Exploratory Data Analysis on Nuclei in Cantonese Dysarthric Speech

Ka Ho Wong [1], *Hoi Kiu Kristy Mok* [1] *and Helen Meng* [1,2]

[1] Human−Computer Communications Laboratory,
Department of Systems Engineering and Engineering Management,
[2] Stanley Ho Big Data Decision Analytics Research Centre,
The Chinese University of Hong Kong, Hong Kong SAR, China

khwong@se.cuhk.edu.hk, hkmok@se.cuhk.edu.hk, hmmeng@se.cuhk.edu.hk

Abstract

Phone substitutions, distortions, deletions, and insertions are some of the main problems in dysarthric speech. This work aims to explore the articulatory error patterns in dysarthric speech, which provides insights for the improvement of automatic dysarthric speech analysis technologies. A set of dysarthric speech is collected and phonetically transcribed manually by different transcribers. Transcriptions are mapped into distinctive feature values. Error rates for each value are extracted. Substitutions and distortions are found to be the major errors in dysarthric speech. Their error patterns are analyzed in this paper. The analysis may provide guidance for labelling dysarthric speech errors, which will be useful to future development of technologies to achieve automated analysis of dysarthric speech.

Index Terms: exploratory data analysis, dysarthric speech

1. Introduction

Dysarthria is a motor speech disorder as due to neurological impairments from trauma or illnesses. These may include head injuries, stroke, Parkinson's disease and brain tumors. Some of these illnesses correlate with old age. With our society rapidly aging, dysarthria is of increasing concern.

To support dysarthric patients, an extensive amount of assessments, treatment and intensive oral training is necessary. This leads to substantial demands on time and effort from highly trained professionals, especially speech therapists. Automatic technologies on intelligibility assessments [1], error analysis and speech recognition [2] may contribute to augmenting the capacity of speech therapists to support and help a large number of patients. In order to develop automatic technologies, we need a reasonable amount of labeled dysarthric speech.

People with dysarthria have difficulties with speech motor control, causing them to experience articulation problems. Various aspects of speech, including speaking rate, voice quality and pitch, may all be affected. Due to their limited control on their tongue, lip and jaw muscles, dysarthria patients have trouble with producing speech in a normal manner. Phone substitutions, deletions, insertions and distortions are some of the main problems that can be observed in dysarthric speech. The objective of this paper is to analyze the deviations of dysarthric speech as compared to normal speech. Investigating the error patterns helps inform the approaches with which we develop automatic technologies for processing dysarthric speech.

This paper is organized as follows: a brief description of CU DYS, a Cantonese dysarthric corpus, is presented in Section 2. Section 3 describes the error trends in nuclei. Finally, Section 4 concludes this paper and presents directions for future work.

2. CU DYS – A Cantonese Dysarthric Speech Corpus

2.1. Corpus Description

CU DYS is a Cantonese dysarthric speech corpus collected by The Chinese University of Hong Kong. It includes speech of 27 subjects with dysarthria (resulting from 16 patients with spinocerebellar ataxia and 11 with cerebral palsy) and 14 control subjects (i.e. non-dysarthric subjects), all native speakers of Hong Kong Cantonese. CU DYS include 15 tasks. Tasks 1 to 3 are word-level, sentence-level, and paragraph-level stimuli respectively. Tasks 4 to 14 cover a range of stimuli for Cantonese which are adapted from the Frenchary Dysarthria Assessment (FDA-2) [3], is based on English. The stimuli assess muscle control in different speech tasks, including reflexes, palatal control, and tongue muscle control. Examples of stimuli are shown in Table 1. This paper will focus on exploratory data analysis of Task 1. In the task,

Table 1: Examples in Tasks 1 to 14 in the assessment of Cantonese dysarthric speech, adapted from the FDA-2 which is based on English.

Task	Stimuli	Syllables	Translation
1	巴士	/baa/ /si/	Bus
2	我好熱	/ngo/ /hou/ /jit/	I am hot
3	巴巴拉史翠珊是我很喜歡的一位美國女明星。…	/baa/ /baa/ /laai/ /si/ /ceoi/ /saan/ /si/ /ngo/ /han/ /hei/ /fun/ /dik/ /jat/ /wai/ /mei/ /gwok/ /neoi/ /ming/ /seng/	Barbra Streisand is my favorite American actress….

