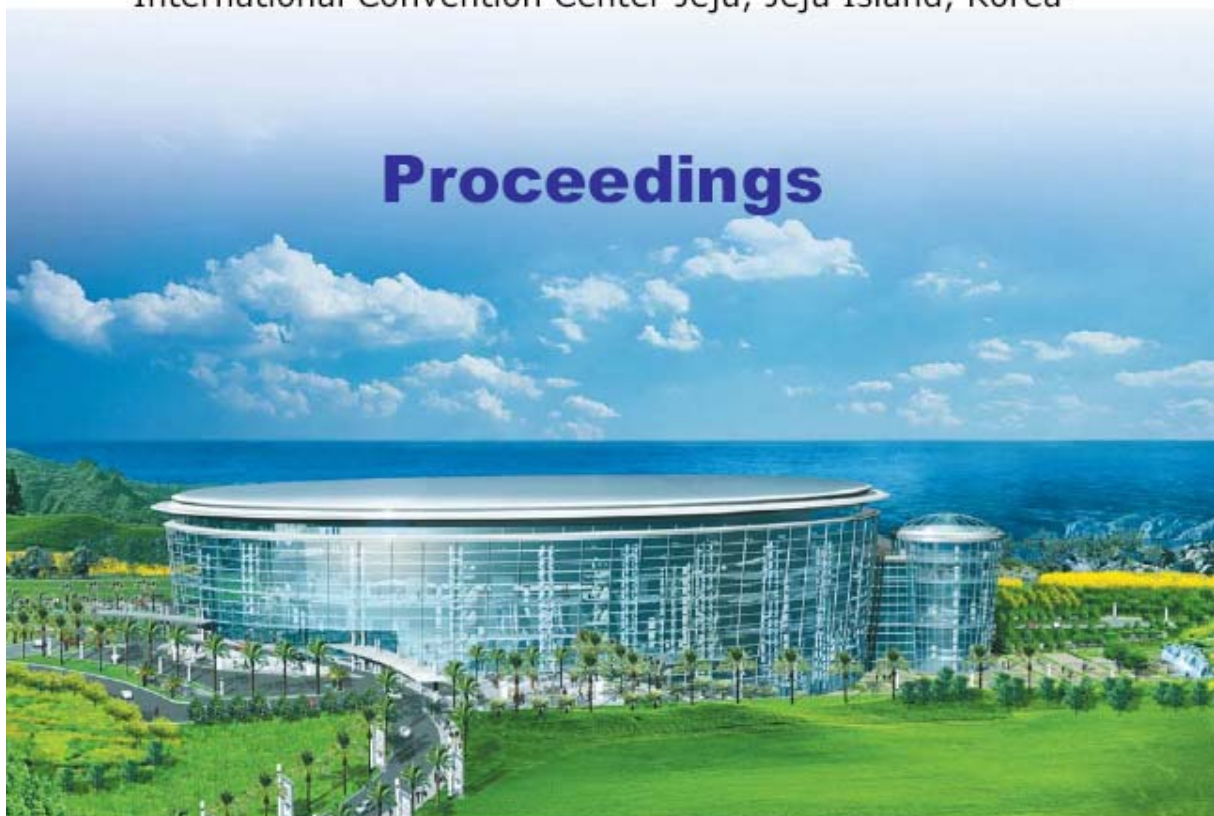


International Speech
Communication Association

INTER_SPEECH 2004 ICSLP

October 4-8, 2004
International Convention Center Jeju, Jeju Island, Korea



Lee, Tzu-Lun, Ya-Fang He, Yun-Ju Huang, Shu-Chuan Tseng & Robert Eklund. 2004.

Prolongation in Spontaneous Mandarin.

Proceedings of Interspeech (ICSLP) 2004, October 4–8 2004, Jeju Island, Korea, vol. III, pp. 2181–2184.

