Quantitative Analysis of Intonation Patterns Produced by Cantonese Speakers with Parkinson’s Disease: A Preliminary Study

Joan K-Y Ma, Tara L. Whitehill

Division of Speech and Hearing Sciences, University of Hong Kong, Hong Kong SAR
joanna@graduate.hku.hk, tara@hku.hk

Abstract
This preliminary study aimed to apply the command-response model to the analysis of disrupted intonation patterns in speakers with Parkinson’s disease (PD). Three Cantonese PD speakers with mild prosodic impairment participated. The speech stimuli were 36 utterances (questions and statements). Productions were analyzed using the model parameters (base frequency, magnitude and onset time of each phrase command, and amplitude and duration of each tone command). The results showed that Cantonese PD speakers marked the question-statement contrast only by adding a question boundary tone command and contrasting the duration of the tone command at the final position, but not by differences in base frequency, as previously reported for non-dysarthric speakers. This study showed promising results for using the command-response model to analyze intonation patterns in Cantonese dysarthric speech. Additional modeling issues for this group of speakers are discussed.

Index Terms: Parkinson’s disease, dysarthria, intonation production, command-response model

1. Introduction
Dysarthria is characterized by deficits in prosody, articulation, resonance, phonation and respiration [1]. Prosodic impairment is one of the most common characteristics in speakers with the hypokinetic dysarthria associated with PD (hereafter, PD speech). Prosody refers to the aspects of the speech signal used in conveying both linguistic (e.g., intonation and stress) and paralinguistic (e.g., anger and sadness) meanings through variations in fundamental frequency (F0), intensity and duration of speech, quality of voice and degree of reduction [2]. In the current study, we focused on how question-statement contrasts are marked in PD speech in Cantonese.

In recent years, the assessment of prosodic impairment using acoustic analysis has received increasing attention [e.g. 3, 4]. Acoustic measures of different aspects of prosody such as timing (e.g. syllable duration and pause proportion), frequency (e.g. average F0, F0 standard deviation (SDF0), median and maximum formant slope) and intensity (e.g. intensity variation) have been used in studies of dysprosody in dysarthric speech. However, there is no commonly accepted set of acoustic measures of prosodic impairment, owing to the multidimensional nature of the prosodic attributes. Penner and colleagues [3] and Kent and Kim [5] argued that measures such as average F0 and SDF0 provide only an overall measure but no information on intonation contours. Measuring only a single point within each syllable or only overall F0 range would miss important frequency changes within a syllable or within a sentence. Linguistic framework was applied in evaluating intonation patterns in dysarthric speech by looking at overall F0 contours [e.g., 3, 6].

In addition to the application of linguistic frameworks, Kent and Kim [5] proposed the use of intonation models for the analysis of intonation in dysarthric speech. Computational models of intonation have been developed to analyze the F0 contours of intonation using a series of well-defined parameters to elucidate the relationship between the surface F0 contour and linguistic representations. Examples of such models include ToBI, command-response model, STEM-ML model and PENTA model. Applying intonation models to the study of intonation patterns in tone languages is complicated by the fact that F0 patterns serve two separate functions (i.e., intonation and tone) at the same time. Specifically, F0 patterns at the syllabic level mark lexical tone, while F0 variations at the sentential level mark intonation. Several of the above-mentioned intonation models have been adapted to Cantonese. In particular, the command-response model has been found to accurately model intonation and have a good correlation with linguistic representations in Cantonese [7, 8]. The command-response model aims at the generation of F0 contours of intonation using a mathematical formulation. The F0 contour of intonation, which employs a logarithmic scale, is described as a linear superposition of a phrase component and an accent component (referred to as the tone component in tone languages) on a baseline level. Additionally, the command-response model has a physiological base specified for each individual component, thus offering the potential to account for disordered intonation patterns. Hence, the command-response model was chosen for the purpose of the current study.