Task	Instruction	Translation
4-14	深呼吸，然後說「烏衣」」十次。要用最快的速度。盡量誇張嘴型。	Take a deep breath, then said, "oo ee" ten times. Use the fastest speed. Try exaggerated lips.

subjects are asked to read a total of 61 words ranging from one to three syllables. 49 stimuli were selected from the Hong Kong Cantonese Articulation Test (HKCAT) [4]. 12 words were added to the task to cover all initials and finals in Cantonese. A final is composed of a nucleus and an optional coda. Some nuclei (e.g. /aa/) can be joined with various codas (e.g. /ng/) to form a final (e.g. /aang/), such as /caang/ "橙" orange. Others nuclei have constrained combinations. For example, /oe/ can only preceed /ng/ (as in /koeng/ "強" strong) and /k/ (as in /koek/ "卻" step back). Figure 1 shows the occurrences of nuclei in Task 1. As Task 1 consists only of single words, the variability due to prosody is reduced. Further details of CU DYS can be found in [5].

2.2. Phonetic Transcription Process

Transcribers are given recordings of dysarthric subjects in sets of 61 utterances, played in randomized order. Each set is transcribed by two independent transcribers, i.e. each utterance in each set will have two transcriptions. All utterances are phonetically transcribed manually according to the Jyutping labelling system [6]. The Jyutping System is used as it is designed for Cantonese sounds and uses alphanumeric characters for transcription. Phonetic transcriptions are used rather than character transcriptions to capture more detailed sound differences of dysarthric speech.

Transcriptions and canonical Jyutping labels are manually and phonetically aligned, as described in [5], to be compatible with the fixed structure of Cantonese syllables. The two sets of transcriptions for each dysarthric subject are then compared with the canonical labels and also between transcribers. Transcriptions are considered "consistent" if both transcribers provide the same labels (regardless of whether they match the canonical labels), and "inconsistent" otherwise.

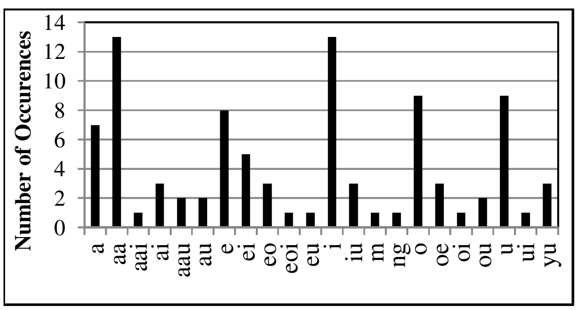


Figure 1: The occurrence of nuclei in Task 1, labelled by Jyupting.

3. Analysis of Nuclei Errors

Syllable nuclei in Cantonese include monophthongs (single vowels) and diphthongs (combination of two vowels). Each Cantonese syllable modeled as the combination of one and only one nucleus. In the analysis of error rates and consistency between transcribers, it was observed that most transcribed phonemes are consistent (80.5%). Consistent transcribed phonemes, which are same as the canonical nuclei, have 72.0%. The remainder consists of four main types of errors, namely substitutions, distortions, deletions and insertions, as illustrated in Table 2. An error is considered substitution when the two transcriptions are consistent but are different from the canonical labels. Distortions refer to transcriptions that are inconsistent. Deletions and insertions refer to nuclei being deleted or inserted respectively. An example of an insertion that we have observed is where /n**aa**m/ ("男", meaning boy) is transcribed as /n **aa au/**. The extra nucleus /au/ is considered as an insertion.

Among the four types, distortions cover 68.8% of the errors. Substitution error rates are also prominent at 30.0%. Noting the high proportions of these two errors, we focus our analysis on both distorted and substituted errors.

3.1.1. Mapping distorted nuclei into distinctive features

Distorted nuclei are difficult to describe by a well-defined phonetic symbol. The two transcriptions of a distorted nucleus can include a match and a mismatch with the canonical nucleus. Figure 2 illustrates an example. /aa/ in the canonical

Table 2: The distribution of nuclei errors

Transcription		Count	Total
Consistent (80.5%)	Match	1789 (72.0%)	2493
	Substitution	211 (8.5%)	
	Deletion	0	
	Insertion	0	
Inconsistent (19.5%)	Distortion (exclude insertion)	484 (19.5%)	
Insertion		9	

