A Musical Template for Phrasal Rhythm in Spoken Cantonese

Ivan Chow[1], Steven Brown[2], Matthew Poon[3] and Kyle Weishaar[4]

[1]Theoretical and Applied Linguistics Laboratory, the University of Western Ontario
[2]Department of Psychology, Neuroscience and Behaviour, McMaster University
[3]School of the Arts, McMaster University
[4]Department of Linguistics and Languages, McMaster University

ichow3@uwo.ca, stebro@mcmaster.ca, poonm5@mcmaster.ca, weishak@mcmaster.ca

Abstract

Although Cantonese lacks stress at the word level, rhythmic patterns are apparent at the sentence level. In order to develop an understanding of this phenomenon, we took a sentence and manipulated the syllabic content of several of its target words in order to observe the consequences for rhythmic structure. Overall, we found that sentence rhythms conformed to simple musical meters. In addition, we found that syllabic durations could become compressed according to small-integer ratios, such as duplets and triplets. Finally, we observed a tendency for sentences to end on a strong beat, a mechanism that we call the Downbeat Rule.

Index terms: speech, rhythm, Cantonese, duration, music, meter, compression

1. Introduction

Cantonese has no perceptible word-level stress: syllables (including particles) never lose their lexical tones, and the duration of syllables in isolated words is invariant [1]. However, at the prosodic phrase level, rhythmic patterns seem to emerge "out of nowhere" [1]. Along with this, syllables in Cantonese are subject to a fair amount of durational variability, and the duration of syllables can be lengthened or shortened without changing their meaning [1].

In a study on the prosodic marking of syntactic junctures in Cantonese, Perry et al. [2] observed a reciprocal relationship between the duration of a pause at a syntactic juncture and the duration of the syllable immediately preceding the pause. This observation led them to propose that the duration of the syllable and the following pause must fit into two conceptual "timing slots". These timing slots are suggested to be isochronous periods, or "beats", to which prosodic units (e.g., syllables, feet, prosodic words) are aligned in speech production.

Research in cognitive science (e.g., [3], [4]) and psycholinguistics (e.g., [5-8]) has pointed to the presence of an "oscillator coupling" mechanism in speech, musical performance, and other sensorimotor activities, something that enables humans to synchronize gestures to beats. Port [9] provided strong evidence for the view that speakers frequently produce speech "in a periodic way, sometimes by coupling their speech production to another speaker or to a metronome pattern, e.g., when chanting or declaiming" (599), although this may apply to natural speech as well. Speakers are capable of aligning the "perceptual centers" (p-centers) of syllables to metronome beats. While there is no current consensus as to which part of a syllable corresponds to the p-center, much recent research points to the time of voicing onset (cf. [10-12]). "According to Fink et al. [13], a metronome stimulus provides a local 'anchoring' during which movements become less variable" [4: 860]. Following Port's theory, we suggest that the rhythmic patterns of utterances spoken against a metronome should be similar to those of natural speech, but with less temporal variability.

Along these lines, we decided to conduct an experiment to elucidate the nature of rhythm in Cantonese by comparing the production of sentences spoken naturally vs. spoken against a metronome beat. There are several important predictions underlying this experiment. First, we predicted that the rhythmic patterns of spontaneous speech would be similar to those of metronomic speech but with relatively greater durational variability. Second, we predicted that the rhythmic patterns produced in the metronome condition would not be a free-for-all but would instead be subject to constraints on how syllables can be aligned to beats. In other words, speakers' renditions would reveal a combination of permissible (frequent) and impermissible (infrequent or totally absent) rhythmic patterns for Cantonese sentences. Third, by adopting an explicitly musical model of syllable durations, we predicted that compressions of syllable durations would occur according to the kinds of small-integer ratios seen in music, such as duplets and triplets that are commonly found in music. Finally, we predicted that the tendency for syllables to undergo compression was again not a random process but that it would be subject to constraints at the level of morpho-syntax and syllable content.

