DOES LEXICAL TONE RESTRICT THE PARALINGUISTIC USE OF PITCH? COMPARING MELODY METRICS FOR ENGLISH, FRENCH, MANDARIN AND CANTONESE.

Daniel HIRST[1,2], *John WAKEFIELD*[3], *H.T. Yoko LI*[3]

[1]Laboratoire Parole et Langage, CNRS & Aix-Marseille University, Aix-en-Provence;
[2]School of Foreign Languages, Tongji University, Shanghai;
[3]Department of English Language and Literature, Hong Kong Baptist University, Hong Kong;

daniel_hirst@lpl-aix.fr; johncw@hkbu.edu.hk; yokoartemis@gmail.com

ABSTRACT

It has long been suspected that the use of pitch for lexical distinctions as in a lexical tone language may restrict the extent to which pitch can be used for other, paralinguistic functions. A recent study found that melody metrics calculated for a comparable corpus of read speech in English and French revealed a significant gender difference which was not found for comparable recordings in Mandarin Chinese. This study applies the same metrics to a comparable corpus of Cantonese. The results confirm our expectation that the gender differences observed in English and French but not in Mandarin Chinese are not observed in Cantonese either.

Keywords: melody metrics, gender, tone, Cantonese, Mandarin Chinese.

1. INTRODUCTION

It is well known that speech prosody in general, and speech melody in particular, are used in speech to carry out a number of different linguistic and paralinguistic functions (e.g. [22, 23]).

It has long been suspected that the pressure of a lexical tone system may restrict the degree to which pitch may be used for paralinguistic functions ([20, 5, 10]). Cruttenden [5] notes:

> *Languages where pitch patterns are produced primarily by lexical tones will have a limited potential for intonational variation.* [p12]

One of the paralinguistic functions of pitch is the expression and marking of gender. The correlation of pitch height with the speaker's sex has an obvious biological basis due to the fact that the male vocal folds are generally longer than the female vocal folds [21]. It is, however, well known that the biological basis only explains part of this effect and that the expression and marking of gender is largely conditioned by sociolinguistic factors which may be different in different societies [6].

2. TONE AND PARALINGUISTIC FUNCTIONS

A recent study [14] measured a number of different melody metrics for two European languages (English and French) and one Asian tone language (Mandarin Chinese). For nearly all the metrics measured (see §5 below for details), there was a significant difference between the measurements for the Mandarin data and those for the European languages.

The most significant effect was that the pitch movements characterised by the different metrics were more ample and more varied in Mandarin than those observed for English and French. There was also a slightly less significant difference between the pitch movements in the French data and those in the English data, the pitch movements in French being slightly more ample and varied than those in English.

In addition to these simple effects of language, there was also a very significant interaction between the language and the gender of the speakers. For the English and French data, the pitch movements were generally more ample and more varied for the female speakers than for the male speakers. For the Mandarin data, the pitch movements were slightly more ample and more varied for the male speakers than for the female speakers but the difference was not significant.

These results were interpreted as a possible effect of the lexical tone system of Mandarin which may have the effect of restricting the paralinguistic use of pitch for the expression of gender differences.

3. CANTONESE.

If it is, indeed, the lexical tone system of Mandarin Chinese which is responsible for the lack of a gender difference in the data analysed in [14], then we might expect to find a similar or even more marked effect from an analysis of comparable data for a language with a more complex tonal system, such as Cantonese.

Cantonese has six lexical tones [1, 25, 18]. Written according to the numbers as used in Jyutping, the tones are: 1 high level; 2 high rising; 3 mid level; 4 low falling; 5 low rising; 6 low level. Because of its complex lexical tone system, which involves both height and orientation, Cantonese severely restricts the speaker's ability to manipulate pitch for anything other than the expression of lexical tone [4, 25].

