A preliminary study on the prosody of broadcast news in Hong Kong Cantonese

Peggy Pik Ki Mok, Holly Sze Ho Fung, Jingwen Li

Department of Linguistics and Modern Languages, The Chinese University of Hong Kong

`peggymok@cuhk.edu.hk, hollyfung@cuhk.edu.hk, joanneljw@gmail.com`

Abstract

Broadcast news is a distinctive register. Previous studies only provided some general descriptions of the prosodic features in broadcast news but with few concrete data. Most of them were also on English news. This study investigated the prosodic features of Cantonese TV broadcast news using acoustic data. Speech using the same materials from two groups was compared: eight Hong Kong professional TV news anchors, and a control group consisting of eight university students. The results show clear differences between the two groups in terms of speech rate, pitch range and variability of syllable duration (speech rhythm). It was found that the news anchors spoke significantly faster than the control group, also with an enlarged pitch range. They also produced more variability in syllable duration. There is clearly more prosodic variation in the news register than ordinary speech. Finally, we provide some possible reasons for these features, as well as directions for future studies.

Index Terms: news, broadcast, prosody, Cantonese

1. Introduction

This study investigates the prosody of Cantonese news broadcast in television read by professional news announcers and compares it to that read by non-professional speakers. Broadcast news has a distinctive style/register in different languages, but not many studies have examined its specific prosodic features. Most previous studies were on English news, while so far no study has examined the prosody of Cantonese broadcast news. This preliminary aims to fill this research gap.

The broadcast news register is easily recognizable even by lay people. Before the 1970s, news readers spoke in a very formal and solemn way, delivering news with much authority. Women were excluded from news broadcast due to their voices being too high-pitched or lack in authority [1]. After the 1970s, with technological advancements in voice capturing techniques and various social changes (e.g., the introduction of television in common households), the style of broadcast news has changed dramatically. Summarizing some previous descriptions on vocal characteristics of newsreaders, [2] suggested that an effective newsreading voice is characterized by a rich, warm, and resonant tone. News readers need to deliver the news clearly and succinctly, and with enthusiasm to capture the audience's attention, while at the same time sounding professionally distant and knowledgeable.

Cotter [3] proposed three factors which affect news prosody: text structure (condensed text written to convey newsworthiness and capture attention); an unseen audience and medium constraints (using voice only to signal various boundaries and emphasis). These factors contribute to the unique prosody of broadcast news register, which, according to [4], is between reading and spontaneous speech, having all the features of reading and some features of spontaneous speech.

Despite the easy identification of the news register, and the literature on general perception and opinions of good practice in newsreading (see summary in [2, 5]), surprisingly few studies have examined the prosodic features of broadcast news in detail. Bolinger [6] noticed that some American radio newscasters intentionally distorted expected sentence stress and put accentual emphasis on words whether semantically justified or not in order to attract attention. Cotter [3] compared news read by radio announcers and that by volunteers, and found that news prosody makes use of speech rate, pauses, pitch movement and other paralinguistic features in an identifiable way, incorporating features of spontaneous and public discourse. [2] compared professional newsreaders with student newsreaders and controls, and found that female newsreaders had a higher pitch than both students and controls; professionals also had greater pitch variability, spoke faster and made fewer errors than both students and controls. The professional newsreaders were also rated significantly higher by the judges on phrasing and overall performance. Moreover, more students reported consciously altering their voices and using more effort in newsreading than the professionals did.

The above studies are all based on broadcast news in English. Zou [7, 8] examined the prosody of broadcast news in Mandarin Chinese. He found that the mean syllable duration is longer in news announcements than in conversation spoken by the same presenters, i.e., the presenters spoke more slowly and clearly in news announcements, compared with their own conversation. Moreover, there is a wider pitch range variation in news announcements than in conversation. His data show that a distinct news register is also present in Chinese. However, it is unclear if these prosodic features in Mandarin news are also applicable to broadcast news in Cantonese, as broadcasters in China are specially trained for the profession with very strict standards on their speech characteristics, while it is not the case in Hong Kong. Also, news prosody can vary between languages [1, 9] and also between different ages of the same language [9, 10]. Therefore, more studies on the prosody of broadcast news in different languages are needed to thoroughly investigate the various prosodic features of this distinct speech style.