2. Methods

Our overall experimental approach to analyzing speech rhythm was to start out with a test phrase having a reliable rhythmic structure (shown below in Figure 1 and transcribed musically in Figure 3), and then perform a series of manipulations of its syllable content, namely replacing components of the test phrase (shown in bold in Figure 1) with words of different syllable count, and examining the impact of these replacements on the rhythmic structure of the phrase.

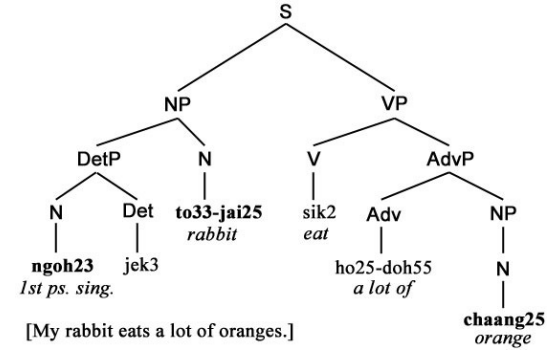

Figure 1. Syntactic structure of the test phrase. Bold words correspond with words that underwent replacement in our experiment, as described in Figure 2. Multisyllabic words are shown as hyphenated here and elsewhere.

Figure 2 shows the test phrase again, with boxes above the target words showing the replacements of the three nouns with alternative nouns having different numbers of syllables. In all, 27 permutations of the test sentence were created by all possible combinations of these replacements, where syntactic structure was kept constant in all cases. This became the sentence corpus of the current study.

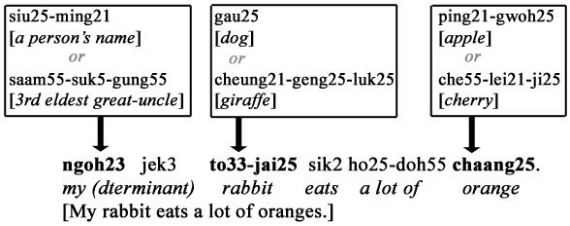

Figure 2. Possible ways of modifying the syllabic content of the test phrase by replacing target nouns with alternative nouns of different syllable count at three locations in the sentence (as indicated by the boxes above the words).

Six participants (three females) participated in this study, as divided into two groups. One group was assigned to the "natural speech" condition, and the other to the "metronome" condition. For the natural speech group, participants were presented with a randomized list of test phrases. They were asked to practice the test phrases until they were fluent. They then uttered the phrases twice as if they were engaging in a normal conversation. Once they were satisfied with their rendition, they moved onto the next test phrase.

In the metronome condition, participants were asked to repeat a randomized list of the test phrases using the rhythmic pattern in which they would normally speak the phrase, while at the same time trying to align their rhythm to a metronome signal played through an on-ear headphone. Eight 30-minute metronome beat tracks were pre-recorded. Their speed ranged from 90 beats per minute (bpm) to 160 bpm, with 10 bpm increments. Participants were given time to practice the test phrases and to choose the metronome speed that most suited them. Once they were confident with the rhythmic patterns, they were given a new randomized list of the same test phrases. Participants were asked to repeat each test phrase several times while trying to align their speech rhythm to the beat. Once they were satisfied with their pattern, they moved onto the next phrase.

In both conditions, recordings were done in an acoustic sound booth using an Apex-181 microphone. In order to simplify the procedure and to minimize performance anxiety, the entire test session was recorded, with the practice sessions later removed. The two best renditions were extracted for speech analysis. Management of voice and beat tracks was carried out using Adobe Audition. In the metronome condition, the metronome signal was played through one track while the voice was recorded onto a second track. This permitted an alignment of the voice and beat tracks for speech analysis.

