<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Perception of vocal emotions produced by Cantonese speakers with hypokinetic dysarthria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Tang, Hau-shuen, Hannah; 鄧巧璇</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Tang, H. H. [鄧巧璇]. (2013). Perception of vocal emotions produced by Cantonese speakers with hypokinetic dysarthria. (Thesis). University of Hong Kong, Pokfulam, Hong Kong SAR.</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2013</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/238532">http://hdl.handle.net/10722/238532</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; The author retains all proprietary rights, (such as patent rights) and the right to use in future works.</td>
</tr>
</tbody>
</table>
Perception of Vocal Emotions produced by Cantonese speakers

with Hypokinetic Dysarthria

Tang Hau Shuen, Hannah

A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2013.
Abstract

Hypokinetic dysarthria is the most common type of speech disorder associated with Parkinson’s disease (PD). Individuals with hypokinetic dysarthria were found to have impaired prosody. The purpose of this study was to investigate the perception of vocal emotions conveyed by Cantonese speakers with hypokinetic dysarthria associated with PD, in comparison to healthy controls. Perceptual ratings of four perceptual dimensions (loudness level, pitch height, pitch variability, speech rate) by listeners for speech samples produced by speakers with PD who intended to convey different emotions were also examined. Speech samples with six different intended emotions (happy, sad, disgust, surprised, angry, neutral) were previously collected from twenty speakers with hypokinetic dysarthria and seven healthy controls. Twenty listeners then identified the perceived vocal emotions and rated the four perceptual dimensions. Results showed that there was significant difference in the identification accuracy of vocal emotions between speakers with PD and healthy controls. In particular, vocal emotion angry conveyed by speakers with PD was significantly difficult for listeners to recognize. Also, listeners perceived the speech samples produced by speakers with PD to be with softer voice, lower pitch height, and smaller pitch variability when compared to those produced by healthy controls. These findings provide empirical evidence for the argument that vocal emotional communication is impaired in Cantonese speakers with hypokinetic dysarthria associated with PD. They also address the focus in treatment for improving expression of vocal emotions for individuals with PD.
Perception of Vocal Emotions produced by Cantonese speakers with Hypokinetic Dysarthria

Parkinson’s disease (PD) is a progressive idiopathic neurological disease of the central nervous system, specifically the basal ganglia. The pathologic changes of PD most often involve loss of nerve cells in the substantia nigra and locus ceruleus, as well as decreased dopamine content in the striatum. The major clinical symptoms of PD include muscle rigidity, tremor at rest, akinesia, and bradykinesia (Hoehn & Yahr, 1967). Dopamine insufficiency within the basal ganglia in patients with PD was found to be the cause of limited muscular control over the physiological support for speech, such as the larynx and the oral cavity (Kegl, Cohen, & Poizner, 1999). Therefore speech disorders are commonly identified in patients with PD and they manifest in approximately 90% of autopsy-confirmed PD cases during the course of the disease (Muller et al., 2001). Hypokinetic dysarthria is the most common type of speech disorder associated with PD (Duffy, 2005). Prosodic insufficiency was the only cluster of deviant speech characteristics found by Darley, Aronson, & Brown (1969a) in their group of PD patients with hypokinetic dysarthria. Prosody refers to the variations in the suprasegmental aspects of speech which are essential for effective communication (MacPherson, Huber & Snow, 2011). There are linguistic prosody (e.g. stress, intonation) and affective prosody (e.g. emotions) (Duffy, 2005). Fundamental frequency (F₀, pitch), intensity (loudness), and durational (tempo/rate) components of syllables, words, phrases, and sentences were found to contribute to prosodic sufficiency in acoustic studies (Duffy, 2005). Dysprosody in the speech of PD patients with hypokinetic dysarthria was found to be attributable to motor deficit (Duffy, 2005) as well as impaired perception of emotional prosody in more advanced stages of the disease (Breitenstein, Daum, & Ackermann, 1998). Speech produced by speakers with hypokinetic dysarthria is generally
described acoustically as having less variability in $F_0$ (monopitch) and intensity (monoloudness) as well as deviated speech rate than the speech produced by healthy age-matched individuals (Duffy, 2005; Harel, Cannizzaro, Cohen, Reilly, & Snyder, 2004).

Although a number of studies have been conducted to investigate the acoustic and perceptual characteristics of the speech produced by patients with PD, little was done on investigating how these patients convey emotions through prosody (Pell, Cheang, & Leonard, 2006; Schröder, 2010) and no related studies were done on the Cantonese population. Pell et al. (2006) found that listeners were significantly less able to recognize intended emotions in speech produced by speakers with PD, especially anger and disgust. Also, the speech produced by PD speakers was frequently perceived as sounding sad or neutral, even when the speakers intended to convey different emotions (Pell et al., 2006). Schröder (2010) concluded that previous data suggested that disturbances of emotional processing, together with motor impairment, contribute to the impaired expression of emotional prosody found in PD.