Prolongation in Spontaneous Mandarin

Tzu-Lun Lee, Ya-Fang He, Yun-Ju Huang, Shu-Chuan Tseng, Robert Eklund*

Institute of Linguistics, Academia Sinica, Taipei

* TeliaSonera Sweden AB, Farsta and NLPLab, Linköping University, Sweden
{ic221, yafang, tsengsc}@gate.sinica.edu.tw

Abstract

This paper presents a corpus-based study on prolongations in spontaneous Mandarin. Prolongations are mainly produced for hesitation, but also for emphasizing a discourse focus and signalling an explicit feedback. 786 prolongation occurrences are investigated in terms of the position, the part of speech and the segment and tone types. Prolongations are often found in word-final, phrase-final and utterance-medial positions. It is more likely to prolong in function words than in content words. However, in the case of monosyllabic words prolongations are more frequently found in function words, but in the remaining cases prolongations are more likely to be found in content words. Prolongations in transitive verbs, adverbs, nouns and particles show particularly high rates, while prolongations in intransitive verbs and aspectual adverbs are really rare. Especially, there is no prolonged adjective. Consonants are rarely prolonged in Mandarin and no particular effect is found for lexical tones.

1. Introduction

Speakers have various ways of signaling hesitation, including silences (unfilled pauses), filled pauses (sound likes “eh”) and so on. One such way to signal hesitation is to prolong speech sounds beyond their normal duration, a phenomenon known as prolongation. While prolongation has been the object of study within stuttering research for a long time, it has been somewhat neglected in the speech of non-stutters. Once considered a tell-tale sign of stuttering, it has later been shown to be a common phenomenon in non-stuttered speech in several languages, such as Swedish, English, Tok Pisin and Japanese. This paper studies prolongation in Mandarin.

2. Data

2.1. Mandarin Conversational Dialogue Corpus

The Mandarin Conversational Dialogue Corpus (MCDC) consists of eight transcribed and annotated Mandarin spontaneous conversations of a length of approximately eight hours [6]. Prolongations annotated in two conversations produced by two male and two female speakers were analyzed (s-1, s-2, s-3, s-4 hereafter). In the MCDC transcription system, all discourse markers and particles are written in capital letters. Discourse markers are words whose original meaning in spoken use disappears and the pragmatic function increases, similar to “well” in English. Discourse particles have different kinds of pragmatic functions in Mandarin, e.g. *EN* can sometimes be used for a prolonged hesitation without a lexicalized meaning, like the filler “uhn” in English. But sometimes its use is equivalent to “well” in English. In this system, discourse particles include the so-called fillers. In our

data, 49 fillers are prolonged. This paper adopts the Pinyin system for the romanized transcription. Mandarin has four marked tones: high level, rising, contour and falling (represented by 1, 2, 3 and 4) and one un-marked neutral tone (represented by 5).

2.2. Mandarin Prolongations (PR)

Segments perceived by the human annotators as markedly longer than in normal and fluent speech are prolongations [1], e.g. Ex1.

Ex1: wo3men5gong1si1 zaaaaiiii4 tai2bei3.
We company in Taipei.
Our company is in Taipei.

In Mandarin sound system, only nasals ([n] and [ŋ]) can be coda consonants, so the coda nasal is often amalgamated with the preceding nucleus. Therefore, a Mandarin PR is also defined as a pair of a vowel and its following coda which are prolonged simultaneously.

2.3. PR Rates

Table 1 summarizes the results of identified PRs in the MCDC. The occurrences of discourse particles (including fillers) and markers are included in the statistics. But in later analyses of phrase position (section 3.3.2), they are not considered.

Table 1: Summary statistics. WT=Word Tokens.

	s-1	s-2	s-3	s-4	Total
WT (#)	3,998	7,219	6,400	4,831	22,448
PR (#)	99	203	218	266	786
PR/WT (%)	2.48%	2.81%	3.41%	5.51%	3.50%

The overall PR rate of 3.5% per word is greater than that in Swedish (1.27%), in Japanese (1.13%) and in American English (0.5%) [2], [4], [3]. The higher PR rate in our data may have to do with the fact that our Mandarin data is long and free conversations. The speakers have enough time to plan their utterances.