Speech analysis was carried out using Praat [14]. For each test phrase, we created a textgrid containing five tiers. The first three tiers were segmented by syllable. The first tier contained romanized transcriptions of the syllables. The second contained tone numbers of the respective syllables, and syllabic duration was automatically entered into the third tier. In the fourth tier, boundaries were placed at the locations of voicing onset, while making reference to where formant patterns started to appear in the spectrogram and where a large increase in intensity often occurred. The duration between p-centers was entered within this interval. For rhythmic patterns collected from the metronome condition, the metronome signal in the first track was extracted with identical time signatures as the slice of the voice-track sentence under analysis. Boundaries were automatically placed at each beat. The resulting beat tier was then appended to the textgrid of the test-phrase analysis as the fifth tier. For the natural speech condition, boundaries were placed into the beat tier at p-centers, between which the measured "inter-p-center" durations were more or less equal. If a similar rhythmic pattern was observed in the metronome condition (which was the case for the majority of the test phrases), boundary locations in the natural speech condition were determined with reference to the corresponding test phrase collected in the metronome condition. We then transcribed the rhythmic patterns of all samples into musical notation, while making reference to the duration between voice onsets (see next section for further discussion).

3. Results

A total of 162 test phrases was analyzed, 81 from each test condition. We have not yet completed statistical analyses of our data and so the results presented here reflect qualitative patterns of rhythm production for our subjects.

The principal difference between the results using the metronome and those without it was the greater temporal variability in the absence of the metronome. That difference aside, the rhythmic results were overall similar between the two conditions. In addition, preboundary lengthening ([15], [16]) was found at the end of phrases in both conditions. The remaining discussion is based on the combined results of the two conditions. The following subsections describe a series of rhythmic mechanisms for Cantonese speech, as demonstrated using musical notation.

3.1. Meter

Figure 3 shows a musical transcription of the dominant pattern of production of the test phrase by subjects.

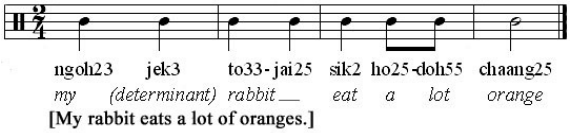

Figure 3. Musical transcription of the rhythmic pattern for the test phrase shown in Figures 1 and 2.

A rhythmic pattern corresponding with duple meter (2/4) was the best representation of the spoken rhythm. As described below, while the manipulations of syllabic content could lead to changes in the rhythmic pattern of the sentence (for example, a change to 3/4 rhythm, see Figure 4), an overall sense of meter was still maintained in these sentences, even without the presence of a metronome. This supports our use of musical models in representing speech rhythm in Cantonese.

In all of our test samples, speakers kept the same meter throughout an entire prosodic phrase. (For utterances consisting of multiple prosodic phrases, there was the possibility for meter changes to occur, such patterns being referred to as "heterometers" [17]. However, this is beyond the scope of the present study.) Depending on the test phrase, syllables were organized into measures of 2, 3, or 4 beats. Anything beyond 4 beats was consistently absent. This agrees with the coupled-oscillator research of Tilsen [4] and Port & Cummins [5] showing that lower-order phases (e.g., $\Phi_{0.33}$, $\Phi_{0.5}$) – which translate to simple time signatures (e.g., 2 or 3 beats per measure) – are favored over higher-order phases (e.g., $\Phi_{0.7}$), which translate roughly to anything beyond 4 beats per measure.

We found that the number of beats in each measure was influenced by syntactic constituency and by the number of syllables per noun. While the nouns in the test phrases were relatively "information-rich", the first syllable of the noun fell on the downbeat and received prominence. In the first rhythmic pattern shown in Figure 4, all nouns (in bold type) were fitted into 3-beat measures, with the first syllable receiving prominence. However, in the second pattern, only the first two nouns fit squarely into its 4-beat measures, while the last noun, che55-lei21-ji25 (*cherry*), was broken up in order to fulfill the Downbeat Rule (see below).