The question-statement contrast in Cantonese has been studied using the command-response model [7, 8]. Ma and colleagues [8] used a set of stimuli with question-statement contrasts and found that questions differ significantly from statements by having (i) a higher base frequency, (ii) an additional boundary tone command towards the end of the final syllable, and (iii) a longer overall duration of tone command for the final syllable. In the present study, similar stimuli were obtained from three native Cantonese PD speakers. The aims of this study were (i) to determine the feasibility of applying the command-response model to the analysis of Cantonese dysarthric speech, and (ii) to determine the acoustic features of impaired intonation production in Cantonese PD speech.

2. Method

2.1. Speakers
Three native Cantonese speakers (CWY, LPS and YYH) with idiopathic Parkinson’s disease took part in the study (all male, age 50 – 60). All three speakers were judged to have mildly impaired prosody based on a sample from a passage reading task. The severity of dysprosody was determined independently by the first and second authors, who are qualified speech-language pathologists with more than five years experience with dysarthric speech. All speakers showed normal language, hearing ability and oro-peripheral structures.

2.2. Stimuli
Two intonation patterns were studied: statements and intonation questions. The carrier phrase /le15 kɔ13 ti32 h4132,
X (This word is X), which can be produced as either a question or a statement, was used. Three sets of words, derived from the syllables /si/, /ji/ and /jju/, were embedded in the above sentence context. Each set consisted of six words that differed only in tone. With two intonations and 18 words, a total of 36 stimuli were collected from each speaker.

2.3. Data collection

Data collection was carried out in a quiet room, with either a Shure (SM48) or an AKG (C525B) microphone connected to a computer running the Audacity software and an Aardvark USB 3 sound card. A 10 cm mouth-to-microphone distance was maintained during recordings. In order to make the speech sample as natural as possible, each subject was engaged in a dialogue with the experimenter during data collection. An additional 24 dialogues with the tonal contrast embedded in the initial or medial positions of the target sentences were included in order to reduce the regularity over repeated productions of the same target sentence structure. The dialogues were presented visually on the screen of a G4 Apple Macintosh running a Hyper-Card (Apple) program. The dialogues were presented visually on the screen of a G4 Apple Macintosh running a Hyper-Card (Apple) program.

In the current study, the parameters Fb, Api and T0i of the command-response model [9]. According to Fujisaki & Hirose [9], the base frequency (Fb) is the F0 baseline of a sentence in the absence of other components. The phrase component is produced by the impulse responses (phrase commands) of the phrase control mechanism. Each phrase command is defined by the magnitude of the command (Ap) and the onset time (T0). There is also a time constant (τ) for the phrase control mechanism. Tone component is a response of the step-wise input signals (tone commands) of the tone control mechanism. The tone commands of both positive and negative polarities are defined by amplitude (Ap), attack time (T0) and offset time (Td). There is a time constant (τ) for the tone control mechanism, and also a ceiling level (γ) of the tone command. Both mechanisms are assumed to be critically-damped second-order linear systems.

In the current study, the parameters Fb, Ap, T0 and Td of the ith phrase command, and Ap, T0 and Td of the jth accent command were used to quantify the F0 pattern. In addition, the duration of each tone command (Dj) was obtained by calculating the difference in onset and offset time (T0 - Td).

FujiParaEditor [10] was used for analysis. The F0 pattern was extracted by an autocorrelation algorithm. Additionally, the F0 contour of the whole sentence was also approximated by estimating the F0 at nine evenly-spaced time points from the beginning to the end of the voiced segment of each syllable using an autocorrelation algorithm in Praat [11]. In cases when unreliable F0 estimates were obtained, F0 measurement was obtained manually instead. As F0 irregularities as a result of factors like glottal fry and poor voice quality (e.g. harshness) are frequently observed in PD speech, the accuracy of the F0 contours extracted by FujiParaEditor was checked against the F0 values obtained from the additional acoustic analysis.