A number of studies were done to establish the perceptual dimensions that correlate or cue the perception of a particular vocal emotion in non-impaired speakers. Previous studies found that speech rate and $F_0$ (the acoustic correlate of perceived pitch) exert the strongest effect on perception of vocal emotions of non-impaired speakers (Banse & Scherer, 1996). Fast speech rate was found to correlate with the perception of angry and frightened emotions whereas slower speech rate was found to correlate with the perception of sad emotion (Banse & Scherer, 1996). Recognition of happy, frightened, and neutral emotions increased with an increase in pitch variation whereas small pitch variation was found to cue the perception of sad and neutral emotions (Breitenstein, Van Lancker, & Daum, 2001).
Aims

The current study aimed to investigate the identification of vocal emotions expressed by speakers with hypokinetic dysarthria associated with PD. Perceptual ratings of loudness level, pitch height, pitch variability, and speech rate by listeners for speech samples from PD speakers with different intended emotions were also examined. Most studies previously done on perception of vocal emotions focused on non-impaired speakers (Banse & Scherer, 1996; Breitenstein et al., 2001; Pittam & Scherer, 1993). All studies previously done on perception of vocal emotions produced by PD speakers focused in English speaking population (Pell, Cheang, & Leonard, 2006; Schröder, 2010). In this study, we would like to compare the identification accuracy of vocal emotions produced by PD speakers and those produced by non-impaired speakers with Cantonese as the speakers’ first language. It was hypothesized that vocal emotions produced by PD speakers would be less accurately identified by listeners when compared to those produced by non-impaired speakers due to deviant speech characteristics associated with hypokinetic dysarthria. It is also hypothesized that the ratings of perceptual dimensions will be different between the two speaker groups due to the limited motor functioning in patients with PD. Furthermore, this study aimed to establish the perceptual characteristics associated with different vocal emotions produced by Cantonese speakers.

Method

Participants

Speakers. The speakers for this study were twenty adults diagnosed with idiopathic Parkinson’s disease (PD) and seven matched healthy control participants. These twenty-seven speakers were recruited under a larger project (Kwan, in preparation). All were native Cantonese speakers or had at least 20 years of speaking Cantonese. They had adequate hearing ability (with
threshold ≤35 dBHL at 500, 1000, 2000, and 4000 Hz for the better ear), normal cognitive ability (with total mark ≥27/30 in the Montreal Cognitive Assessment) (Nasreddine et al., 2005), normal expressive and receptive language ability (as screened by the short form of Cantonese Aphasia Battery) (Yiu, 1992), and no depression (as screened by the Becks Depression Inventory – Hong Kong Cantonese Version) (Chinese Behavioural Science Corporation, 2000). Individuals in each group (PD and HC) were comparable in age (PD: $M = 59.6$, $SD = 6.34$; HC: $M = 57.86$, $SD = 5.67$) and years of education (PD: $M = 9.45$, $SD = 2.78$; HC: $M = 8.86$, $SD = 1.46$). Speakers in the PD group were diagnosed by neurologists as having Parkinson’s disease Stage 2 to Stage 3 using the Hoehn & Yahr Scale (Hoehn & Yahr, 1967) but with no other neurological disease. They were all judged to demonstrate hypokinetic dysarthria by two qualified speech-language pathologists experienced with the speech of individuals with PD. All speakers in the PD group had taken their prescribed medication and were at the optimal “on” status when they participated in the tasks. The speakers in the HC group had no diagnosis of Parkinson’s disease or other neurological disorder (self-reported) and no vocal pathology or other speech disorder (as judged by an experienced speech-language pathologist).

**Listeners.** Twenty university students (10 male and 10 female) ranging in age from 19 to 24 years ($M = 20.95$, $SD = 1.50$) were recruited as listeners. All listeners were native Cantonese speakers (speaking Cantonese as first language and have been educated in local schools in Hong Kong for at least 12 years) and reported normal hearing ability. They had limited exposure to individuals with Parkinson’s disease and to disordered speech. They were not students majoring in music nor had any professional vocal training in the past five years. Year 2 to Year 4 Speech and Hearing Sciences students were not recruited due to their exposure to disordered speech.
Stimuli. All stimuli were previously collected as part of the larger project (Kwan, in preparation). Six emotions were studied: happy, sad, anger, disgust, surprise, and neutral. These six emotions were included in reference to the study by Dara, Monetta, & Pell (2008). A total of 162 speech stimuli were collected (27 speakers × 6 emotions = 162 stimuli). Speakers were asked to produce the same sentence “出門口記得門窗同鎖門” (“Remember to close the window and lock the door when leaving home”), which is neutral in content, in the six different emotions. With the inclusion of a neutral category, listeners will be less likely to use exclusion rules which may result in artificially high rates of correct guesses (Scherer, 1986). It may be argued that the speakers were asked to produce the sentences with “portrayed” vocal emotions, therefore the speech materials collected were unnatural and may be different from spontaneous productions. However, it has been pointed out that naturally occurring vocal emotional productions are generally controlled by socio-cultural rules of emotional expression (Banse & Scherer, 1996), therefore portrayed vocal emotional expression is unlikely to be strongly different from those produced spontaneously.

