HHS Public Access

Author manuscript

Int Conf Speech Database Assess. Author manuscript; available in PMC 2017 March 28.

Published in final edited form as:

Int Conf Speech Database Assess. 2015 October ; 2015: 86–89. doi:10.1109/ICSDA.2015.7357870.

Analysis of Intonation Patterns in Cantonese Aphasia Speech

Tan Lee[1], **Wang Kong Lam**[1], **Anthony Pak Hin Kong**[2], and **Sam Po Law**[3]

[1]Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR

[2]Department of Communication Sciences and Disorders, University of Central Florida, Orlando, FL, USA

[3]Division of Speech and Hearing Sciences, University of Hong Kong, Hong Kong SAR

Abstract

This paper presents a study on intonation patterns in Cantonese aphasia speech. The speech materials were spontaneous discourse recorded from seven pairs of aphasic and unimpaired speakers. Hidden Markov model based forced alignment was applied to obtain syllable-level time alignments. The pitch level of each syllable was determined and normalized according to the given tone identity of the syllable. Linear regression of the normalized pitch levels was performed to describe the intonation patterns of sentences. It was found that aphasic speech has a higher percentage of sentences with increasing pitch. This trend was found to be more prominent in story-telling than descriptive discourses.

Index Terms

Acoustic signal analysis; tonal language; aphasia; tone normalization; Cantonese

I. Introduction

Aphasia is a collection of language disorders caused by damage to specific regions of a human brain. It most commonly affects language and speech production of the patient. Language impairments associated with aphasia can be present in phonology, morphology, semantics, syntax, and pragmatics [1], [5]. In terms of speech production, aphasia is often manifested by phonemic errors, articulation distortions, and speech disfluencies [2], [19]. It has been suggested that speech features and paralinguistic symptoms in addition to linguistic characteristics are useful in differentiating disordered speech from unimpaired speech and typifying syndromes of aphasia [7], [11], [13], [9]. Particularly, acoustical analysis is believed to have good potential in providing objective evidences to support medical diagnosis. In the past, acoustical analysis of aphasic speech was often limited to a small number of short utterances that were carefully elicited by oral reading. The acoustical measurements were made manually from speech waveforms and spectrograms. There is a strong demand for effective deployment of state-of-the-art spoken language technologies to support large-scale study.

Cantonese is a major Chinese dialect spoken by tens of millions of people in Southern China, Hong Kong and Macau areas. Prosody of Cantonese speech has attracted tremendous

interests from speech researchers, mainly because Cantonese has a complicated tone system. However, efforts on disordered Cantonese speech are rarely seen. Recently a large database of aphasia Cantonese speech, namely Cantonese AphasiaBank, has been developed as part of an inter-disciplinary project on multi-modal and multi-level analysis of Chinese aphasic discourse [10]. In [12], a hidden Markov model (HMM) based continuous speech recognition system of Cantonese was used for automatic time alignment of parallel aphasia speech and unimpaired speech. Statistical analysis of supra-segmental duration measurements showed that aphasia utterances contain more frequent insertion of pauses and significantly shorter speech chunks than unimpaired speech. A follow-up investigation revealed that the duration difference between content words and function words in aphasia speech is consistently smaller than that in unimpaired speech.

The present study is focused on another important dimension of prosody, namely intonation. We investigate the sentence-level intonation patterns and make comparison between aphasia and unimpaired speech. The intonation pattern is represented by linear regression over normalized pitch levels of all syllables in the sentence. It is found that aphasia utterances tend to have more sentences with pitch increasing.

II. Background

A. Cantonese Dialect

Like Mandarin or Putonghua, Cantonese is a monosyllabic and tonal language. Each Chinese character is pronounced as a syllable that can be divided into an Initial part and a Final part. Each Cantonese syllable carries a specific lexical tone. When the tone changes, the syllable may correspond to a different character with different meaning. There are 20 Initials, 53 Finals and 9 tones in Cantonese [4]. Figure 1 depicts the pitch patterns of the Cantonese lexical tones. A syllable carrying entering tone must end with an occlusive code /p/, /t/ or /k/, which makes the syllable duration substantially shorter than a syllable with non-entering tone. In terms of pitch level, each of the entering tones coincides with one of the non-entering tones. Therefore in most transcription systems, e.g., the LSHK's JyutPing system [16], only six distinct tone labels are used. They are numbered from 1 to 6 as shown in Figure 1.

B. Cantonese AphasiaBank

Cantonese AphasiaBank is a multi-modal and multi-level corpus developed jointly by University of Central Florida and University of Hong Kong [10]. The corpus contains speech recordings from 120 unimpaired and 96 aphasic individuals, all being native speakers of Cantonese. The speakers are divided into three age groups (18 - 39; 40 - 59; 60 and above) and two education levels. The recordings were elicited as spontaneous oral narratives on 8 different tasks. The tasks include single picture description, sequential picture description, procedure discourse, story telling and personal monologue, with details given as in Table 1.