In modelling of the F0 contour, automatic calculation by FujiParaEditor was used for initial analysis, followed by manual adjustment using analysis-by-synthesis. For each stimulus, the Fb was set as the lowest F0 value within the utterance. The number of phrase and tone commands was determined by linguistic constraints such as syllable boundary, and the magnitude and timing of each command was approximated to generate the closest F0 contour resembling the actual utterance. A set of tone command patterns for Cantonese has been previously proposed [12]. They are positive for tone 55, negative followed by positive for tone 25, zero for tone 33, negative for tone 21, negative followed by zero for tone 23 and negative for tone 22. Previous research has shown that the final-rise in F0 in questions can be generated by an addition of tone command (i.e., boundary tone command) [7]. However, these default commands were not always appropriate in modelling PD speech. In some cases, an additional command needed to be added or a command might need to be omitted in order to achieve a close approximation to the F0 contour. This judgement was always made by comparing the re-synthesized signal, using the LPC-based re-synthesis tool of FujiParaEditor, to the original signal, perceptually. An accurately modelled signal should sound the same as the original signal and the F0 contour of the modelled signal should closely approximate the F0 contour of the original signal. In the command-response model, parameters α and β are assumed to be constant within a stimulus, and were set as the default value of the program (2 and 20 respectively). The parameter γ was set as 0.9 in this experiment.

3. Data analysis

The F0 pattern of the stimuli was analysed according to the command-response model [9]. In this model, the F0 pattern of a sentence is considered as a linear superposition of the phrase component (denoting the global F0 changes, i.e., intonation) and the tone component (denoting the local F0 changes, e.g. tone) on a baseline frequency.

Each tone component is a response of the step-wise input signals (tone commands) of the tone control mechanism. The tone commands of both positive and negative polarities are defined by amplitude (Ap), attack time (T0) and offset time (Td). There is a time constant (τ) for the tone control mechanism, and also a ceiling level (γ) of the tone command. Both mechanisms are assumed to be critically-damped second-order linear systems.

In the current study, the parameters F0, Ap and T0 of the ith phrase command, and Ap, T0 and Td of the jth accent command were used to quantify the F0 pattern. In addition, the duration of each tone command (Dj) was obtained by calculating the difference in onset and offset time (T0 - Td).

FujiParaEditor [10] was used for analysis. The F0 pattern was extracted by an autocorrelation algorithm. Additionally, the F0 contour of the whole sentence was also approximated by estimating the F0 at nine evenly-spaced time points from the beginning to the end of the voiced segment of each syllable using an autocorrelation algorithm in Praat [11]. In cases when unreliable F0 estimates were obtained, F0 measurement was obtained manually instead. As F0 irregularities as a result of factors like glottal fry and poor voice quality (e.g. harshness) are frequently observed in PD speech, the accuracy of the F0 contours extracted by FujiParaEditor was checked against the F0 values obtained from the additional acoustic analysis.

In modelling of the F0 contour, automatic calculation by FujiParaEditor was used for initial analysis, followed by manual adjustment using analysis-by-synthesis. For each stimulus, the Fb was set as the lowest F0 value within the utterance. The number of phrase and tone commands was determined by linguistic constraints such as syllable boundary, and the magnitude and timing of each command was approximated to generate the closest F0 contour resembling the actual utterance. A set of tone command patterns for Cantonese has been previously proposed [12]. They are positive for tone 55, negative followed by positive for tone 25, zero for tone 33, negative for tone 21, negative followed by zero for tone 23 and negative for tone 22. Previous research has shown that the final-rise in F0 in questions can be generated by an addition of tone command (i.e., boundary tone command) [7]. However, these default commands were not always appropriate in modelling PD speech. In some cases, an additional command needed to be added or a command might need to be omitted in order to achieve a close approximation to the F0 contour. This judgement was always made by comparing the re-synthesized signal, using the LPC-based re-synthesis tool of FujiParaEditor, to the original signal, perceptually. An accurately modelled signal should sound the same as the original signal and the F0 contour of the modelled signal should closely approximate the F0 contour of the original signal. In the command-response model, parameters α and β are assumed to be constant within a stimulus, and were set as the default value of the program (2 and 20 respectively). The parameter γ was set as 0.9 in this experiment.