Speech data collection was carried out in a quiet room with an ambient noise level of 30 – 34 dB SPL. The speakers first listened to a demonstration of vocal emotions produced by an actor through headphones (Sennheiser HD 212Pro) attached to a laptop computer (Apple MacBook Pro using iTunes, sound card S/PDIF Op-cal Digital Audio Output). Then they were asked to produce the same semantically neutral sentence in each of the 6 emotional tones in the order of neutral, sad, disgust, angry, happy, and surprised. Speech samples were recorded using an audio recorder (ZOOM H4n Handy Recorder with AKG MicroMic C520 vocal). A mouth-to-microphone distance of 10 cm was maintained during recordings.
Procedure. The perceptual tasks were carried out in a sound-attenuated booth with an average ambient noise level of 29.8 dB measured by a sound level meter. The stimuli were played to the listeners through headphones (Sennheiser HD428) attached to a laptop computer (eMachines KAV60). All five perceptual tasks were carried out using Adaptive Visual Analog Scales software (AVAS) (Marsh-Richard, Hatzis, Mathais, Venditt, & Dougherty, 2009) played by a laptop computer (Fujitsu LH530). There were 180 trials for each task (162 stimuli + 18 repeated trials for reliability measures = 180 trials). Task 1 was the identification of emotions. In this task, the listeners heard the stimuli which consisted of the six different emotional tones (happy, sad, surprised, disgust, angry, neutral). Following each trial, the listeners were asked to choose from an English multiple-choice list (Appendix A) of the six emotions shown on the computer screen to indicate which emotion best corresponded to the emotion which the speaker was trying to convey when they were producing the sentence. Written Chinese translation of the six emotions (開心, 傷心, 驚訝, 厭惡, 憤怒, 中立) was given to the listeners before the beginning of the tasks. Tasks 2 to 5 were perceptual ratings of four perceptual speech dimensions including loudness level, pitch height, pitch variability, and speech rate of the speech samples. A 20-cm visual-analog scale (VAS) was used for all of the four perceptual rating tasks. The left and right anchors for each perceptual speech dimension are summarized in Table 1. After each trial, the listener indicated his/her perceptual rating by marking with an arrow on the VAS (Appendix B).
Table 1

*Anchors on the Visual Analog Scale for ratings of the four Perceptual Dimensions*

<table>
<thead>
<tr>
<th>Task</th>
<th>Perceptual dimensions</th>
<th>Left anchor</th>
<th>Right anchor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Loudness level</td>
<td>Very Soft Voice</td>
<td>Very Loud Voice</td>
</tr>
<tr>
<td>3</td>
<td>Pitch height</td>
<td>Very Low Pitch</td>
<td>Very High Pitch</td>
</tr>
<tr>
<td>4</td>
<td>Pitch variability</td>
<td>Monopitch</td>
<td>Excess Pitch Variability</td>
</tr>
<tr>
<td>5</td>
<td>Speech rate</td>
<td>Very Slow</td>
<td>Very Fast</td>
</tr>
</tbody>
</table>

Note: Numerical rating generated by the AVAS Software is given in parenthesis.

Discrete visual analog scale (DVAS) (Appendix A) was used for Task 1 and VAS (Appendix B) was used for Tasks 2 to 5. Each task began with a detailed instruction and two practice trials. The listener was seated in a chair approximately 80-cm from the screen of the laptop computer. A mouse was given to the listener for selecting the responses. The stimuli were played manually by the investigator and a new trial was played when the listener had rated the previous trial. No limitation was imposed for the time taken to respond. Listeners were allowed to take a break whenever they feel fatigued.

Each listener first began with Task 1. The order of Tasks 2 to 5 was counterbalanced to control for possible order effects. The order of presentation of the 180 stimuli was randomized into five different orders. These five orders were allocated randomly to the five perceptual tasks for each listener. Each listener rated a total of 900 trials (180 stimuli x 5 perceptual tasks) and it took 1 hour 45 minutes to 2 hours for each listener to complete all the tasks. A HKD$20 gift coupon was given to each listener upon completion of all tasks.
Data Analysis. Identification accuracy of different emotions was calculated as the percentage of responses correctly identified by the listeners as the intended emotion. A mean percentage across all twenty listeners was calculated for each stimulus. The data output from Task 2 to Task 5 (using VAS) was a numerical rating ranging from 0 to 100 (0 = left anchor and 100 = right anchor). A mean rating across all twenty listeners was calculated for each stimulus. Five $2 \times 6$ two-way analysis of variance (ANOVA) (Speaker Group [PD, HC] × Intended Emotion [happy, sad, surprised, disgust, angry, neutral]) were done to examine the effects of the speaker group and the intended emotion on the identification accuracy and the ratings of the four perceptual dimensions. Post hoc comparisons were done to compare the perceptual ratings for the four perceptual dimensions. Independent-sample $t$-tests were used to find if there is any significant difference between the two speaker groups on the identification accuracy as well as the perceptual ratings of each intended emotion.