Cantonese has the apparently universal characteristics of prosody, such as downdrift [4, 9], the manipulation of pitch range or level [4, 19], and the marking of syntactic phrasing [11]. It also allows the use of rising tones to form declarative clause questions [11, 18]. However, Cantonese is restricted in its use of intonation to express a wide range of discourse related meanings, as found in languages such as English. This is probably why it has more than 30 sentence-final particles (SFPs), which are used to express these types of discourse meanings. Yau [24] suggested that

> *there is a mutual compensation between [SFPs] and intonation patterns and that the more a language relies on the use of [SFPs] in expressing sentential connotations, the less significant will be the role played by intonation patterns, and vice versa.* [p 51]

He concluded that Cantonese and English represent the two extremes of this continuum, and other linguists have supported this claim by saying that, as far as they know, Cantonese has more SFPs than any other language studied thus far (e.g., [16, 17].

4. RECORDINGS

[14] used the 40 five-sentence continuous passages taken from the Eurom1 corpus, provided as a deliverable of the European SAM (Speech Assessments and Methodology) project [3]. The passages were originally recorded in the 1980's by ten speakers (five male and five female) for each language, but at that time, each speaker read only a limited number of the 40 passages (15 per speaker for English and 10 for French). In addition, the original recordings of the corpus were under copyright, owned by the different laboratories and universities which had produced the recordings, which restricted the distribution of the corpus.

For [14], new recordings of this corpus were made for English (EN), French (FR) and Mandarin (ZH), with all 40 passages read by at least 10 speakers (EN: 5m, 6f; FR: 4m, 7f; ZH: 5m, 5f). The recordings and associated annotations are distributed by the *SLDR (Speech and Language Data Repository : http:sldr.org)* as part of the *OMProDat* database [15], under an open-database license .

For our study of Cantonese (YUE), the 40 five-sentence passages were translated into Cantonese from the English and Chinese versions of the corpus. The text of the passages was adapted to the Hong Kong environment when necessary.

For the recordings, we adopted the same strategy as [14]: all 40 of the 5-sentence passages were read by 10 native speakers of Hong Kong Cantonese. 5 of the speakers were male and 5 were female. They ranged in age from 20 to 22. All of them were either in their final year of undergraduate studies, or had just graduated.

The Cantonese recordings and associated annotations will be added to the *OMProDat* database and distributed by the *SLDR*.

5. MELODY METRICS

In order to eliminate the microprosodic effects of the specific phonemes which make up each utterance (in particular the consonants, cf [8]), the fundamental frequency curves of the recordings were modelled using the Momel algorithm [12, 13], which assumes that a pitch contour can be adequately represented by a sequence of target points, each contiguous pair of which is linked by a continuous smooth monotonic quadratic interpolation (defining a quadratic spline function).

This, in turn, assumes that the shape of a pitch contour is entirely determined by the temporal and frequential values of the relevant target points. The algorithm uses an asymmetric version of robust regression to optimise the modeling of raw fundamental frequency curves with a quadratic spline function.

Figure 1 illustrates the output of the Momel pitch modelling algorithm for the sentence *What can I have for dinner tonight?,* the first sentence of passage R1 of the English version of the Eurom1 corpus. The fundamental frequency curve for this 1.5 second utterance consists of 150 values of f0, with a time step of 0.01 seconds (the green curve). The curve is modelled by the Momel algorithm as a sequence of 7 target points which define the continuous smooth quadratic spline curve shown in red in this figure.

These target points are taken as the primary dependent variable in this study.

In order to reduce the inter-subject variability, in particular the difference between male and female speakers, the target points were scaled using the *OMe* (Octave-Median) scale as proposed in [7] with the formula:

(1) ome = log2(Hz/median)

where *median*, here, is the median value of f_0 for the whole five-sentence passage.