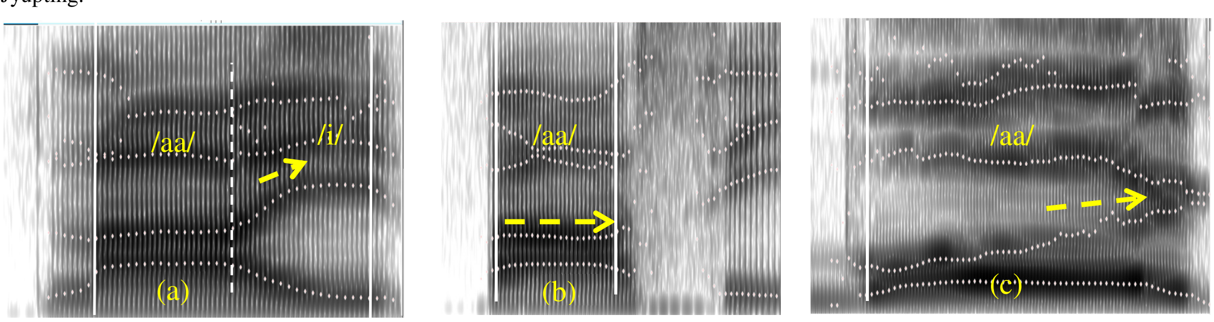
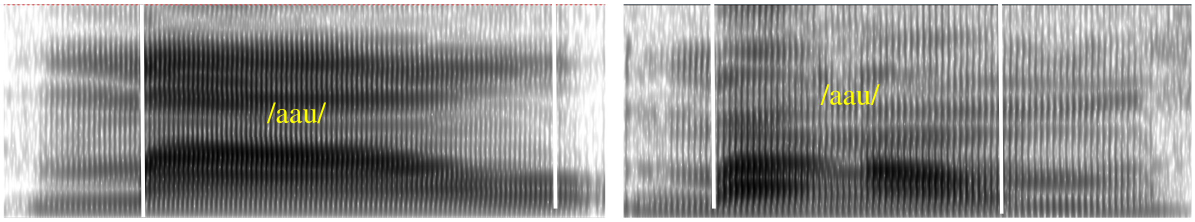


Figure 2: Spectrograms of a distorted nucleus. The dotted line shows the tracked formant [8]. The white vertical lines show the nucleus boundaries. The arrows indicate the trend of formant trajectory.
(a) **/laai/** ("拉" pull) from a control subject
(b) **/baa/** ("巴" bus) from a control subject
(c) **/baa/** from a dysarthric subject, where the nucleus is perceived by two transcribers to contain /aai/ and /aa/ respectively.



(a) /b/

Figure 3:

(a) /m**aau**/ (“貓” cat) from a control subject

(b) /m**aau**/ (“貓” cat) from a dysarthric subject, but /aau/ is transcribed as /aa/ and /a/

syllable /b**aa**/ (“巴” bus) was labeled by our transcriber as /aa/ (match) and /aai/ (substitution). Another possibility of a distorted nucleus, where the two transcriptions different, both do not match with the canonical nucleus (Figure 3).

To understanding the underlying articulation, we map distorted phones into distinctive feature vectors. Distinctive features (DFs) are a set of binary features to mark the distinction between two phonemes in a language [7]. There are 21 DFs in the set. All phones are mapped into DF vectors by table lookup. Four values are possible for each DF: positive “+”, negative “–”, irrelevant “x” or unspecified “/”. Irrelevant refers to a DF that does not contribute to producing the phone, whereas unspecified means that the DF does not influence the recognition of the phone.

As not all DFs are applicable to Cantonese nuclei, relevant features are selected for use in this analysis. These include [SYLLABIC], [ANTERIOR], [LABIAL], [HIGH], [LOW], [FRONT], [BACK], [ROUNDED], [CONTINUANT], [TENSE] and [VELAR]. Distorted nuclei are mapped accordingly into DFs for analysis. For diphthongs, two DF vectors are used to represent the transition between two vowels (Figure 4b). To maintain consistency comparison, monophthongs are also represented by two DFs (Figure 4a) where the two DF vectors are the same, indicating no change in articulation within the production of a monophthong. All canonical and transcribed nuclei are also mapped into DF vectors as in Figure 4.

3.1.2. *DF Analysis*

Canonical DF values and transcription DF values were compared. The comparison relationship is shown in Table 3. Only positive and negative values are considered.