Reliability. Inter-rater reliability between the twenty listeners was measured using Intraclass Correlation Coefficient (ICC). ICC (2, 20) for the twenty listeners was .93 (95% confidence interval = .91-.94) for Task 1 (emotion identification), .98 (95% confidence interval = .97-.98) for Task 2 (loudness level), .96 (95% confidence interval = .95-.97) for Task 3 (pitch height), .95 (95% confidence interval = .93-.96) for Task 4 (pitch variation), and .96 (95% confidence interval = .95-.97) for Task 5 (speech rate). Intra-rater reliability for each listener was measured using the 3 speakers whose speech samples were repeated in each of the five perceptual tasks. Cohen’s Kappa was calculated for 10% of the listeners for Task 1. The mean intra-rater reliability for Task 1 was .73. Pearson’s $r$ was calculated for 10% of the listeners for Task 2 to 5. A mean Pearson’s $r$ was calculated for each of the four tasks. The mean intra-rater reliability was .91 for Task 2, .74 for Task 3, .73 for Task 4, and .77 for Task 5.
Results

Identification Accuracy

Figure 1 shows the mean identification accuracy (%) of different intended vocal emotions produced by the two speaker groups, PD and HC. It was found that for speech samples produced by the speakers with PD, the highest mean identification accuracy was found for intended emotion sad (74.50%), followed by surprised (69.25%), and the lowest mean identification accuracy was found for happy (22.25%). For the speech samples produced by the HC speakers, the highest mean identification accuracy was found for intended emotion angry (81.43%), followed by sad (80.00%), and the lowest mean identification accuracy was found for disgust (16.43%).

![Bar graph showing the mean identification accuracy (%) of different intended vocal emotions produced by the two speaker groups (PD and HC).](image)

The mean identification accuracy of different intended emotions ranged from 22.25% to 74.50% for PD group with mean overall accuracy of 47.75% ($SD = 33.74$) and it ranged from 16.43% to 80.00% for HC group with mean overall accuracy of 61.19% ($SD = 36.17$). For mean identification accuracy, the two-way ANOVAs (2 Speaker Group × 6 Intended Emotion) showed significant main effects of speaker group, $F(1, 150) = 7.82, p < .05$, with lower mean identification accuracy in vocal emotions produced by the PD group ($M = 47.75, SD = 33.74$)
than that produced by the HC group (\( M = 61.19, SD = 36.17 \)), and of intended emotion, \( F(5, 150) = 16.32, p < .001 \). There was significant interaction effect between speaker group and intended emotion, \( F(5, 150) = 2.54, p < .05 \). Independent-sample \( t \)-tests were performed to compare the mean identification accuracy (%) between the two speaker groups (PD and HC) for each intended emotion. Significant results were found between the PD group (\( M = 34.00, SD = 31.77 \)) and the HC group (\( M = 81.43, SD = 28.54 \)) for intended emotion angry, \( t(25) = 3.48, p < .005 \), suggesting there is significantly lower identification accuracy for the intended emotion angry produced by the PD group. The differences in mean identification accuracy (%) between the two speaker groups for all other intended emotions were not significant (\( p > .05 \)). The interaction between speaker group and intended emotion is shown in Figure 2.

\[ \text{Figure 2. Profile plot showing the main effects of speaker groups and that of intended emotion on the mean identification accuracy (%) and interaction effect between speaker group and intended emotion. Blue line represented HC group while Green line represented PD group.} \]

Figure 3 shows the percentages of different perceived emotions for the speech samples produced by the two speaker groups, PD and HC. The percentages for each intended emotion in
the speech samples were the same (i.e. 100% divided by six intended emotions = 16.67%). The speech samples produced by both the PD group and the HC group were most frequently perceived as neutral (36.04% for PD group and 26.55% for HC group) and least frequently perceived as happy (7.71% for PD group and 8.57% for HC group). The speech samples produced by the PD group appeared to be more frequently perceived as sad and neutral and much less frequently perceived as angry when compared with the HC group. The percentages appeared comparable for other perceived emotions (happy, surprised, and disgust).

Figure 2. Bar graph showing the percentages (%) of different perceived vocal emotions produced by the two speaker groups (PD and HC).

Perceptual Ratings

Two-way ANOVAs (2 Speaker Group × 6 Intended Emotion) were done for the listeners’ ratings of each perceptual dimension (loudness level, pitch height, pitch variability, and speech rate) on the VAS (from 0 to 100). For loudness level, there were significant main effects of speaker group, $F(1,150) = 36.69, p < .001$, where listeners perceived lower loudness level in vocal emotions produced by the PD group ($M = 44.35, SD = 15.96$) than those produced by the HC group ($M = 56.88, SD = 16.81$), and of intended emotion, $F(5,150) = 27.93, p < .001$, but there was no significant interaction effect between speaker group and intended emotion, $F(5,150)$
= .71, \( p = .62 \). For pitch height, there were significant main effects of speaker group, \( F(1,150) = 43.01, p < .001 \), where listeners perceived lower pitch height in vocal emotions produced by the PD group (\( M = 44.02, SD = 14.00 \)) than those produced by the HC group (\( M = 57.31, SD = 14.18 \)), and of intended emotion, \( F(5,150) = 16.42, p < .001 \), but there was no significant interaction between speakers group and intended emotion, \( F(5,150) = .43, p = .83 \). For pitch variability, there were significant main effects of speaker group, \( F(1,150) = 31.57, p < .001 \), where listeners perceived lower pitch variability in vocal emotions produced by the PD group (\( M = 40.29, SD = 16.23 \)) than those produced by the HC group (\( M = 50.45, SD = 16.05 \)), and of intended emotion, \( F(5,150) = 40.83, p < .001 \), but there was no significant interaction between speaker group and intended emotion, \( F(5,150) = .73, p = .61 \). Lastly, for speech rate, there were significant main effects of speaker group, \( F(1,150) = 4.74, p < .05 \), where listeners perceived slower speech rate in vocal emotions produced by the PD group (\( M = 46.71, SD = 15.73 \)) than those produced by the HC group (\( M = 51.02, SD = 13.98 \)), and of intended emotion, \( F(5,150) = 24.06, p < .001 \), but there was no significant interaction between speaker group and intended emotion, \( F(5,150) = .64, p = .67 \). Independent-sample \( t \)-tests were done to examine the effect of speaker group on the ratings of each perceptual dimension for each intended emotion. Post-hoc tests were done to compare the ratings of each perceptual dimension between different intended emotions. These results would be presented below according to each intended emotion.