4. Results

4.1. Qualitative analysis

Qualitative analysis was conducted to examine the modelling accuracy by inspecting the model parameters of each stimulus produced by all three speakers. Generally speaking the stimuli could be modelled accurately with the same number of phrase and tone commands as have been found for non-dysarthric speakers [8]. However, three types of variation were observed. The first, which was only found in the production of questions, reflected disturbances in intonation production. In these cases (6% of stimuli), the speakers did not mark questions with a final-rise in F0, that is, there was no question boundary tone command for these questions. The second inconsistency was observed in the production of tones (8%), where the tone command used to model the final syllable was not consistent with the previously proposed value associated with Cantonese tones [12]. For example, tone 33 was modelled with a positive tone command instead of zero tone command. The perceptual identity of these tones was ambiguous according to the judgement of the first author. Third, in 3% of the stimuli, positive tone commands were used to model word 3 (/tsi/) and word 4 (/hi/) instead of negative tone commands as have been found for non-dysarthric speakers [12]. However, these default commands were not always appropriate in modelling PD speech. In some cases, an additional command needed to be added or a command might need to be omitted in order to achieve a close approximation to the F0 contour. This judgement was always made by comparing the re-synthesized signal, using the LPC-based re-synthesis tool of FujiParaEditor, to the original signal, perceptually. An accurately modelled signal should sound the same as the original signal and the F0 contour of the modelled signal should closely approximate the F0 contour of the original signal. In the command-response model, parameters α and β are assumed to be constant within a stimulus, and were set as the default value of the program (2 and 20 respectively). The parameter γ was set as 0.9 in this experiment.

4.2. Quantitative analysis

The mean value for each parameter (F0, Ap, T0 and Dj) of each intonation for each tone was calculated for each speaker. Two-way ANOVAs were used to analyse the data for each parameter separately, with Speaker as the between-subject factor, and Intonation as the within-subject factor.
4.2.1. Base Frequency

Statistical analysis showed a significant main effect for Speaker \( F(2, 30) = 162.08, p < 0.001 \), but not for Intonation \( F(1, 30) = 0.66, p > 0.05 \). Post-hoc analysis showed that the F0 value of CWY (mean = 136.9) was significantly higher than the other two speakers (Tukey HSD test, \( p < 0.05 \) for both); while the F0 value of YYH (mean = 128.70) was significantly higher than that of speaker LPS (mean = 87.11) (Tukey HSD test, \( p < 0.05 \)). No significant interaction effect was found between Intonation x Speaker \( F(2, 30) = 0.14, p > 0.05 \).

4.2.2. Phrase component

Two phrase commands were used to model the stimuli in this experiment, in consideration of the prosodic structure of the speakers’ productions and the modelling accuracy. Therefore, a three-way ANOVA was used to compare the magnitude of the phrase commands, as the difference between the two phrase commands was also considered. The results showed a significant main effect for Speaker \( F(2, 60) = 3.42, p < 0.05 \). Post-hoc analysis showed that speaker YYH had a significantly higher average \( A_n \) than speaker CWY (Tukey HSD test, \( p < 0.05 \)). No significant contrast was found between speakers CWY and LPS, nor between speakers YYH and LPS (Tukey HSD test, \( p > 0.05 \) for both). The main effect of the Magnitude of the phrase command showed that the average value of \( A_{t1} \) (mean = 0.32) was significantly higher than that of \( A_{t2} \) (mean = 0.17) \( F(1, 60) = 92.86, p < 0.001 \). However, there was no significant difference in \( A_p \) values between questions and statements \( F(1, 60) = 0.27, p > 0.05 \).

There was a significant interaction effect between Speaker x Magnitude of the phrase command \( F(2, 60) = 14.86, p < 0.001 \). Post-hoc analysis showed that the decrease in \( A_p \) value from \( A_{t2} \) to \( A_{t1} \) was significantly higher for speaker LPS than the other two speakers (Tukey HSD test, \( p < 0.05 \) for all).

The onset times of the two phrase commands were compared across Speaker and Intonation. Results showed a significant contrast between Speaker for both \( T_{i0} \) \( F(2, 30) = 7.97, p < 0.005 \) and \( T_{i0} \) \( F(2, 30) = 82.96, p < 0.001 \). Post-hoc analysis showed that the \( T_{i0} \) for speaker LPS was earlier than that of CWY (Tukey HSD test, \( p < 0.05 \)); and a significant contrast in \( T_{i0} \) was found between all three speakers (Tukey HSD test, \( p < 0.001 \) for all). A significant main effect of Intonation was observed for \( T_{i0} \) \( F(1, 30) = 4.35, p < 0.05 \) but not for \( T_{i0} \) \( F(1, 30) = 0.013, p > 0.05 \). The \( T_{i0} \) for statements was earlier than that of questions.