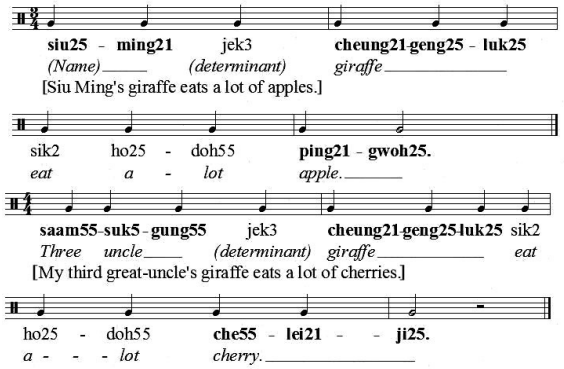

Figure 4. A change in meter based on syntactic constituency. Replacement of target nouns with disyllabic or trisyllabic nouns results in meter changes compared to the original test phrase (shown in Figure 3).

Based on these types of results, we speculate that, although speakers make reference to syntax in the organization of rhythmic patterns, rules that reflect musical well-formedness may be even more important. In addition, although there was a tendency for a syntactic constituent (particularly nouns) to fall within a single measure, patterns that segmented syntactic constituents into successive measures were not uncommon. See section 3.5 for a discussion of optimality-based rule rankings.

3.2. The Downbeat Rule

The Downbeat Rule describes the tendency for Cantonese sentences to end on strong beats. It states that the first or last syllable of the last noun of a phrase falls onto the downbeat of the last measure. Rather than dealing with meter per se, the Downbeat Rule is concerned with the ending of a sentence.

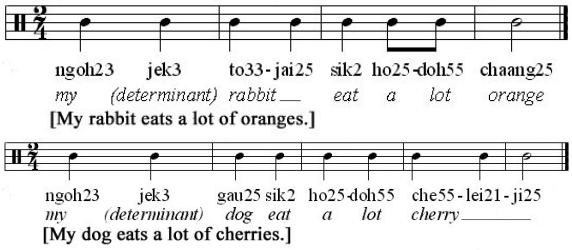

Figure 5. Rhythmic patterns that conform to the Downbeat Rule.

Figure 5 shows two well-formed rhythmic patterns that abide by the Downbeat Rule. For the first sentence, where the last noun is a monosyllabic word, all the rhythmic patterns we collected ended with the last syllable on the downbeat. Yet, when the last noun consisted of more than one syllable, only the first or the last syllable typically fell on the downbeat. A hypothetical exception is seen in Figure 6, where the middle syllable of a trisyllabic word falls on the downbeat.

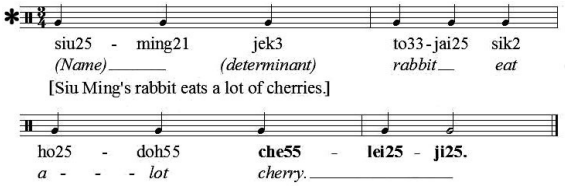

Figure 6. A hypothetical pattern that violates the Downbeat Rule. The middle syllable of a trisyllabic word falls on the downbeat.

Since patterns like this were missing from our collection of samples, we speculate that this is because they are in violation of the Downbeat Rule. In addition, the last noun would be segregated between two measures.

3.3. Syllable compression according to small-integer ratios

Although the majority of rhythmic patterns we observed were isochronous throughout, we frequently observed that syllable durations became compressed (cf. [18], [19]), and did so according to the types of small-integer ratios typically found in music, such as duplets, triplets, and quadruplets. Figure 7 provides one example, which is the basic test phrase from Figures 1, 2 and 3.

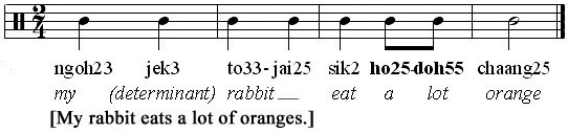

Figure 7. Durational compression of the disyllabic noun ho25-doh55 into a musical duplet.