**Happy.** Table 2 shows the listeners’ perceptual ratings on a VAS for the four perceptual dimensions of the speech stimuli with intended emotion happy produced by the two speaker groups (PD and HC). Significant results were found in the mean pitch height, \( t(25) = 2.58, p < .05 \), between the two speakers group when they intended to convey the happy emotion. It indicated that when the speakers intended to convey the emotion of happy, listeners perceived
the speech samples produced by PD speakers as having lower pitch height than those produced by the HC speakers. No significant differences were found for the other perceptual dimensions ($p > .05$).

Table 2

*Listeners’ perceptual dimension ratings for speech stimuli with intended emotion Happy produced by the two speaker groups (PD and HC)*

<table>
<thead>
<tr>
<th>Perceptual Dimensions</th>
<th>Mean ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PD</td>
</tr>
<tr>
<td>Loudness level</td>
<td>48.39 (12.90)</td>
</tr>
<tr>
<td>Pitch height</td>
<td>49.20 (12.34)</td>
</tr>
<tr>
<td>Pitch variability</td>
<td>46.77 (12.25)</td>
</tr>
<tr>
<td>Speech rate</td>
<td>55.44 (10.21)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

**Sad.** Table 3 shows the listeners’ perceptual ratings on a VAS for the four perceptual dimensions of the speech stimuli with intended emotion sad produced by the two speaker groups (PD and HC). Non-significant results ($p < .05$) were found for all four perceptual speech dimensions when comparing the perceptual ratings of speech samples between the two speaker groups when they intended to convey the sad emotion. However, for intended emotion sad, lowest mean ratings were found for all the four perceptual dimensions for both speaker groups when comparing to the other intended emotions. Post hoc comparisons showed that intended emotion sad had significantly lower loudness level, lower pitch height, and smaller pitch variability than all other intended emotions except neutral (all $p < .001$), and significantly slower speech rate than all other intended emotions (all $p < .05$). It indicated that listeners perceive the
speech samples with intended emotion sad as having the softest voice, the lowest pitch height, the smallest pitch variability, and the slowest speech rate when compared to all other intended emotions (except neutral).

Table 3

*Listeners’ perceptual dimension ratings for speech stimuli with intended motion Sad produced by the two speaker groups (PD and HC)*

<table>
<thead>
<tr>
<th>Perceptual Dimensions</th>
<th>Mean ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PD</td>
</tr>
<tr>
<td>Loudness level</td>
<td>30.20 (12.73)</td>
</tr>
<tr>
<td>Pitch height</td>
<td>34.15 (11.67)</td>
</tr>
<tr>
<td>Pitch variability</td>
<td>24.91 (9.92)</td>
</tr>
<tr>
<td>Speech rate</td>
<td>29.23 (10.98)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

**Surprised.** Table 4 shows the listeners’ perceptual ratings on a VAS for the four perceptual dimensions of the speech stimuli with intended emotion surprised produced by the two speaker groups (PD and HC). Significant results were found for loudness level, \( t(25) = 2.07, p < .05 \), and pitch height, \( t(25) = 2.20, p < .05 \), when comparing the mean perceptual ratings of speech samples produced by the two speaker groups when they intended to convey the surprised emotion. It indicated that when conveying the surprised emotion, speech samples produced by PD speakers were perceived as having softer voice and lower pitch height than those produced by HC speakers. Moreover, speech samples with intended emotion surprised have the highest mean pitch height among all vocal emotions produced by PD speakers. The speech samples with intended emotion surprised have the highest mean pitch variability among all vocal emotions
produced by both speaker groups. Post hoc comparisons showed that intended emotion surprised had significantly larger pitch variability than all other intended emotions (all $p \leq .001$).

Although intended emotion surprised had higher pitch height than all other emotions, post hoc comparisons showed that it was only significant when compared to intended emotions sad, disgust, and neutral ($p < .001$) but was not significant when compared to happy and angry ($p > .05$).