4.2.3. Tone component

The number of tone commands used to model the local F0 changes of tones in each syllable depends on the tone of that syllable, as stated earlier. The amplitude (\( A_{t} \)) and duration (\( D_{t} \)) for the second syllable was not compared, as the word /kho\ was modelled with a zero tone command. Statistical analysis comparing the amplitude of each tone command showed that the main effect of Speaker was significant for \( A_{t1} \) \( F(2, 30) = 24.72, p < 0.001 \) and \( A_{t3} \) \( F(2, 30) = 24.41, p < 0.001 \), but not for \( A_{t5} \) \( F(2, 30) = 1.73, p > 0.05 \) and \( A_{t3} \) \( F(2, 30) = 1.53, p > 0.05 \). Post-hoc analysis showed that speaker YYH (mean = 0.43 for \( A_{t1} \) and 0.38 for \( A_{t3} \)) showed significantly higher \( A_{t1} \) and \( A_{t3} \) than speakers CWY (mean = 0.26 for \( A_{t1} \) and 0.27 for \( A_{t3} \)) and LPS (mean = 0.32 for \( A_{t1} \) and 0.29 for \( A_{t3} \) (Tukey HSD test, \( p < 0.001 \) for all). No statistically significant differences were found in \( A_{t1} \) and \( A_{t3} \) between speakers CWY and LPS (Tukey HSD test, \( p > 0.05 \) for both). No significant main effect was noted for Intonation for all four tone commands \( A_{t1}, A_{t3}, A_{t5}, A_{t3}; p > 0.05 \) for all). Additionally, no significant interaction effect was noted between Speaker x Intonation for all tone commands \( p > 0.05 \) for all).

The duration of each tone command was calculated by the difference in time between the onset (\( T_{j1} \)) and the offset (\( T_{j0} \)) of each tone command. In the case of the final syllable in questions, where two tone commands were used to model a single syllable, comparison was carried out between (i) the duration of the original tone command in questions and statements, and (ii) the total duration of the original tone command and the boundary tone command in questions and the duration of the original tone command in statements. The result showed a significant main effect of Speaker for the duration of all the tone commands \( p < 0.05 \) for all). Post-hoc analysis showed that speaker LPS showed a significantly shorter \( D_{t1} \) than speakers CWY and YYH (Tukey HSD test, \( p < 0.001 \) for both). However, speaker LPS showed a significantly longer \( D_{t3} \) than speakers CWY and YYH, a significantly larger \( D_{t4} \) than speaker CWY and a significantly larger \( D_{t5} \) than speaker YYH (Tukey HSD test, \( p < 0.05 \) for all). No significant main effect of Intonation was found for \( D_{t1}, D_{t3}, D_{t4} \) (\( p > 0.05 \) for all). A significant main effect of Intonation was observed in comparing the original tone command of the final syllable \( (D_{t5}) \) \( F(2, 30) = 8.97, p < 0.05 \). The \( D_{t5} \) of statements was significantly longer than that of questions. However, when the duration of the entire syllable at the final position were considered, the combined duration of the original tone command and the boundary tone command in questions were significantly longer than the duration of the original tone command in statements \( F(1, 24) = 49.56, p < 0.001 \).

5. Discussion

The current study was the first to apply a computational model of intonation to the analysis of dysarthric speech. The results showed that questions and statements produced by Cantonese speakers with dysarthria could be accurately modelled using the command-response model, as the F0 contour of the original signal was closely approximated using the model parameters \( F_{b}, A_{t0}, T_{0i}, A_{tj}, T_{1j} \) and \( T_{2j} \). Although the model showed great potential for analyzing intonation patterns in speakers with prosodic impairment, several issues need to be considered. First, the extraction of F0 contour in dysarthric speech is not always straightforward. Deviant voice quality (such as rough voice and breathy voice) is one of the most severely affected areas in PD speech [1]. Additionally, glottal fry was commonly observed in the current data set. These abnormal voice qualities, which are characterized by F0 irregularities, may cause degradation in the quality of the extracted F0 [13]. As a result, it is important to determine that the F0 contour extracted is accurate. In the current analysis, the extracted F0 contour was checked against F0 values obtained from additional acoustic analysis.