2. Method

The present study compares the news announced by professional journalists (anchors) in television news broadcast with those read by non-professional speakers (controls). The controls read exactly the same materials as the anchors for direct comparison.

2.1. Speakers and materials

Eight news anchors (four male, four female) from a mainstream local television broadcasting company in Hong Kong (TVB) were chosen. It is unknown if the company has any special voice/speech training for their news anchors, but the company is known for its insistence on a high standard of pronunciation, e.g., their anchors preserve some contrasts that

are already lost or used interchangeably by most Hong Kong people. Approximately one-minute of news announced by each anchor based on four different stories each was chosen from the 'News at Six-Thirty' programme aired between 23 March and 23 July 2013, videos of which could be found in YouTube. Criteria for choosing the anchors include their popularity and the availability of videos in YouTube. Table 1 summarizes the information about the speech samples of the eight anchors.

Table 1. *Summary of news materials by the anchors.*

Female				Male			
Anchors	# σ	# IP	# S	Anchors	# σ	# IP	# S
Cheng	246	20	5	Fong	278	16	5
Chow	246	19	5	Lau	262	22	6
Lau	266	15	6	Ng	279	19	5
Law	277	22	5	Pun	303	24	5

σ = syllables; IP = intonational phrases; S = sentences

The control group consists of eight local university students, also four male and four female, all native speakers of Hong Kong Cantonese. They were aged between 18 to 27, and none of them reported any speech or hearing impairment. Two speakers received course credits and another speaker was paid for their participation. The other five speakers participated on a voluntary basis.

2.2. Procedure

A pilot trial with one speaker (not in the control group) indicated that it would be very difficult and very slow for the control speakers to read the news materials fluently on the spot. In order to solve this problem, the control speakers were given the news materials a couple of days in advance and were asked to practise them well before the actual recording. Therefore, the control speakers knew the purpose of the experiment. This is unavoidable given the logistic constraints. This also means that the control speakers were performing at their best as 'mock anchors', based on their own perception of the news register.

The recording took place in a quiet room at the Chinese University of Hong Kong. A solid state recorder was placed approximately 20 cm away from the speakers. The speech materials were recorded with a sampling rate of 44100 Hz and 16 bits. The control speakers were asked to read aloud the stimulus sentences displayed on a computer screen one by one in a random order. If they made a mistake, they were prompted to re-read the sentence again immediately, so the recorded materials contain no speech errors or disfluencies. Two repetitions of the materials were collected.

In order to control for the gender difference, the control speakers only read the news materials produced by the anchors of the same sex.

2.3. Data analysis

As the sentences written for the news scripts were very condensed and very long, the sentences were divided into intonational phrases for analysis (see Table 1). Three types of data were examined to compare the prosodic features between anchors and controls: speech rate, pitch range and speech rhythm. Speech rate was calculated based on the average number of syllables per second. Since there is much individual variation between speakers' habitual pitch levels, pitch range

is a better indicator of any stylistic pitch variation in news prosody than mean pitch is. We used two methods to estimate the pitch range. As Cantonese is a tone language, a good estimation of the pitch range can be obtained by comparing the F0 values at the midpoint of all Tone 1 syllables (a high level tone [55]) with those of Tone 4 (a low falling tone [21]), which respectively represent the highest and lowest pitch levels in a speaker's voice. The pitch level at the end of Tone 4 falls so low that it often results in creakiness. Measuring F0 values at the midpoint can ensure more valid data. In addition to the pitch differences between T1 and T4, we also calculated the pitch ratios between T1 and T4 as a normalized measure for all speakers.

We used the acoustic metrics on syllable duration, VarcoS and nPVIS, to compare speech rhythm. VarcoS shows the standard deviation of syllable duration in an utterance normalized for speech rate [11], while nPVIS shows the normalized pairwise variability of syllable duration [12]. VarcoS measures durational variability globally, while nPVIS measures variability locally between each pair of unit. Mok [13, 14] have shown that these metrics on syllable duration can be more robust than metrics on consonantal and vocalic durations in reflecting rhythmic differences. Moreover, the syllable structure of Cantonese is very simple. Segmenting the speech streams into syllables is quite easy and straight forward. It is noted that the rhythmic metrics have been under much criticism in recent years [15], but it is justified in our case because we compared exactly the same speech materials between anchors and controls. Cantonese is a typical syllable-timed language [13, 14, 16]. It will be interesting to see if more durational variability would be employed by the news anchors to enhance liveliness and to capture audience's attention.