*Table 4*

*Listeners’ perceptual dimension ratings for speech stimuli with intended emotion Surprised produced by the two speaker groups (PD and HC)*

<table>
<thead>
<tr>
<th>Perceptual Dimensions</th>
<th>Mean ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PD</td>
</tr>
<tr>
<td>Loudness level</td>
<td>47.26 (11.55)</td>
</tr>
<tr>
<td>Pitch height</td>
<td>54.53 (14.13)</td>
</tr>
<tr>
<td>Pitch variability</td>
<td>59.14 (11.71)</td>
</tr>
<tr>
<td>Speech rate</td>
<td>56.60 (10.37)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

**Disgust.** Table 5 showed the listeners’ perceptual ratings on a VAS for the four perceptual dimensions of the speech stimuli with intended emotion disgust produced by the two speaker groups (PD and HC). Significant results were found for the perceptual dimensions of loudness level, $t(25) = 3.11$, $p < .05$, pitch height, $t(25) = 4.22$, $p < .005$, and pitch variability, $t(25) = 3.95$, $p < .005$, when comparing the speech samples with intended emotion disgust produced by the two speaker groups. It indicated that when speakers intended to convey the disgust emotion, listeners perceived the speech samples produced by PD speakers as having
softer voice, lower pitch height, and smaller pitch variability when compared to those produced by HC speakers.

Table 5

*Listeners’ perceptual dimension ratings for speech stimuli with intended emotion Disgust produced by the two speaker groups (PD and HC)*

<table>
<thead>
<tr>
<th>Perceptual Dimensions</th>
<th>Mean ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PD</td>
</tr>
<tr>
<td>Loudness level</td>
<td>43.88 (12.26)</td>
</tr>
<tr>
<td>Pitch height</td>
<td>39.78 (8.75)</td>
</tr>
<tr>
<td>Pitch variability</td>
<td>36.10 (10.11)</td>
</tr>
<tr>
<td>Speech rate</td>
<td>44.92 (14.00)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

**Angry.** Table 6 showed the listeners’ perceptual ratings on a VAS for the four perceptual dimensions of the speech stimuli with intended emotion angry produced by the two speaker groups (PD and HC). Significant results were found for the perceptual dimensions of loudness level, \( t(25) = 4.15, p < .005 \), pitch height, \( t(25) = 3.26, p < .005 \), and pitch variability, \( t(25) = 2.10, p < .05 \), when comparing the speech samples with intended emotion angry produced by the two speaker groups. It indicated that when speakers intended to convey the angry emotion, listeners perceived the speech samples produced by PD speakers as having softer voice, lower pitch height, and smaller pitch variability than those produced by HC speakers. Moreover, speech samples with intended emotion angry have the highest mean loudness level and highest mean speech rate among all intended emotions for both speaker groups. Post hoc comparisons showed that intended emotion angry had significantly higher loudness level than all other
intended emotions (all \( p < .001 \)). Although intended emotion angry had faster speech rate than all other intended emotions, post hoc comparisons showed that it was only significant when compared to intended emotions sad, disgust, and neutral (\( p < .001 \)) but not significant when compared to intended emotions happy and surprised (\( p > .05 \)).

Table 6

*Listeners’ perceptual dimension ratings for speech stimuli with intended emotion Angry produced by the two speaker groups (PD and HC)*

<table>
<thead>
<tr>
<th>Perceptual Dimensions</th>
<th>Mean ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PD</td>
</tr>
<tr>
<td>Loudness level</td>
<td>63.04 (11.40)</td>
</tr>
<tr>
<td>Pitch height</td>
<td>51.63 (13.44)</td>
</tr>
<tr>
<td>Pitch variability</td>
<td>49.32 (11.27)</td>
</tr>
<tr>
<td>Speech rate</td>
<td>57.17 (12.64)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

**Neutral.** Table 7 showed the listeners’ perceptual ratings on a VAS for the four perceptual dimensions of the speech stimuli with intended emotion neutral produced by the two speaker groups (PD and HC). Significant results were found for perceptual dimensions of loudness level, \( t(25) = 2.25, p < .05 \), pitch height, \( t(25) = 3.28, p < .005 \), and pitch variability, \( t(25) = 2.63, p < .05 \), when comparing the speech samples with intended emotion neutral produced by the two speaker groups. The results indicated that when the speakers intended to convey the emotion of neutral, listeners perceived the speech samples produced by PD speakers as having softer voice, lower pitch height, and smaller pitch variability than those produced by the HC speakers. Post hoc comparisons showed that intended emotion neutral had significantly
lower loudness level and smaller pitch variability than all other intended emotions except sad ($p < .001$). It indicated that listeners perceive the speech samples with intended emotion neutral as having the softest voice and the smallest pitch variability when compared to all other intended emotions except sad.