2.4. PR Functions

Expectedly, the main function a PR has is to allow the speaker more time for the language planning, i.e. hesitation. However, we find more functions for non-filler PRs: emphasis and response. An emphasis PR is normally associated with a discourse focus on the corresponding word, e.g. Ex 2.

Ex 2: wo3 jue2de2 pa2 shan1 hai2 bu2 cuoooo4.
I think climb mountain still not bad.
I think that mountain-climbing is not too bad.

When a PR signals a feedback, we define it a response PR. They are often produced in the form of discourse particles such as O and OH etc., usually in the utterance-initial position.

Table 2: Functions of Non-filler PRs. * Including fillers.

		Hesitation	Emphasis	Response	Total*
s-1	PR (#) PR (%)	37/99 37.37%	23/99 23.23%	11/99 11.11%	99
s-2	PR (#) PR (%)	138/203 67.98%	43/203 21.18%	2/203 0.99%	203
s-3	PR (#) PR (%)	109/218 50.00%	109/218 50.00%	0 0.00%	218
s-4	PR (#) PR (%)	141/266 53.01%	94/266 35.34%	9/266 3.38%	266
	Total*	425/786 54.07%	269/786 34.22%	22/786 2.80%	786

Table 2 illustrates the results. The ratio of hesitation PRs is much greater than that of the emphasis and response PRs. However, it is clearly supported by the data that prolongation is also used to express a special discourse focus, in addition to hesitation. As shown in Table 2, about the half of the PRs produced by s-3 are identified to have the hesitation function and the other half have the emphasis function.

3. Results

3.1. PR position in the word, phrase and utterance

Table 3 summarizes results of an analysis of PR position in three different syntactic units: word, phrase and utterance. For the analysis of word position, 435 monosyllabic words are not taken into account. And for the analysis of phrase position, 131 discourse particles and 66 markers are excluded.

Table 3: PR position in word, phrase and utterance.

	Initial	Medial	Final
PR words (#)	14	2	335
PR words (%)	1.78%	0.25%	42.62%
PR phrases (#)	180	109	300
PR phrases (%)	22.90%	13.87%	38.17%
PR utterances (#)	89	599	98
PR utterances (%)	11.32%	76.21%	12.47%

As shown in Table 3, word initial and medial PRs are rare in general. Prolonged syllables are often located in the word-final and phrase-final positions, 42.62% and 38.17% respectively. But with regard to the position in utterances, PRs are mainly found in utterance-medial positions. However, for prosodic boundaries such as words and phrases, Mandarin shows a clear tendency for final prolongations.

The PR ratio among word-initial, -medial and -final position is approximately 4-1-95. This ratio is very different from the 30-20-50 ratio both in Swedish [2] and in American English [3], but comparable to the 10-5-85 ratio in Japanese [4] and the 15-0-85 ratio in Tok Pisin [1].

3.2. Word classes of PRs

Table 4 shows the rates of PRs in content and function words. In the overall word tokens, content words are more frequently produced than function words in our data (58.65%-41.35%).

However, the number of PR function words is almost the same as that of PR content words. Nevertheless, referring the numbers to the overall word tokens, we find that prolongation is more likely to be found in function words (PR per word token is 4.18%) than in content words (2.96%). This result is similar to that for Tok Pisin [2], but different from that for Japanese (no preference was found) [4] and for Swedish [2].

Table 4: The PR rates of word classes. WT=Word Tokens.

	WT	PR words	PR words/WT (%)
PR content words (%)	58.65%	49.62%	2.96%
PR function words (%)	41.35%	50.38%	4.18%

3.3. Word classes and PR position

3.3.1. Word classes and PR position within word

Table 5 shows results of the distribution of PR function and content words in relation to their position within the word.

Table 5: Word classes and PR position in the word.