Durational compression seems to be a mechanism by which Cantonese speakers manipulate rhythmic patterns in order to fulfill both the Meter and the Downbeat rules. However, there are constraints that guide this process. When analyzing trisyllabic combinations – e.g., *cheung21-geng25-luk25* (giraffe: NNN) or *sik2-ho25-doh25* (eat a lot: VAdvAdv) – we observed that trisyllabic nouns were always realized as short-short-long (SSL) patterns, not LSS, whereas phrases consisting of a monosyllabic verb plus a disyllabic adverb or noun were realized as LSS (as shown in the third measure of Figure 7), not as SSL.

An alternative manner of realizing such trisyllabic combinations was in the form of a triplet occupying two beats (referred to as a "polyrhythm" in music). However, the virtual absence of yet another alternative pattern, the syncopated pattern (SLS), led us to the next rule.

3.4. The No Syncopation Rule

Among all the rhythmic patterns collected in this study, syncopated patterns (in which a beat or an accent is displaced so that a strong beat becomes weak) were consistently missing. Because of the importance of the downbeat mechanism for spoken Cantonese, syncopation would make speech sound unnatural. Hence, it seems to be avoided, even in metronomic speech.

3.5. Overview of the Rhythmic Template

We propose that rhythmic patterns in spoken Cantonese are governed by rules organized according to Optimality Theory [20], such as METER, DOWNBEAT, SYNTACTIC CONSTITUENCY, NO SYNCOPATION, and so on. Details of these rules and their relative ranking are currently being analyzed. We suspect that different speakers may use different strategies (different rankings) in the generation of

rhythmic patterns, which may lead to slightly different (albeit acceptable) patterns and thus to rhythmic variability among speakers.

4. A Musical Template

Linguists, cognitive scientists, and musicologists alike have noticed similarities in rhythmic features between speech and music (e.g., [21]). The fact that syntactic constituency is of lesser importance than rules of musical well-formedness leads us to look at speech rhythm in a new light. We believe that phonological units are *not* analogous to timing units nor are they necessarily isochronous at any given level. Metrical phonology ([22], [23]) sees each hierarchical level as linear, i.e., consisting of units of the lower level, distributed in an isochronous manner. Phonological theories on Mandarin (a language of the same language family as Cantonese) consider prosodic phrases as being divided strictly into prosodic feet of two or three syllables ([24]). However, these theories fail to account for the *relative durations* of syllables and the fact that some syllables can occupy full beats while others occupy fractions of beats as a result of compression. Again, in our musical model of speech rhythm, these compressions correspond for the most part with small-integer ratios, such as duplets and triplets.

5. Further issues and future research directions

Since the model we put forward here is based on test phrases of one syntactic structure, it is critical to include phrases of different syntactic structures and modalities in order to analyze the application of the rhythmic template to all utterances. As we examine more types of phrases, we anticipate that the rules that constitute this template may need to be revised. For example, we envision that the Downbeat Rule may need to be revised for phrases that do not end with nouns, e.g., those that end in adverbs (e.g., mei22 *yet*), particles (e.g., ah33, la55, me33), or even phrases in different modalities (questions, for example).

Although Cantonese is categorized as a "syllable-timed" language, results from our study indicate that syllables are not always isochronous. Rather, they are fitted into timing units of different strengths within the rhythmic structure. Brown and Weishaar [17] showed that the rhythm of English (a "stress-timed" language) can be described with a very similar model. We plan on including more languages of different rhythmic types (e.g., French, Japanese) in our study in order to determine how such a template may account for the rhythmic patterns of these languages.

6. Conclusions

The present study indicates that the rhythmic structure of Cantonese is non-linear. Although Flynn [1] points out that phrasal rhythm in Cantonese seems to "emerge out of nowhere", speech rhythm is not random and unpredictable. Instead, reproducible phrase patterns are observed across speakers and they correspond more or less with rules of musical well-formedness.