Table 7

*Listeners’ perceptual dimension ratings for speech stimuli with intended emotion Neutral produced by the two speaker groups (PD and HC)*

<table>
<thead>
<tr>
<th>Perceptual Dimensions</th>
<th>Mean ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PD</td>
</tr>
<tr>
<td>Loudness level</td>
<td>33.32 (11.05)</td>
</tr>
<tr>
<td>Pitch height</td>
<td>34.85 (8.14)</td>
</tr>
<tr>
<td>Pitch variability</td>
<td>25.48 (5.71)</td>
</tr>
<tr>
<td>Speech rate</td>
<td>36.89 (11.69)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

**Discussion**

This study aimed to investigate if there was any difference between the identification accuracy of vocal emotions produced by speakers with PD and those produced by control speakers (HC). It also aimed to examine if there were differences in the perceptual speech dimensions (loudness level, pitch height, pitch variability, and speech rate) in the speech samples produced by speakers with PD and HC, and how the differences may be used to explain the differences in the identification accuracy of vocal emotion between the two speaker groups. Results revealed that the mean overall identification accuracy of vocal emotions produced by the PD group was significantly lower than that of the vocal emotions produced by the matched
healthy speakers. In particular, the speech samples with intended emotion angry produced by the PD group had significantly lower identification accuracy than those produced by the HC group. For perceptual ratings, speech samples produced by the PD group had significantly softer voice, lower pitch, lower pitch variability (more monotonous), but comparable speech rate when compared with speech samples produced by the HC group in general. This is consistent with the perceptual profile for Cantonese speakers with PD established by Whitehill, Ma, and Lee (2003). In their study, Whitehill et al. (2003) found that monopitch and monoloudness were within the six most deviant speech dimensions for the group of speakers with PD. Whitehill et al. (2003) also found that unusually rapid or unusually slow speech rate were only found in a small percentage of PD speakers. Comparable speech rate between the two speaker groups in the current study is also consistent with the previous study.

**Identification of emotions produced by different speaker groups**

It was found that for speech samples produced by the speakers with PD, the highest mean identification accuracy was found for the intended emotion sad (74.50%), followed by surprised (69.25%), and the lowest mean identification accuracy was found for happy (22.25%). For the speech samples produced by the HC speakers, the highest mean identification accuracy was found for the intended emotion angry (81.43%), followed by sad (80.00%), and the lowest mean identification accuracy was found for disgust (16.43%). The identification accuracy of vocal emotions produced by HC speakers was consistent with previous studies on recognition of vocal emotions conducted in different countries with different languages as reviewed by Pittam & Scherer (1993). The results from these studies concluded that sadness and anger are best recognized and disgust is most poorly recognized. However, the identification accuracy of vocal emotions produced by speakers with PD was not consistent with these previous studies,
suggesting differences in the expression of vocal emotions between speakers with PD and healthy controls, thus leading to lower identification accuracy for vocal emotions produced by PD speakers.

The results showed that the mean identification accuracy of vocal emotions was significantly lower for the PD group than that for the HC group, suggesting that vocal emotions produced by speakers with PD were less accurately perceived as compared to the vocal emotions produced by healthy controls. The lower identification accuracy for vocal emotions may be contributed by both the disease-specific motor impairments as well as the disturbances of emotional processing found in patients with PD (Schröder, 2010). Monopitch, low pitch and monoloudness were some of the most deviant speech characteristics encountered in hypokinetic dysarthria in Cantonese speakers and speakers of other languages (Darley et al., 1969a; Whitehill et al., 2003). The motor impairments associated with PD, such as tremor, rigidity and bradykinesia are associated with the deviant speech characteristics found in the speech of patients with hypokinetic dysarthria (Duffy, 2005). On the other hand, the experiment carried out by Möbes and colleagues (2008) examined the role of disturbances of emotional processing in the expression of emotional prosody by patients with PD. It was found that for patients with mild to moderate PD, their pitch and intensity modulation differed significantly from controls when requested to produce emotional prosody prompted by a visual cue. However, the production was nearly normal when they were requested to imitate emotional speech pre-recorded by a professional speaker. This suggested that the difficulty in modulating pitch and intensity was not only limited by the phonation or speech motor capacity of the patients with PD, but also contributed by disturbances in emotional processing. A number of previous studies concluded that deficits in recognizing emotions from vocal expressions were common in patients
with focal lesions affecting the basal ganglia when compared to other lesion sites (Cancelliere & Kertesz, 1990; Weddell, 1994). In the early stages of PD, the damage is relatively confined to the basal ganglia (Alexander, DeLong, & Strick, 1986). This suggested that the speakers with PD in this study may also have deficit in recognizing vocal emotions due to damage in basal ganglia because of PD, which is consistent with previous studies (Breitenstein et al., 2001; Pell & Leonard, 2003; Yip, Lee, Ho, Tsang, & Li, 2003). As speech production relies on auditory feedback for determining whether the perceptual speech characteristics are salient enough for listeners, deficits in recognizing own production of vocal emotions in speakers with PD may explain their reduced ability to express vocal emotions when compared to healthy controls.

Pell et al. (2006) found that Parkinson’s disease was associated with flat emotional and non-emotional speech and that listeners experienced difficulties recognizing most of the emotions, especially anger and disgust, expressed by people with Parkinson’s disease with the exception of sadness. The results from Cantonese speakers with PD in the current study are not entirely consistent with the results obtained by Pell and colleagues (2006) on English speakers with the same six intended emotions. In our study, although anger was also found to be significantly difficult for listeners to recognize in the speech of speakers with PD, identification of other emotions were not significantly different when compared with the identification accuracy of the healthy controls. However, the results were consistent with other studies (Caektebeke, Jennekens-Schinkel, Van der Linden, Buruma, & Roos, 1991; Penner, Miller, Hertrich, Ackermann, & Schumm, 2001) indicating that speakers with PD were poor at conveying anger through prosody.