	1-syll. words	2-syll. words		> 2-syll. Words		
		initial	final	initial	medial	final
PR (#)	435	13	308	1	2	27
PR (%)	55.34%	1.65%	39.19%	0.13%	0.25%	3.44%
PR content words (%)	23.41%	1.27%	21.37%	0.13%	0.25%	3.18%
PR function words (%)	31.93%	0.38%	17.81%	0.00%	0.00%	0.25%

As shown in Table 5, more than a half of the PRs occur in monosyllabic words. Excluding monosyllabic words, the PR rate is high in the word-final position, especially in the disyllabic word-final position.

It seems that the PR words with more than two syllables are a little more in content words than in function words. In addition, monosyllabic words show a higher rate in function words. The reason for these results is that function words are mostly monosyllabic in Mandarin.

3.3.2. Word classes and PR position in the phrase

Table 6 shows results of the distribution of PR function and content words in relation to their position within the phrase.

Table 6: Word classes and PR position in the phrase.

	Initial	Medial	Final
PR (#)	180	109	300
PR (%)	3.99%	0.57%	95.44%
PR content words (%)	14.50%	7.51%	24.55%
PR function words (%)	8.40%	6.36%	13.61%

In Table 6, the PR rates of content words in all three phrasal positions are higher than those of function words, especially, the phrase-final position. Given the results of Table 5 and 6, it is more likely to find PR content words in final position despite the fact that generally function words are more frequently produced than content words.

3.3.3. Word classes and PR position in the utterance

Table 7 shows results of the distribution of PR function and content words in relation to their position within the utterance.

Table 7: Word classes and PR position in the utterance.

	Initial	Medial	Final
PR (#)	89	599	98
PR (%)	11.32%	76.21%	12.47%
PR content words (%)	1.15%	41.22%	7.25%
PR function words (%)	10.18%	34.99%	5.22%

Different from the high PR rates in final position in the word and phrase, PRs in the utterance tends to occur in medial position. As reported for Mandarin [5], a majority of PRs occur when hesitating, and repetitions are usually produced in the case of restarts or repetitions in utterance-medial positions. Thus, the PR rate is high in both content and function words.

3.4. Part of speech of PRs

Furthermore, we investigate whether the part of speech (POS) of the PR words plays a role in the distribution of PRs. The automatic word segmentation and syntactic tagging system developed for modern Mandarin by the Chinese Knowledge Information Processing Group (CKIP) of Academia Sinica is used to process the data [7]. The POS group ‘‘T’’ used in Table 8 is different from that in the CKIP system. It includes only question particles and expletives used for composing compounds with adverbs, adjectives and nouns, but it excludes interjections, responding and mood expressions.

Table 8: POS of the PR words (A=adjective, ADV=adverb, PA=predicate adjective, Vi=intransitive verb, Vt=transitive verb, N=noun, PROUN=pronouns, DET=determiner, M=measure, C=conjunction, P=preposition, POST=postposition T=expletive and question particle, ASP=aspectual adverb).

Word classes	POS	Total
PR content words (%)	A	0.00%
	ADV	14.38%
	PA	3.05%
	Vi	1.91%
	Vt	18.96%
PR function words (%)	N	11.32%
	PROUN	5.47%
	DET	3.69%
	M	3.56%
	C	5.22%
	P	3.82%
	POST	0.51%
	T	2.67%
	ASP	0.38%
	Particle	16.67%
	Marker	8.40%

In Table 8, transitive verbs, adverbs, nouns and particles show particularly high rates among the POS of PRs, while intransitive verbs and aspectual adverbs are really rare. Especially, there is no prolonged adjective. It is also found that the four frequent PR POS groups are more likely to be prolonged in the following contexts: transitive verbs which are disyllabic compound words or are modified by adverbs,

particles used as fillers, adverbs followed by predicate adjectives or subjects, and disyllabic compound nouns.

3.5. Segment types

In addition to the lexical and syntactic analyses of PR words, we furthermore analyze the segment and tone types of the PR syllables. The result is summarized in Table 9 in three groups: prolonged vowels, prolonged vowel-coda pairs, prolonged onset and prolonged coda. Clearly, vowels and vowel-coda pairs are more frequently prolonged than consonants for all four speakers.