From the scaled target points, the mean and standard deviation for each passage were then calculated for:

- **pitch** value of target points on the OMe scale
- **interval** absolute difference (in octaves) from the previous target
- **rise, fall** difference from previous target for rise and fall separately
- **slope** absolute difference in octaves from previous target divided by distance in seconds
- **rise-slope, fall-slope** slope for rise and fall separately

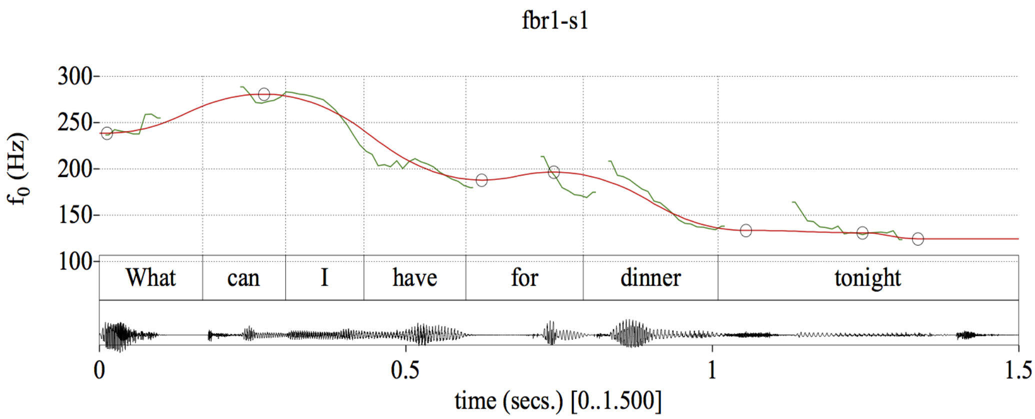
In all, 14 values were calculated for each passage. All the values were offset to the speaker's median f_0 by (1).

6. RESULTS

A linear discriminant analysis was carried out on the metrics for the four languages, English (*EN*), French (*FR*), Cantonese (*YUE*) and Mandarin Chinese (*ZH*).

As can be seen from Table 1 the overall discrimination by language was 73.9% (= (359+231+277+331)/1198).

Figure 1. Fundamental frequency and Momel target points.



Both Mandarin and Cantonese were well distinguished from English and French (88.2% correct = (359+159+52+23+277+5+16+331)/1198) as well as from each other (96.7% correct = (271+331)/(277+331+5+16)).

Table 1. Linear discriminant analysis on the melody metrics for English, French, Cantonese and Mandarin.

	Predicted			
	EN	FR	YUE	ZH
EN	***359***	159	41	4
FR	52	***231***	41	47
YUE	18	7	***277***	5
ZH	8	25	16	***331***

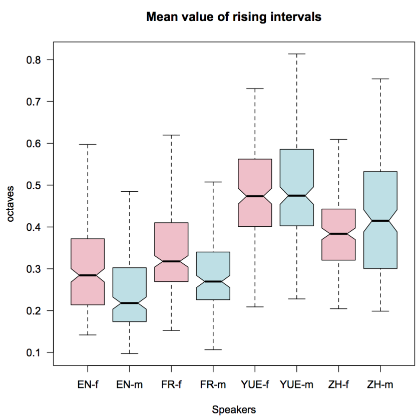
Table (2) shows the significance level of an Anova analysis for each of the 14 parameters for *language* (L) and *gender* (G) as well as for the interaction *language*gender* (L*G). A detailed statistical analysis of these results, in progress, will be reported in future work, but, as a preliminary illustrative example, the box-plot in Figure (2) shows that *pitch rises* are, on average, of greater amplitude for Cantonese than for Mandarin, for Mandarin than for French, and for French than for English.

Table 2. Significance levels of Anova for each parameter and interaction. [-] : n.s.; [*] : p < 0.05 [***] : p < 0.001.

	mean			sd		
	L	G	LG	L	G	LG
pitch	***	***	***	***	***	***
interval	***	-	*	***	***	***
rise	***	***	***	***	***	***
fall	***	***	***	***	***	***
slope	-	-	-	-	-	-
rise-slope	-	-	-	-	-	-
fall-slope	***	***	***	***	***	***

For English and French, as observed by [14], there is a significant gender difference: for both languages the pitch rises of the female speakers are significantly greater than for the male speakers.

Figure 2: Mean value of rising intervals for male and female speakers for the four languages, on an octave scale. The bottom middle and top of each box corresponds to the 1st, 2nd (= median) and 3rd quartiles of the distribution, while the whiskers show the minimum and maximum values.