Research on speech rhythm based on coupled oscillators or similar models indicates that speech can be analyzed as combinations of synchronized gestures. The metronome stimulus provides an "anchoring" so that the timing of these gestures becomes less variable. In our study, we found that rhythmic patterns in natural and metronomic speech were very similar, although natural speech exhibited greater temporal variability. Based on the musical transcriptions of the observed rhythmic patterns, we developed a music-inspired template to account for phrasal rhythmic patterns in spoken Cantonese. Rules that make reference to terminologies from both linguistics and music are combined to develop a template capable of predicting phrasal rhythm.

7. References

[1] Flynn, C.-Y.-C. 2003. *Intonation in Cantonese*. Munich: Lincom Studies in Asian Linguistics.

[2] Perry, C. Wong, R. K.-S. & Matthews, S. 2009. Syllable timing and pausing evidence from Cantonese. *Language and Speech* 52(1). 29-53.

[3] Patel, A. D., Iversen, J. R., Chen, Y. & Repp, B. H. 2005. The influence of meter and modality on synchronization with a beat. *Experimental Brain Research* 163. 226-238.

[4] Tilsen. S. 2009. Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science* 33. 839-879.

[5] Cummins, F. & Port, R. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26. 145-171.

[6] O'Dell, M. & Nieminen, T. 2008. Coupled oscillator model of speech rhythm. *Proceedings Speech Prosody 2008*. 1075-1078.

[7] Van Lieshout, P. 2004. Dynamical systems theory and its application in speech. In B. Maassen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (eds.), *Speech motor control in normal and disordered speech*, 51-82. Oxford: Oxford University Press.

[8] Golstein, L., H. Nam, Saltzman, E. & Chitoran, I. 2008. Coupled oscillator planning model of speech timing and syllable structure. *Proceedings of the 8th Phonetic Conference of China and the International Symposium on Phonetic Frontiers*. Beijing.

[9] Port, R. 2003. Meter and speech. *Journal of Phonetics* 31. 599-611.

[10] de Jong, K. 1994. The correlation of P-center adjustments with articulatory and acoustic events. *Perception & Psychophysics* 56(4). 447-460.

[11] Patel, A. D., Löfqvist, A. & Naito, W. 1999. The acoustics and kinematics of regularly timed speech: A database and method for the study of the p-center problem. *Proceedings of the 14th International Congress of Phonetic Sciences* 1. 405-408.

[12] Scott, S. 1998. The point of P-centres. *Psychological Research* 61. 4-11.

[13] Fink, P., Foo, P., Jirsa, V. & Kelso, J. 2000. Local and global stabilization of coordination by sensory information. *Experimental Brain Research* 134. 9-20.

[14] Boersma P. & D. Weenink. 2006. *Praat: Doing phonetics by computer*. [Computer Program]. Institute of Phonetic Sciences, University of Amsterdam.

[15] Gussenhoven, C. & Rietveld, A. C. M. 1992. Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics* 20(3). 283-303.

[16] Swerts, M. G. J., 1997. Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America* 101(1). 514-521.

[17] Brown, S. & Weishaar, K. 2010. Speech is "heterometric": The changing rhythms of speech. *Proceedings of the Speech Prosody 2010 Conference*. Chicago, IL.

[18] Lehiste, I. 1970. *Suprasegmentals*. Cambridge: MIT Press.

[19] Nooteboom, S. G. 1972. *Production and perception of vowel duration: A study of durational properties of vowels in Dutch*. Eindhoven: Philips Research Laboratories.

[20] Prince, A. & Smolensky, P. 2004. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell Publishing.

[21] Ramus, F., Nespor, M. & Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3). 265-292.

[22] Nespor, M. & Vogel, I. 1986. *Prosodic phonology*. Dordrecht: Foris Publications.

[23] Liberman, M. 1975. *The intonational system of English*. Ph.D. Thesis, M.I.T., Cambridge, MA. Published by Indiana University Linguistics Club.

[24] Duanmu, S. 1996. Pre-juncture lengthening and foot binarity. *Studies in the Linguistic Science* 26(1/2). 95-115.