Table 9: Types of prolonged segments.

	s-1	s-2	s-3	s-4
Vowel & Vowel-coda	93/5,141 1.81%	182/9,795 1.86%	196/9,076 2.16%	232/6,654 3.49%
Onset Consonant	2/4,514 0.04%	1/8,164 0.01%	4/7,384 0.05%	11/5,427 0.20%
Coda Consonant	4/1,415 0.28%	20/2,642 0.76%	18/2,090 0.86%	23/1,644 1.40%

The most frequently prolonged segments are listed in Table 10.

Table 10: Most frequently prolonged segments. The ratio is the number of each prolonged segment divided by the number of the segment produced in the corpus.

s-1	s-2	s-3	s-4
[ɔ]	[ɔ]	[ɔ]	[on]
16/62 25.81%	7/53 13.21%	27/61 44.26%	1/2 50%
[on]	[y]	[o]	[ɔ]
2/31 11.76%	10/132 7.58%	15/190 7.89%	20/57 35.09%
[y]	[o]	[a]	[o]
4/79 5.06%	8/183 4.37%	35/645 5.43%	19/129 14.73%
[ɔ]	[l]	[əŋ]	[au]
27/565 4.78%	31/728 4.26%	10/220 4.55%	22/199 12.43%
[l]	[iɛ]	[iɛ]	[an]
9/357 2.25%	6/162 3.70%	6/155 3.87%	14/203 6.90%

The most frequently prolonged segment is [ɔ] which often appears in particles. The fact that the adverb ‘ran2 hou4’ (then) and the transitive verb ‘you3’ (have) are frequently prolonged leads to the result that [o] is also frequently prolonged. Two vowel-coda pairs [on] and [əŋ] often occur in particles, whereas the vowel [l] often appears in transitive verbs and conjunctions. Interestingly, the most frequently prolonged segments in Swedish are mainly consonants [2] and the only vowel [o] found prolonged appears in the preposition. In Mandarin, segmental changes are often found in vowels rather than in consonants.

3.6. Tone types

Table 11 shows the rate of PR relative to Mandarin tones.

Table 11: PR rates in tones.

	Tone 1	Tone 2	Tone 3	Tone 4	Tone 5	Total
total (#)	911	781	930	1,864	303	4,789
s-1 PR (#)	8	3	11	28	3	99
PR (%)	0.88%	0.38%	1.18%	1.50%	0.99%	2.07%
total (#)	1,623	1,626	1,903	3,531	579	9,262
s-2 PR (#)	29	15	21	87	15	203
PR (%)	1.79%	0.92%	1.10%	2.46%	2.59%	2.19%
total (#)	1,376	1,551	1,877	3,106	612	8,522
s-3 PR (#)	38	20	33	56	17	218
PR (%)	2.76%	1.29%	1.76%	1.80%	2.78%	2.56%
total (#)	1,039	1,124	1,279	2,267	493	6,202
s-4 PR (#)	47	26	22	92	18	266
PR (%)	4.52%	2.31%	1.72%	4.06%	3.65%	4.29%
Total	4,949	5,082	5,989	10,768	1,987	28,775
	122	64	87	263	53	786
	2.47%	1.26%	1.45%	2.44%	2.67%	2.73%

Tone 4 is the most frequently produced tone type in the overall data, but not the most frequently prolonged tone type. In Table 11, we find that the speaker s-1 prolongs syllables much less frequently than the other three speakers and speaker s-4 produces more PRs than the other three speakers for all five tone types. This result reflects individual differences in producing prolongation. Also, we find that tone 1, tone 4 and tone 5 are more likely to be prolonged than tone 2 and tone 3.

4. Discussion

In this section, we will discuss the relationship between the function and the POS of PRs. The distribution of POS of PRs divided into groups of PR functions is summarized in Table 12.

Table 12: The PR function versus part of speech.

