

Tone or Toneless: The Prosody of Mandarin Filled Pauses

Yuan Zhao & Dan Jurafsky

[yuanzhao@stanford.edu](mailto:yuanzhao@stanford.edu); [jurafsky@stanford.edu](mailto:jurafsky@stanford.edu)

*Department of Linguistics, Stanford University, California 94305, U.S.A.*

Understanding the interaction between prosody and lexical tones in tonal languages like Mandarin is important for the construction of a phonetically-based theory of intonational phonology. We propose a new way to examine this interaction through the study of filled pauses. Previous research has shown that filled pauses such as *uh* and *um* in English normally have a low F0 and level or falling tone (O'Shaughnessy 1992) and the pitch of filled pauses can be predicted from the preceding F0 peak and the speaker's pitch range (Shriberg & Lickley 1993). These results suggest that there might be a speaker-specific and context-dependent pitch target for filled pauses. Recent research in Mandarin has shown that neutral tone in Mandarin, as a reduced form, has its own pitch target and at the same time is influenced by the preceding prosodic contexts (Chen & Xu, in press). A key goal of our research is to ask what kind of F0 information speaker of a tonal language stores in the mental lexicon for these filled pauses. One possibility is that they have a speaker-specific pitch value with less variation across contexts, as is true for English filled pauses. A second possibility is that filled pauses are lexical items realized with neutral tone in Mandarin. Yet another possibility is that filled pauses, like other syllables, have acquired canonical lexical tones with strong contextual variations. A final possibility is what is stored is some combination of phonological tones and speaker-specific pitch targets. In this paper, we examine the tonal features of Mandarin filled pauses and investigate the similarities and differences between the tones of filled pauses and the neutral tone.

Mandarin has four filled pauses: the distal (*nage* 'that') and proximal (*zhege* 'this') demonstratives, as well as cross-linguistically more common reduced forms *uh* and *mm* (Zhao & Jurafsky 2005). In this study, we extracted and annotated 203 filled pauses used by three Beijing Mandarin speakers (one male, two females) from the 1998 LDC CallFriend Mandarin Telephone Conversation Corpus. Average F0, duration and 10 time-normalized F0 values of the filled pauses and their preceding/following syllables were automatically obtained.

We found that Mandarin filled pauses are produced with a speaker-specific constant pitch value across the four different types of filled pauses. The F0 of the filled pauses is on average 45-50% above the speaker's baseline. As for the lexical tones of the four filled pauses, we found that *uh* is realized with a low level tone (slope= -10.6 Hz/sec) and *mm* is realized with a rising tone (slope=31.2 Hz/sec); demonstrative filled pauses preserve their original falling tone shape (slopes

for distal and proximal are -54.3 and -111.5 Hz/sec, respectively), but shift significantly downward in pitch register towards the constant mid-range pitch value. As for whether filled pauses have neutral tone, we found that, similar to neutral tone, there is a strong carryover effect between the F0 of the offset of the preceding syllable and the onset F0 of the filled pauses. This is the case even though filled pauses are often preceded by glottalization and thus a small intonational phrase boundary. Again like neutral tone, no anticipatory effect was found between the offset of filled pauses and the following syllable.

In general, our results suggest that the representation of filled pauses may include elements of both canonical tones and a speaker-specific constant pitch target. This together with the fact that filled-pause demonstratives are significantly lower in pitch than their lexical counterparts might support the hypothesis that the phonological representation of wordforms should include fine-grained phonetic details such as F0 information in addition to broad phonological categories (Johnson 1997; Pierrehumbert 2001).

References:

- Chen, Yiya. and Xu, Yi. (in press). Production of weak elements in speech -- Evidence from F0 patterns of neutral tone in standard Chinese. To appear in *Phonetica*.
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In Johnson & Mullennix (Eds). *Talker Variability in Speech Processing*. San Diego: Academic Press. pp. 145-16
- O'Shaughnessy, D. (1992). Recognition of hesitations in spontaneous speech. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, San Francisco, pp. 521-524.
- Pierrehumbert, Jenet B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee and P.Hopper (Eds.) *Frequency and Emergence of Linguistic Structure*. Amsterdam: Benjamin.
- Shriberg, Elizabeth & Robin Lickley (1993). Intonation of clause-internal filled pauses. *Phonetica* 1993, 50: 172-179.
- Zhao, Yuan & Dan Jurafsky (2005). A preliminary study of Mandarin filled pauses. Proceedings of Disfluency in Spontaneous Speech 05, Aix-en-Provence, pp. 179-182.

Thematic area: Session 1--variation, phonetic detail and phonological modeling

Presentation preference: oral or poster