3. Results

3.1. Speech rate

Table 2 shows the average speech rate of the eight anchors and the control speakers. It is obvious that the anchors spoke faster than the controls [t(14) = 7.484, p < 0.0001]. It is worth noting that the controls were already performing at their best as 'mock anchors', and no disfluency was included in their recordings. Nevertheless, they still spoke significantly slower than the anchors.

Table 2. *Average speech rate (# syllable / second).*

Source	Anchors	Controls*
Cheng	6.10	4.66
Chow	5.75	4.76
Lau	5.68	4.53
Law	5.19	4.53
Fong	5.51	4.97
Lau	5.62	4.84
Ng	6.19	4.90
Pun	5.75	4.93

* The control data are averaged across the reading of the same anchor's materials by four control speakers of the same sex.

3.2. Pitch range

Figure 1 shows the pitch differences in Hz of the anchors and those averaged across the four control speakers reading the materials by the same anchor. Each bar represents one

anchor's materials. Female and male data are shown separately. It can be seen that female anchors have a significantly larger pitch difference than female controls do [t(6) = 7.147, p < 0.001]. The pitch range is also generally larger for male anchors than male controls, although the difference is not significant [t(6) = 1.669, p = 0.146]. The larger pitch range indicates that the (female) professional anchors used more pitch variation in their news reading. It is interesting to note that there is also more individual variation in pitch range among the anchors than among the controls.
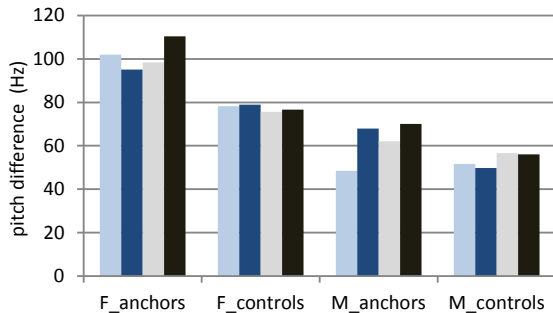

Figure 1: *Pitch difference of anchors and controls.*

In addition to the differences in Hz, we have also calculated the ratios of T1/T4 as a normalized measure of pitch range for both anchors and controls to verify the patterns observed above. Figure 2 shows the pitch ratios of both groups of speakers. Again, anchors have a larger pitch range than controls do, and the difference is significant for the female speakers [t(6) = 5.344, p=0.002], and is tending towards significance for male speakers [t(6) = 2.075, p = 0.083].
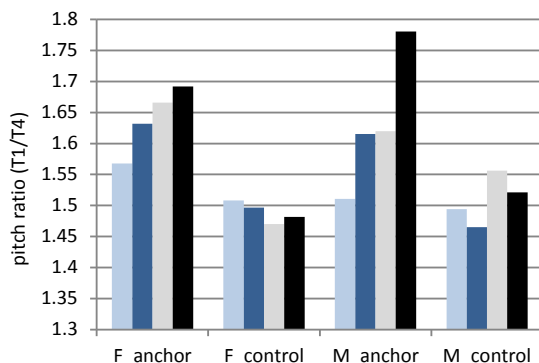

Figure 2: *Pitch ratio of anchors and controls.*

3.3. Rhythmic metrics

Table 3 shows the values of VarcoS and nPVIS for the anchors and those averaged across the four control speakers reading the materials by the same anchors. Anchors have significantly higher variability of syllable duration than controls do for both rhythmic metrics: VarcoS [t(14) = 5.90, p < 0.0001] and nPVIS [t(14) = 4.503, p = 0.001]. Moreover, if we examine the values more closely, those produced by the anchors are always higher than those averaged among the controls, even though they read exactly the same news materials. It is also true if we compare the anchor's data with each individual control speaker.

