs revealed that FP rates are related to demographic factors such as sex. Men produced significantly higher rates of FPs than women [6]. Our study does not find similar results in Mandarin; sex does not have a significant effect on the use of FPs in Mandarin (t=-.315, p<.755). Although sex alone does not have significant effect on the use of FPs, the interaction between sex and place of growing up turns out to be significant (F=3.823, p<.019). The mean number of FPs used by a southern female speaker is the highest and that of a northern male speaker is the lowest.

In addition, we found no significant effect of education level on the use of FP, although we did find a trend for speakers with more education to use fewer FPs than those with less

education.

4. Conclusion

The paper presents preliminary results of a study on Mandarin FPs. Through comparing three corpora, four major FPs were identified. It revealed that apart from *uh* and *mm*, Mandarin speakers intensively use demonstratives as one major type of FPs. Demonstratives and *uh/mm* FPs differ in their distribution among syntactic contexts. Namely, demonstratives are more frequently used in nominal-searching environments, while *uh* and *mm* are more likely to be used clause-initially. The two types of FPs also contrast greatly in the amount of surrounding disfluencies. Demonstratives were found to be more likely accompanied by repetitions, false starts etc., while *uh/mm* were only rarely accompanied by other disfluencies. In addition, southern speakers tend to make use of more FPs in conversation than northern speakers, which may suggest dialectal influence on the use of FPs in Mandarin.

This paper provides only a preliminary picture of Mandarin FPs. Much more work needs to be done in the future.

5. Acknowledgements

This research was partially supported by Stanford University through a Deans' program summer research fellowship. Thanks to Huihsin Tseng and Rebecca Starr on help scripting.

6. References

[1] Clark, Herbert & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speech. *Cognition, 84,* pp. 73-111.

[2] Clark, Herbert & Thomas Wasow. 1998. Repeating words in spontaneous speech. *Cognitive Psychology 37*, pp. 201-242.

[3] Eklund, Robert & Elizabeth Shriberg. 1998. Crosslinguistic disfluency modelling: A comparative analysis of Swedish and American English human-human and human-machine dialogues. *Proceedings of ICSLP'98*, 30 November-5 December 1998, Sydney, vol.6, pp. 2631-2634.

[4] Hayashi, M. and Yoon, K. to appear. "A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble." *Studies in Language.*

[5] Quimbo, Felix C M., Tatsuya Kawahara & Shuji Doshita. 1998. Prosodic analysis of fillers and self-repair in Japanese speech. *Proceedings of ICSLP'98*, 30 November-5 December 1998, Sydney, pp.3313-3316.

[6] Shriberg, Elizabeth. 1996. Disfluencies in Switchboard. *Proceedings of International Conference on Spoken Language Processing*, Philadelphia, Addendum, pp. 11-14.

[7] Tseng, Shu-Chuan. 2003. Repairs and repetitions in spontaneous Mandarin. *Preceedings of DiSS'03,*5-8 September 2003, Goteborg, Sweden, pp.73-76.

[8] Wasaw, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change, 9,* pp.81-105.

[9] Wilkins, David P. 1993. Interjections as deictics. *Journal of Pragmatics, 18,* pp. 119-158.