- Nuclei Substitutions

For the phone substitution errors shown in Table 2, we analyze the errors of the corresponding DF values, which are shown in Table 4. As can be seen, most DF achieves a matching rate of over 60%, with the exception of [ANTERIOR] [- → +] (mostly from the prompts of /ng/ “五” five to /m/ “唔” um), [VELAR] [+ → -] (also /ng/ “五” five to /m/ “唔” um) and [TENSE] [- → +] (/f**a**n/ “分” score → /f**aa**n/ “反” opposite), [HIGH] [+ → -]

DF	/aa/ DF1	/aa/ DF2	/aau/ DF1	/aau/ DF2
SYLLABIC	+	+	+	+
ANTERIOR	x	x	x	x
LABIAL	-	-	-	+
HIGH	-	-	-	+
LOW	+	+	+	-
FRONT	-	-	-	-
BACK	-	-	-	-
ROUNDED	-	-	-	+
CONTINUANT	+	+	+	+
TENSE	+	+	+	+
VELAR	x	x	x	x

(a) A monophthong (b) A diphthong

Figure 4:DF vectors for a monophthong and a diphthong. Monophthongs are also represented by two DFs like diphthongs to maintain consistency comparison, especially when comparing a canonical diphthong (2 DFs) and a transcribed monophthong (1 DF).

Table 4: The error distribution of distinctive features in nuclei substitution.

DF	Match	DF value substitution					
		+ → -	+ → /	+ → x	- → +	- → /	- → x
[SYLLABIC]	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
[ANTERIOR]	0.0%	9.5%	0.0%	0.0%	90.5%	0.0%	0.0%
[LABIAL]	64.5%	19.4%	0.0%	0.0%	16.1%	0.0%	0.0%
[HIGH]	58.4%	20.1%	9.1%	0.0%	12.4%	0.0%	0.0%
[LOW]	68.9%	8.4%	0.0%	0.0%	13.6%	9.1%	0.0%
[FRONT]	76.1%	6.0%	0.0%	0.0%	8.9%	9.1%	0.0%
[BACK]	69.9%	11.7%	9.1%	0.0%	9.3%	0.0%	0.0%
[ROUNDED]	74.4%	18.5%	0.0%	0.0%	7.1%	0.0%	0.0%
[CONTINUANT]	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
[TENSE]	49.2%	20.3%	0.0%	0.0%	30.5%	0.0%	0.0%
[VELAR]	0.0%	90.5%	0.0%	0.0%	9.5%	0.0%	0.0%

Table 5: The error distribution of distinctive features in nuclei distortion, where one transcription equals the canonical nuclei .

DF	Match	DF value substitution					
		+ → -	+ → /	+ → x	- → +	- → /	- → x
[SYLLABIC]	99.7%	0.0%	0.0%	0.3%	0.0%	0.0%	0.0%
[ANTERIOR]	0.0%	45.5%	0.0%	0.0%	54.5%	0.0%	0.0%
[LABIAL]	66.9%	14.1%	0.0%	0.0%	18.8%	0.0%	0.3%
[HIGH]	69.9%	11.3%	1.6%	0.3%	16.9%	0.0%	0.0%
[LOW]	70.8%	16.2%	0.0%	0.0%	11.1%	1.6%	0.3%
[FRONT]	79.2%	7.8%	0.0%	0.3%	11.2%	1.6%	0.0%
[BACK]	77.0%	8.8%	1.6%	0.0%	12.3%	0.0%	0.3%
[ROUNDED]	69.8%	12.8%	0.0%	0.0%	17.2%	0.0%	0.3%
[CONTINUANT]	99.7%	0.0%	0.0%	0.3%	0.0%	0.0%	0.0%
[TENSE]	57.8%	19.3%	0.0%	0.0%	22.7%	0.0%	0.3%
[VELAR]	0.0%	54.5%	0.0%	0.0%	45.5%	0.0%	0.0%

Table 6: The error distribution of distinctive features in nuclei distortion, where both transcriptions are different from canonical nuclei.