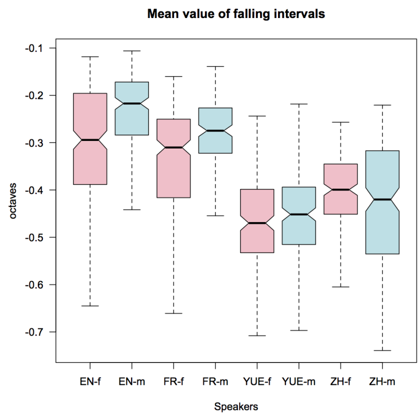
Note that, since all the values are calculated on the octave-median scale, the greater amplitude of the pitch rises for female speakers is not simply an effect of higher pitch.

For Mandarin, the gender difference is reversed: the male Chinese speakers produced slightly larger pitch rises than the female speakers but the difference is probably not significant as can be seen from the fact that the 'notches' on the boxes overlap for these two.

For Cantonese, as can be seen, there seems to be no significant difference between the amplitude of the pitch rises for male and for female speakers.

A similar pattern to that for rising pitch movements is also found for falling pitch movements, as shown in Figure 3.

Figure 3: Mean value of falling intervals for male and female speakers of the four languages using the Octave-Median scale. See Figure 1 for details.



The (absolute) mean amplitude of the falling pitch movements follows the same order as observed for the rises: Cantonese > Mandarin > French > English. Once again, there is a signicant gender difference for English and French, with the female speakers making significantly more ample falling pitch movements (on an octave scale) than the male speakers.

For Mandarin, just as for the rising intervals, the order is reversed, the male speakers making larger pitch falls than the female speakers; but, once again, the overlapping notches of the boxes are an indication that this difference is probably not significant. This time, for the Cantonese speakers, the falling intervals for the female speakers are slightly greater than those for the male speakers but, once again, the notches of the boxes are largely overlapping, which normally indicates that the difference is not significant.

We have illustrated the results of the Anova analyses with the mean values of rising and falling intervals but it is worth noting that, for *all* the parameters and interactions which appear as significant in Table 1, there is a similar pattern with the same ordering: *Cantonese* > *Mandarin* > *French* > *English* and with a gender difference *female* > *male* which appears to be significant for English and French but not for Mandarin or Cantonese.

7. CONCLUSIONS

In [14], it was suggested that the differences between Mandarin Chinese on the one hand and French and English on the other hand, in particular the lack of a gender effect in Mandarin, might be due to the pressure of lexical tone restricting the paralinguistic use of pitch for non-lexical sociolinguistic effects such as gender marking.

In this paper, we carried out a similar test using data from Cantonese, a language with an even richer lexical tone system than Mandarin. The results confirmed our expectations that the paralinguistic gender effects, which were observed in English and French but not in Mandarin, were not observed in Cantonese either.

There are, naturally, far too few subjects to allow us to draw firm conclusions from these results, but the fact that all the parameters measured showed the same pattern could be taken to indicate that the results are fairly robust.

There are also, of course, other possible explanations for our results. We have compared two Indo-European languages from Europe with two Sinitic languages from Asia. It is possible that the differences we observed are not caused by the lexical tone at all but by characteristics specific to the language families to which these languages belong or by cultural differences between European and Asian civilisations.

In order to test these possibilities, we intend, in future work, to extend the paradigm adopted here to include non-tonal Asian languages (such as Japanese and Korean), non Sinitic tone languages (such as Vietnamese and Yoruba) as well as non-Indo-European languages from Europe (such as Finnish and Basque).

8. REFERENCES

[1] Bauer, R. S., Benedict, P. K. 1997. *Modern Cantonese phonology*. Berlin: Mouton de Gruyter.

[2] Bauer, R. S., Cheung, K., Cheung, P., Ng, L. 2004. Acoustic correlates of focus-stress in Hong Kong Cantonese. In S. Burusphat (Ed.), *Papers from the Eleventh Annual Meeting of the Souteast Asian Linguistics Society 2001*: Tempe, Arizona: Arizona State University: Program for Southeast Asian Studies Monograph Series Press: 29–49.