Second, in the modelling of the phrase and tone components, although an infinite number of phrase or tone commands can be used to achieve the closest approximation, the number should be constrained to ensure that the parameters obtained can be interpreted linguistically [13]. However, in modelling the intonation of PD speech, not all commands could easily be explained linguistically. For example, in three utterances produced by speaker CWY, a sudden rise in F0 was observed in the third and fourth syllables. As a result, these tone 22 syllables were modelled by a positive tone command, which deviated from the proposed polarities of tone commands in Cantonese [12]. This was not unexpected as physical constraints are imposed on the speech production mechanism of the PD speakers. However, this may be an obstacle to the linguistic interpretation of the
model parameters. In the current study, deviation of phrase or tone commands was only observed in a small portion of the dataset and was dealt with by careful documentation and qualitative analysis. However, all the speakers in the current study were judged to have mild prosodic impairment. Further investigation is needed for speakers with more severe impairment.

Another objective of this study was to investigate the production of question-statement contrasts in Cantonese PD speakers by employing a quantitative model. As mentioned earlier, by using the command-response model, Ma, Ciocca and Whitehill [8] previously showed that questions and statements produced by non-dysarthric Cantonese speakers are mainly contrasted by (i) differences in base frequency \(F_0\), (ii) the addition of question boundary tone command, and (iii) the shortening of the original tone command and the increase in the combined duration of the original tone command and question boundary tone command at the final position of questions. By analyzing the same intonation contrast produced by Cantonese PD speakers, we found that, although the PD speakers do mark the difference between questions and statements by (ii) and (iii), no statistically significant difference in \(F_0\) was observed between their question and statement productions. Although previous studies have argued that the final-rise in \(F_0\) in questions is the most significant cue in the perception of questions [14], the significance of \(F_0\) contrast in the non-final portion of questions has also been reported [15]. In cases when the final-rise in \(F_0\) was not clearly presented in a question, a stimulus with a higher \(F_0\) level was more likely to be perceived as a question than that one with a lower \(F_0\) level.

Speaker variability was also explored, as dysarthric speech is known to be heterogeneous. Two speakers (CWY and YYH) had relatively high \(F_0\) values for male speakers while speaker LPS showed a \(F_0\) value compatible with non-dysarthric males [15]. Although abnormal pitch is not commonly reported of hypokinetic dysarthria, elevated \(F_0\) levels have been reported [1]. As mentioned earlier, the tone 22s of the third and fourth words were modelled with positive tone commands in three of the utterances produced by speaker CWY. This finding was unexpected as PD speech is commonly characterized by monopitch [1]. One explanation is the poor laryngeal control in dysarthric speakers [1], which resulted in the \(F_0\) variability. Another observation was that speaker LPS showed a significantly larger decrease from \(Ap_2\) to \(Ap_3\) than the other two speakers. Fujisaki and Hirose [9] described the occurrence of the second phrase command as re-phrasing, with the second phrase command superimposed on the first phrase command. The small \(Ap_2\) of speaker LPS suggested limited rephrasing on the phrase curve of \(Ap_1\), resulting in a more monotonous-sounding intonation, as confirmed via perceptual judgement, than the other two speakers.

In conclusion, the present study aimed to determine the feasibility of applying the command-response model to the analysis of intonation in PD speech. The preliminary data showed a high modelling accuracy. The model parameters could largely be used to elucidate the relationship between the acoustic measures and their linguistic representations. Additionally, the contrast between linguistic structures (statements and questions) could be compared quantitively using the model parameters. The command-response model showed good potential for further application in the analysis of dysarthric speech. Speakers with different degrees of prosodic impairment will be included in future studies.

6. Acknowledgements
We gratefully acknowledge Hong Kong Parkinson’s Disease Association for assistance in subject recruitment, and Katherine Cheung and Susanne So for their help in data collection.

7. References