DF	Match	DF value substitution					
		+ → -	+ → /	+ → x	- → +	- → /	- → x
[SYLLABIC]	99.5%	0.0%	0.0%	1.0%	0.0%	0.0%	0.0%
[ANTERIOR]	NIL	NIL	NIL	NIL	NIL	NIL	NIL
[LABIAL]	57.5%	27.5%	0.0%	0.0%	14.5%	0.0%	0.5%
[HIGH]	61.0%	22.5%	0.5%	0.0%	15.5%	0.0%	0.5%
[LOW]	57.3%	16.5%	0.0%	0.0%	25.3%	0.5%	0.5%
[FRONT]	72.8%	13.0%	0.5%	0.5%	13.3%	0.0%	0.0%
[BACK]	70.0%	15.5%	0.0%	0.0%	13.5%	0.5%	0.5%
[ROUNDED]	57.0%	28.0%	0.0%	0.0%	14.5%	0.0%	0.5%
[CONTINUANT]	99.0%	0.5%	0.0%	0.5%	0.0%	0.0%	0.0%
[TENSE]	46.5%	19.5%	0.0%	0.5%	33.0%	0.0%	0.5%
[VELAR]	NIL	NIL	NIL	NIL	NIL	NIL	NIL

(/s**yu**/ “書” book → /s**e**/ “些” some). These suggest that intervention strategies can focus on these

- Nuclei Distortions

For the distorted nuclei, we also analyzed their DF errors. There are two possible cases: the first case is that the canonical nucleus is equal to one of the transcriptions, and hence we compare the DF vectors between the canonical and inconsistent labels. The result is shown in Table 5. The error distribution is similar to Table 4 except for [ANTERIOR] and [VELAR]. In Table 4, most [ANTERIOR] are [- → +] and [VELAR are [+ → -]. Many /ng/ are substituted into /m/. In Table 5, both /ng/ and /m/ can be distorted to another label (i.e. /ng/ → /m/ and /m/ → /ng/).

Another type of distortion case is where both transcriptions are different from the canonical nuclei. These may imply more serious distortion for the canonical label. We compare the canonical DF vector and with both distorted DF vectors. The results are shown in Table 6. First, no [ANTERIOR] and [VELAR] errors are found. /m/ is mainly confused with /ng/ and vice versa. Second, the matching rate of most DFs are generally lower when compare Tables 4 and 5 but [ROUNDED] is dropped (74.4% in Table 4 and 69.8% in Table 5), especially [ROUNDED] [+ → -] (/c**oe**ng/ “窗” window → /c**eo**n/ “春” spring) .

In general, [SYLLABIC] and [CONTINUANT] show the highest matching rate. All nuclei are mainly confused with other nuclei but rarely a consonant or even a syllabic consonant (/ng/ and /m/).

4. Conclusions and Future Work

In this study, we have analyzed a set of nuclei and explored the error patterns of their production in dysarthric speech. Syllabic nuclei are likely to be distorted in dysarthric speech. The match rate of [ROUNDED] is dropped much from 74.4% in substituted phones to 57.0% in distorted phones (Table 6). [TENSE] is one of the most distorted features (match rates: 46.5% to 57.8%), causing confusion between /aa/ and /a/.

In the future, the dominant distortions from this analysis will be targeted for the improvement of speech technologies for dysarthric speech.

5. Acknowledgements

This project is partially sponsored by a grant from the Hong Kong SAR Government General Research Fund (reference no. GRF415513).

6. References

[1] Myung Jong Kim, Younggwan Kim, and Hoirin Kim, "Automatic Intelligibility Assessment of Dysarthric Speech Suing Phonologically-Structured Spare Linear Model," *Audio, Speech, and Language Processing*, vol. 23, no. 4, April 2015.

[2] Toru Nakashika et al., "Dysarthric Speech Recognition Using a Convolutive Bottlenect Network," in *International Conference on Signal Processing*, HangZhou, China, 2014.

[3] P. Enderby, *Frenchay Dysarthria Assessment-2*. San Diego College Hill Press, 2008.

[4] P. S. P. Cheung, A. Ng, and C. K. S. To, "Hong Kong Cantonese Articulation Test," Hong Kong, 2006.

[5] Ka-Ho Wong et al., "Development of a Cantonese Dysarthric Speech Corpus," in *Interspeech*, Dresden, Germany, 2015.

[6] The Linguistic Society of Hong Kong. (1993) The Jyutping Scheme. [Online]. http://www.lshk.org/node/47

[7] K. H. Wong, Y. T. Yeung, Patrick C. M. Wong, G.-A. Levow, and H. Meng, "Analysis of Dysarthric Speech Using Distinctive Feature Recognition," in *the 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*, Dresden, Germany, 2015.

[8] P. Boersma and D. Weenink. (2014) Praat: Doing Phonetics by Computer [Computer Program]. [Online]. http://www.praat.org