[3] Chan,D., Fourcin,A., Gibbon,D., Granstrom,B., Huckvale,M., Kokkinakis, G., Kvale, K., Lamel, L., Lindberg, B., Moreno, A., Mouropoulos, J., Senia, F., Trancoso, I., Veld, C., Zeiliger, J. 1995. Eurom - a spoken language resource for the EU. In *Eurospeech95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology.*, 1, 867870, Madrid., September 1995 18-21.

[4] Cheung, K.-H. 1986. *The Phonology of Present Day Cantonese* (Unpublished doctoral dissertation). University College, London.

[5] Cruttenden, A. 1986. *Intonation*. Cambridge: Cambridge University Press.

[6] Daly, Nicola & Paul Warren. 2001. Pitching it differently in New Zealand English: speaker sex and intonation patterns. *Journal of Sociolinguistics*, 5(1), 85-96.

[7] De Looze, C., Hirst, D.J. 2010. L'echelle OME (Octave-MEdiane): une echelle naturelle pour la melodie de la parole. in *Actes des XXVIIIes Journes d'Etude sur la Parole*, Mons, Belgium, May 25-28 2010.

[8] Di Cristo, Hirst,D.J. 1986. Modelling French micromelody: analysis and synthesis. *Phonetica*, 43(1- 3), 11–30.

[9] Flynn, C.-Y.-C. 2003. *Intonation in Cantonese*. Muenchen: Lincom Europa.

[10] Fox, A. 2000. *Prosodic Features and Prosodic Structure. The Phonology of Suprasegmentals*. Oxford: Oxford University Press.

[11] Fox, A., Luke, K.-K., & Nancarrow, O. 2008. Aspects of intonation in Cantonese. *Journal of Chinese Linguistics*, *36*(2), 321–367.

[12] Hirst, D.J. 2007. A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation, In *Proceedings of the XVIth International Conference of Phonetic Sciences*: 1233-1236, Saarbrucken, 2007.

[13] Hirst, D.J. 2011. The analysis by synthesis of speech melody: from data to models, *Journal of Speech Sciences*, 1(1): 55-83.

[14] Hirst, D.J. 2013. Melody metrics for prosodic typology: comparing English, French and Chinese. in *Interspeech 2013: Proceedings of the 14th Annual Conference of the International Speech Communication Association,* Lyon, August 2013. 572-576.

[15] Hirst,D.J., Bigi,B., Cho,H., Ding,H., Herment,S., Wang,T. 2013. Building OMProDat, an open multilingual prosodic database. *Proceedings of TRASP, Tools and Resources for the Analysis of Speech Prosody*, satellite workshop of *Interspeech* 2013, Aix-en- Provence, August 30, 2013. 11-14.

[16] Leung, C. 2005. 當代香港粵語語助詞的研究 *[A Study of the Utterance Particles in Cantonese as Spoken in Hong Kong]*. Hong Kong: Language Information Sciences Research Centre, City University of Hong Kong.

[17] Luke, K. K. 1990. *Utterance Particles in Cantonese Conversation*. Amsterdam: John Benjamins.

[18] Ma, J. K.-Y., Ciocca, V., Whitehill, T. L. 2011. The perception of intonation questions and statements in Cantonese. *The Journal of the Acoustical Society of America*, *129*(2), 1012–1023.

[19] Pennington, M. C., & Ellis, N. C. 2000. Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal*, *84*(3), 372–389.

[20] Pike, K.L. 1946. *Tone Languages. A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion*. Ann Arbor: The University of Michigan Press.

[21] Titze, I. 1989. Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America* 85: 1699-1707.

[22] Wang, S.-Y. 1969. The many uses of f0. in A.Valdman (ed.) *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre,* The Hague: Mouton, 487-503.

[23] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46: 220-251.

[24] Yau, S. 1980. Sentential connotations in Cantonese. *Fangyan*, (1), 35–52.

[25] Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.