Table 3. *Rhythmic metrics on syllable duration.*

Source	Anchors		Controls	
	VarcoS	nPVIS	VarcoS	nPVIS
Cheng	28.59	30.17	23.31	25.05
Chow	25.20	28.44	22.89	25.77
Lau	28.11	27.95	24.22	25.84
Law	29.24	32.05	24.97	27.90
Fong	30.62	33.61	23.73	26.31
Lau	27.62	31.42	25.72	30.02
Ng	29.68	30.83	25.39	28.10
Pun	28.42	31.01	25.18	28.06

4. Discussion

The study has found clear differences in the prosodic features between news read by professional anchors and by control speakers. Some differences can be expected, e.g., speech rate, while others are rather interesting, e.g., speech rhythm.

The best versions produced by the control speakers are still significantly slower than those produced by the anchors. In order to speak clearly as 'mock anchors', the controls produced the materials carefully and thus had slowed down the speech rate, as what most people would do. However, the anchors could speak clearly and quickly, as they need to deliver as much information as possible in a short period of time. The demands on time necessitate a faster speech rate. So in addition to the three factors (text structure, unseen audience and medium constraints) proposed by Cotter [3] which influences news prosody, time pressure is another important factor contributing to the unique prosody of broadcast news.

In addition to saying things more slowly, listening to the recordings tells us that another reason why the controls were slower is that they made more pauses and/or paused longer than the anchors.

Furthermore, given that news anchors can speak clearly and quickly, it will be interesting to examine their speech articulatorily to see if there are differences between their articulation in news reading and that of control, and also between the news speech and the 'normal' speech by the same anchors. Do anchors simply move their articulators faster or do they have different articulatory strategies in news reading? For example, would there be more gestural overlaps in their news reading? A faster speech rate usually results in more gestural overlaps [17], but the demand for clarity would result in less overlaps [18]. Would the anchors increase the magnitude of their gestures in order to accommodate these contradicting demands? Interesting patterns are likely to be found.

Professional news readers need to capture and sustain the attention of the audience, and to highlight different important information in the scripts that are already very condensed. A useful strategy is to increase variability in speech. Our data show that the anchors did so in both pitch and duration.

The pitch ranges produced by the female anchors were significantly larger than those by the female controls. Thus, the 'tone space' is larger for the anchors, and their lexical tones would be more distinct than those by the controls. In addition to having more distinct tones, more fluctuating intonation patterns would also result in a larger pitch range. It is quite likely that both aspects have jointly contributed to their larger pitch range. Further studies should devise ways to evaluate the contribution of lexical tone and intonation

separately.

It is interesting to notice that although the male anchors had a similar pattern of generally having a larger pitch range than the male controls did, the difference is not significant. The pitch range for male speakers is naturally smaller than that of female speakers, so the increase in pitch variability would probably be proportionally smaller as well, although Figure 2 shows one male anchor having a higher ratio than the female anchors. Probably, the more subtle increase is harder to show up statistically.

A further reason for the insignificant result is that one male control speaker had a pitch range larger than other male speakers, including three male anchors. The authors were familiar with this control speaker. When listening to his recordings, it is obvious that he had prepared the materials well and spoke with more pitch variation in the news reading than in his ordinary speech. He performed well as 'mock anchors'. This shows us that even lay speakers have a hunch of the pitch patterns and variation in the news register, which are confirmed statistically by our female data.

Last, but certainly not the least, our results indicate that the anchors had also increased the variability of syllable duration in their news reading. This finding is particularly interesting given that Cantonese is a very typical syllable-timed language, even more so than French and Italian [19]. One may expect that variability of syllable duration would be the least likely aspect to be manipulated. Our significant findings confirm that even such unlikely aspect is used in news prosody. All the anchors consistently had higher variability than the controls as a group had, and also consistently higher variability than each individual control speaker.

When we listened to the recordings, it is obvious that the anchors were skillful in compressing some common or familiar phrases, while speaking more clearly and slowly for other important information. For example, the title and the name of the President of China, Xi Jinping (國家主席習近平) produced by one female anchor had the following characteristics: she had parsed the whole phrase into two parts (the title with four syllables and the name with three syllables). The duration for the first part with four syllables is 531ms, while it is 552ms for the second part with only three syllables. In addition to such strategic compression and extension, since the speech rate is so fast, the anchors may be actively using syllable duration for phrasing and creating boundaries. Also, they may produce strong focus when important terms come up, resulting in longer syllable durations in certain phrases. All this explains very well why the rhythmic metrics for syllable duration are significantly higher for all anchors than for the controls. Much skill and experience in information selection is needed in order to manipulate this prosodic feature well. Further studies on Cantonese news prosody should especially examine this interesting feature.