		Emphasis	Hesitation	Response
PR content words (%)	A	0.00%	0.00%	0.00%
	ADV	3.69%	10.69%	0.00%
	PA	2.42%	0.51%	0.13%
	Vi	1.02%	0.89%	0.00%
	Vt	8.02%	10.94%	0.00%
	N	5.73%	5.60%	0.00%
PR function words (%)	PROUN	1.91%	3.56%	0.00%
	DET	1.27%	2.42%	0.00%
	M	1.15%	2.42%	0.00%
	C	0.76%	4.45%	0.00%
	P	1.27%	2.54%	0.00%
	POST	0.13%	0.38%	0.00%
	T	0.51%	2.16%	0.00%
	ASP	0.13%	0.25%	0.00%
	Particle	5.98%	7.51%	3.18%
	Marker	0.38%	8.02%	0.00%
Total	34.48%	62.34%	3.18%	

As shown in Table 12, most words are prolonged for hesitation approaching 62.34%, but the PR occurrences for expressing an emphasis also make up a high percentage (34.48%). PRs for response mostly occur in particles. Among function words, PRs used for hesitation dominate the others for all POS. In content words, however, there are some exceptions: predicate adjectives and intransitive verbs are

prolonged more often for emphasizing than hesitating. This may be associated with the word order. Predicate adjectives and intransitive verbs are modified and preceded by adverbs. Adverbs show a high PR rate in hesitation, therefore the following predicate adjectives and intransitive verbs are produced relatively rare for hesitation accordingly.

Among transitive verbs, the copular “shi4” is the most prolonged word, and almost only prolonged for hesitation. The transitive verb “suo1” (speak) is sometimes used as a conjunction-like particle attached to an adverb or a preposition. Thus, it is frequently prolonged. In addition, the PR function words like adverbs and conjunctions present an inclination, because a number of repetitions and restarts are often made in adverbs and conjunctions and they are associated with hesitation.

5. Conclusion

This paper presents detailed analyses of prolongations in spontaneous Mandarin. Especially, we show that the function of prolongations in Mandarin is more than to hesitate. To emphasize something or to express a strong feedback are two functions we found in our data. Compared to prolongations in other languages, different preferences are found with regard to the lexical, syntactic and segmental features.

6. Acknowledgements

The work presented in this paper is partially supported by the Ministry of Education of Taiwan (grant 91-E-FA06-4-4) and National Science Council (grant NSC-92-2411-H-001-075).

7. References

- [1] Eklund, Robert, "Crosslinguistic disfluency modeling: A comparative analysis of Swedish and Tok Pisin human-human ATIS dialogues", *Proc. ICSLP'00*, Beijing, vol. 2, pp. 991-994, 2000.
- [2] Eklund, Robert, "Prolongation: A dark horse in the disfluency stable", *Proc. ISCA tutorial and research workshop on Disfluency in Spontaneous Speech*, Edinburgh, UK, pp5-8, 2001.
- [3] Eklund, Robert & Elizabeth Shriberg, "Crosslinguistic disfluency modeling: A comparative analysis of Swedish and American English human-human and human-machine dialogues", *Proc. ICSLP'98*, Sydney, pp. 2631-2634, 1998.
- [4] Den, Yasuharu, "Some strategies in prolonging speech segments in spontaneous Japanese", *Proc. Disfluency in Spontaneous Speech Workshop*, Göteborg University, Sweden, pp. 87-90, 2003.
- [5] Tseng, Shu-Chuan, "Taxonomy of spontaneous speech phenomena in Mandarin conversation", *Proc. of ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, Tokyo, pp. 23-26.
- [6] Tseng, Shu-Chuan and Liu, Yi-Fen, *Annotation of Mandarin Conversational Dialogue Corpus*, Tech. Rep. no.02-01, Chinese Knowledge Information Processing Group, Academia Sinica, 2002.
- [7] Chen, Keh-Jiann, Huang, Chu-Ren, Chang, L.-P. and Hsu, H.-L., "ACADEMIA SINICA BALANCED CORPUS: design methodology for balanced corpora", *Proceedings of the Eleventh Pacific Asia Conference on Language, Information and Computation*, pp. 167-176.