All recordings were manually transcribed using the Child Language ANalyses computer program (CLAN; [17]). CLAN has been widely used in studies of conversational interaction, language learning, and language disorders. The CLAN editor supports functions such as

annotation of morpho-syntactic features on linguistic transcripts, adding codes to files, linking transcripts to media, and automatic computation of different indices (e.g., MLU or TTR). The orthographic transcription of each recording was prepared in the form of a sequence of Chinese characters. Non-speech events, such as fillers or unintelligible jargons and speech segments, are also tagged with specific codes.

C. Speech Materials

Seven pairs of age- and gender-matched aphasia and unimpaired speakers were selected from the Cantonese Aphasi-aBank. They include five pairs of male and two pairs of female. Among the seven aphasia speakers, five are anomic aphasia and the other two are transcortical sensory. They are considered to be fluent speakers from speech pathology perspective. There were 112 recordings from the 14 speakers. The average lengths of recordings are 48.6 second and 48.2 second for aphasia and unimpaired speakers respectively.

III. Method

A. Automatic Speech Alignment

Automatic time alignment at sub-syllable level was performed on the speech materials using HMM forced alignment technique. The acoustic feature vector was composed of 13 MFCC parameters, and their first and second derivatives, which were computed with a short-time frame size of 25 msec and frame shift of 10 msec. Gender-dependent acoustic models were trained with about 20, 000 utterances from the CUSENT database [14]. Each Initial unit was modeled with a 3-state HMM and each Final was modeled with a 5-state HMM. Each HMM state was represented by a Gaussian mixture model with 16 mixture components.

For each speech recording, the Chinese characters in the corresponding orthographic transcription were first converted to Cantonese syllables. Each syllable was then further decomposed into an Initial and a Final. The process of forced alignment of speech is similar to that of automatic speech recognition. The given sequence of Initial and Final units are used to define a highly restricted search space that contains only one legitimate HMM sequence. The goal of search is to determine the model-level time alignment that best matches the input utterance [20]. The time boundaries of syllables and words could be derived straightforwardly from the time alignments of Initials and Finals. An example of forced alignment result is shown as in Figure 2.

B. Marking sentence boundaries

Sentence boundaries were marked manually by inspecting both the orthographic transcriptions and audio recordings. This is an important step because intonation patterns would be established for individual sentences. In principle, a sentence is a sequence of words which conveys a complete thought. However, in this study, some sentences do not complete in meaning. The speaker might start another sentence and switch to a new topic. The speaker might also quote what was supposed to be spoken by the character in the story. In these cases, it was considered that a new sentence was started.

Conjunctions like ‘because’, ‘then’, ‘therefore’, ‘finally’ hint that a new sentence starts. Normally a sentence contains one verb only, except the case when the verb subcategorizes for a clause, e.g., ‘discover’, ‘decide’, ‘believe’, ‘estimate’, etc. In some of the utterances, especially those by aphasic speakers, long pauses or breaks are included. We do not consider these pauses or breaks as sentence boundaries if the speech fragments before and after are expressing the same idea and involve only one verb.

As a result, a total of 1,072 sentences were marked in the 112 recordings. Among them, 419 sentences were from aphasia speakers. The average sentence lengths are 10 syllables and 13 syllables for aphasia and unimpaired speech, respectively.

C. Pitch Estimation

Pitch values were estimated every 10 msec from the speech signals. We used three existing pitch estimation algorithms: PEFAC [3], [8], RAPT [18], [3] and YIN [6]. These algorithms are based on different principles and therefore considered to be complementary to each other. Given the three estimated pitch values, the average of the two closest values was taken as the pitch of the respective frame. Knowing the time alignment of a syllable, we used the median of the pitch values from all voiced frames to represent the pitch level of the syllable.

D. Pitch Normalization

The pitch level of a syllable is determined by the underlying lexical tone. For example, a syllable carrying Tone 1 is expected to have a higher pitch than one with Tone 3. In continuous speech, the pitch of a syllable depends also on the sentential context. In order to separate the long-term trend of pitch change from the local tone variation, a method of pitch normalization was proposed by Li [15] in an effort on Cantonese prosody modeling. The basic idea is to convert the raw pitch value of a syllable into a value that is independent of the syllable's tone identity. This can be done by multiplying a conversion ratio. For example, the pitch level of a syllable carrying Tone 1 can be converted into an equivalent pitch of Tone 3 by multiplying a ratio of 0.8.

We make reference to [15] and adopt the conversion ratios in Table 2 to convert the pitch levels of all syllables in the utterance,

$$c = f \times R_k \quad (1)$$

where f is the original pitch of a syllable, c is the converted pitch, $k = 1, 2, \ldots, 6$ is the tone identify of the syllable, and R_k is respective conversion ratio. The ratios were obtained by Li [15] through statistical analysis.

After the conversion, we could consider as if all syllables were carrying the same tone. In other words, the normalized pitch values are independent of the tone identities. The change of these pitch values would reflect the global trend over the sentence.

E. Derivation of Intonation Patterns

The intonation pattern of each sentence was obtained by performing linear regression on the normalized pitch values of all syllables in the sentence. The objective is to find the best fit by minimizing the sum of squared errors. Each intonation pattern is described by two parameters: the slope m and the offset b. A positive value of m indicates that the sentence has a tendency of increasing pitch, and a negative value hints a tendency of decreasing pitch.