Our study is a preliminary study on the news prosody in Hong Kong Cantonese. It provides concrete acoustic data as evidence, supplementing the many general descriptions of features of news prosody in the literature. Nevertheless, one shortcoming is that the data, although well controlled for comparison, is quite limited. Further studies using a larger data set are needed to corroborate the findings in this study.

5. Conclusions

A distinct register of broadcast news is found in Cantonese. Professional news readers speak significantly faster than control speakers do. They also have more stylistic variations in both pitch and syllable duration. It will be interesting to compare the news register in different languages in the future to confirm which prosodic features are common for news reading in general, and which features are language-specific devices to cater for this special speech style.

6. References

[1]   Price, J. (2008). New News Old News: A Sociophonetic Study of Spoken Australian English in News Broadcast Speech. *Arbeiten aus Anglistik und Amerikanistik,* 33(2), 285-309.

[2]   Neil, E., Worral, L., Day, A., & Hickson, L. (2003). Voice and speech characteristics and vocal hygiene in novice and professional broadcast journalists. *Advances in Speech-Language Pathology,* 5(1), 1-14.

[3]   Cotter, C. (1993). Prosodic Aspects of Broadcast News Register. In *Proceedings of the Nineteenth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Semantic Typology and Semantic Universals* (pp. 90-100). Berkeley.

[4]   Levin, H., Schaffer, C., & Snow, C. (1982). The prosodic and paralinguistic features of reading and telling stories. *Language and Speech,* 25(1), 43-54.

[5]   Warhurst, S., McCabe, P., & Madill, C. (2013). What Makes a Good Voice for Radio: Perceptions of Radio Employers and Educators. *Journal of Voice,* 27(2), 217-224.

[6]   Bolinger, D. (1982). The network tone of voice. *Journal of Broadcasting,* 26, 726-728.

[7]   Zuo, Y., Li, X., & Hou, M. (2006). Comparison of News Announcing and Talking Styles in Broadcast Speech, *International Symposium on Chinese Spoken Language Processing (ISCSLP 2006).* Singapore.

[8]   Zou, Y. (2007). A Formal Study on Prosody of Presenter's Spoken Language Based on Broadcast Speech Corpus (in Chinese). PhD thesis, Communication University of China.

[9]   Boula de Mareüil, P., Rilliard, A., & Allauzen, A. (2012). A Diachronic Study of Initial Stress and other Prosodic Features in the French News Announcer Style: Corpus-based Measurements and Perceptual Experiments. *Language and Speech,* 55(2), 263-293.

[10] Zou, Y., Wang, Y., & He, W. (2012). Diachronic contrastive analysis on read speech in broadcast news: evidence from pitch and duration, *The International Symposium on Chines Language Processing (ISCSLP 2012)* (pp. 291-295). Hong Kong.

[11] Dellwo, V. (2006). Rhythm and speech rate: a variation coefficient for ΔC. In P. Karnowski & I. Szigeti (Eds.), *Language and Language-Processing* (pp. 231-241). Frankfurt am Main: Peter Lang.

[12] Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 515-546). Berlin: Mouton de Gruyter.

[13] Mok, P. (2011) The acquisition of speech rhythm by three-year-old bilingual and monolingual children: Cantonese and English. *Bilingualism: Language and Cognition*, 14: 458-472.

[14] Mok, P. (2013) Speech rhythm of monolingual and bilingual children at 2;06: Cantonese and English. *Bilingualism: Language and Cognition,* 16: 693-703.

[15] Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica,* 66(1-2), 46-63.

[16] Mok, P. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics,* 2, 148-154.

[17] Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics,* 24(2), 263-282.

[18] Tasko, S. M., & Greilick, K. (2010). Acoustic and Articulatory Features of Diphthong Production: A Speech Clarity Study. *Journal of Speech, Langauge and Hearing Research,* 53, 84-99.

[19] Mok, P. & Dellwo, V. (2008) Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. In *Proceedings of the 4th Speech Prosody (2008).* 423-426. Campinas, Brazil.