Further investigation on the listeners’ perception revealed that speech samples produced by the PD group were more frequently perceived as sad and neutral when compared to those
produced by the HC group. This result is in consistent with studies showing that speech samples produced by patients with PD were often misinterpreted as sounding sad or neutral (Pell et al., 2006; Schröder, 2010). This can be explained by the prosodic characteristics of speech produced by speakers with PD (monopitch, reduced pitch, monoloudness) which converge with the parameters for identifying sad and neutral speech (Pell et al., 2006).

**Perceptual characteristics of different intended emotions**

Among all six intended emotions, speech samples with intended emotion angry produced by the PD group were found to have significantly lower identification accuracy than those produced by the HC group. Results for the perceptual ratings revealed that for the intended emotion angry, speech samples produced by the PD group were found to have significantly lower loudness level, lower pitch height, lower pitch variability, but comparable speech rate when compared with the speech samples produced by the HC group. From previous studies, it was found that vocal emotion of anger produced by non-impaired speakers was generally characterized by an increase in mean fundamental frequency (F₀) and mean intensity (Pittam & Scherer, 1993). Some studies also included increases in F₀ variability and increases in rate of articulation as the acoustic signal for expression of Anger (Pittam & Scherer, 1993). It is generally accepted that F₀ is the acoustic variable underlying pitch and intensity is the acoustic variable underlying loudness (Pittam & Scherer, 1993). The lower pitch height (which is responsible for reduced mean F₀), lower pitch variability (which is responsible for reduced F₀ variability), and lower loudness level (which is responsible for reduced intensity) in the speech samples with intended emotion angry produced by speakers with PD hindered the dominant acoustic characteristics for expression of anger, therefore leading to the lower identification accuracy when compared with the HC group.
For intended emotion happy, sad, surprised, disgust, and neutral, results revealed that there were no significant differences in the identification accuracy between the two speaker groups PD and HC. For happy and sad, it may be explained by their lower F<sub>0</sub> variability, lower mean intensity, and lower intensity variability when compared to angry (Scherer, 2003), which may be more easily attained for patients with PD due to their limited motor capacity. There is limited information on the acoustic features essential for the accurate perception of the remaining emotions.

Vocal emotions produced by PD speakers in this study were most frequently perceived as neutral and sad. From the results of this study, it was found that intended emotion of sad and neutral had similar perceptual characteristics – listeners perceived neutral and sad with the softest voice and smallest pitch variability when compared to other intended emotions (except sad or neutral). It may be concluded that the lower loudness level and smaller pitch variability in the speech of PD speakers when compared to HC speakers contributed to the high frequency of their vocal emotions being perceived as neutral and sad.

**Implications and limitations**

This study provides empirical evidence for the argument that vocal emotional communication is impaired in Cantonese speakers with hypokinetic dysarthria associated with PD. It also highlighted the communicative barriers encountered by those affected by the disease when expressing vocal emotions, especially anger. When speakers with PD are unable to express vocal emotions, treatment should focus on increasing the loudness, pitch height, and pitch variability of speech. Moreover, the perceptual speech dimensions should be “fine-tuned” according to the characteristics required for the accurate perception of an emotion. For example, angry is associated with highest loudness level and faster speech rate than sad, disgust, and
neutral, while sad and neutral are associated with lowest loudness level and smallest pitch variability.

Future studies may attempt to address the following concerns: 1) Listeners’ perceptual judgment and ratings were based on simulated vocal emotions which may be different from spontaneous productions in daily life; 2) clear definition of emotions, e.g. whether anger represents “hot” (rage) or “cold” anger (irritation), were not provided to the speakers. This may lead to listeners having different perceptual judgments; and 3) the study did not include speakers with a wide range of severity of hypokinetic dysarthria, so that comparison between these groups was unable to be performed, and the results may not be generalizable to all Cantonese speakers with PD.

Conclusions

Few studies have been carried out to investigate the expression of vocal emotions by English speakers with PD (e.g., Pell et al., 2006; Schröder, 2010). The current study was the first to investigate the perception of vocal emotions by Cantonese speakers with PD. Vocal emotions produced by Cantonese speakers with PD were in general more difficult to recognize than those produced by healthy controls, particularly when expressing anger. The speech produced by speakers with PD generally was perceived as having a lower loudness level, a lower pitch height, and being more monotonous when compared with the speech of matched controls. These deviant speech characteristics were hypothesized to be related to limited motor capacity as well as a possible deficit in processing and recognizing vocal emotions.
Acknowledgments

I would like to express my sincere gratitude to Professor Tara Whitehill and Mrs Lorinda Kwan for their supervision and guidance throughout this dissertation project. I would also like to thank Mrs Lorinda Kwan for kindly providing the speech samples previously collected under her larger project as well as the demographic data of the speakers. I am also deeply thankful for the help from the twenty listeners who spent much time and effort in rating the speech samples.
References


Appendix A

Print screen image of the English multiple-choice list for Task 1 (emotion identification)
Appendix B

Print screen images of the visual analog scale for perceptual ratings of Task 2 to 5 (loudness level, pitch height, pitch variation, speech rate)