An example is shown as in Figure 3. The original and the normalized pitch values are marked in red cross and green square respectively. There are two sentences shown in the Figure, one with a declining pitch and the other with an increasing trend.

For all recordings of each subject, the number of sentences with positive slopes is counted. We compare the percentages of positive-sloped sentences between aphasia and unimpaired speakers on different tasks and sentence lengths.

IV. Results

Table 3 shows the percentages of sentences with positive slopes of intonation in aphasia and unimpaired speech utterances. Out of the 419 manually marked sentences from aphasia speakers, 108 (25.78%) show increasing pitch. In unimpaired speech, there are only 69 out of 653 sentences (10.57%) having positive slopes. The significant difference could be partly due to the fact that sentences in aphasic speech are relatively short and short sentences do not show declining pitch in general. Nevertheless, if we focus on only short sentences (8 syllables or shorter), the percentage of sentences with positive slopes in aphasia speech is much higher than that in unimpaired speech.

Table 4 shows the percentages of positive-slope intonations for story-telling and descriptive discourses separately. It can be seen that sentences in story-telling discourses more frequently have rising-pitch intonation.

V. Conclusion

By analyzing the speech materials from a limited number of speakers, it was observed that aphasia speech contains more sentences with rising-pitch intonation than unimpaired speech. The difference is quite significant. The method of linear regression on tone-normalized pitch values is effective to show the contrast, despite that the derived intonation patterns may be over-simplified and inaccurate. We plan to extend the study by including more speakers and carrying out statistical significance test. Our ultimate goal is to apply intonation analysis to objective assessment of aphasia speech.

Acknowledgments

This study is supported in part by a grant funded by the National Institutes of Health to Anthony Pak-Hin Kong (PI) and Sam-Po Law (Co-I) (project number: NIH-R01-DC010398).

References

1. Basso, Anna. Aphasia and its therapy. Oxford Univ Press; 2003.

2. Blumstein, Sheila E., Cooper, William E., Goodglass, Harold, Statlender, Sheila, Gottlieb, Jonathan. Production deficits in aphasia: a voice-onset time analysis. Brain and language. 1980; 9(2):153–170. [PubMed: 7363061]

3. Brookes, Mike, et al. Voicebox: Speech processing toolbox for matlab. 1997. Software, available [Mar 2011] from www.ee.ic.acuk/hp/staff/dmb/voicebox/voicebox.html

4. Ching PC, Lee T, Lo WK, Meng H. Cantonese speech recognition and synthesis. Advances in Chinese Spoken Language Processing. 2006:365–386.

5. Albyn Davis, George. Aphasiology: Disorders and clinical practice. Pearson College Division; 2007.

6. De Cheveigné, Alain, Kawahara, Hideki. YIN, a fundamental frequency estimator for speech and music. The Journal of the Acoustical Society of America. 2002; 111(4):1917–1930. [PubMed: 12002874]

7. Gandour, Jack, Petty, Soranee Holasuit, Dardarananda, Rochana. Dysprosody in Broca's aphasia: A case study. Brain and Language. 1989; 37(2):232–257. [PubMed: 2765857]

8. Gonzalez, Sira, Brookes, Mike. Signal Processing Conference, 2011 19th European. IEEE; 2011. A pitch estimation filter robust to high levels of noise (PEFAC); p. 451-455.

9. Helm-Estabrooks, N., Albert, ML. Manual of aphasia therapy. Pro Ed; Austin, TX: 1991.

10. Kong APH, Law SP, Lee ASY. The construction of a corpus of Cantonese-aphasic-discourse: a preliminary report. Poster presented at the American Speech-Language-Hearing-Association Convention. 2009

11. Lee A, Kong APH, Law S. Using forced alignment for automatic acoustic-phonetic segmentation of aphasic discourse. Procedia Social and Behavoiral Sciences. 2012; 61:92–93.

12. Lee T, Kong A, Chan V, Wang H. Analysis of auto-aligned and auto-segmented oral discourse by speakers with aphasia: A preliminary study on the acoustic parameter of duration. Procedia-Social and Behavioral Sciences. 2013; 94:71–72.

13. Lee T, Kong APH, Wang H. Duration of content and function words in oral discourse by speakers with fluent aphasia: Preliminary data. Frontiers in Psychology. 2014

14. Lee T, Lo WK, Ching PC, Meng H. Spoken language resources for Cantonese speech processing. Speech Communication. 2002; 36(3-4):327–342.

15. Li, Yu Jia. PhD thesis. The Chinese University of Hong Kong; 2003. Prosody Analysis and Modeling for Cantonese Text-to-Speech.

16. Linguistic Society of Hong Kong. Hong Kong Jyut Ping Characters Table. Linguistic Society of Hong Kong Press; Hong Kong: 1997.

17. MacWhinney, B. The CHILDES project: Tools for analyzing talk. Lawrence Erlbaum; Hillsdale, NJ: 2003.

18. Talkin D. A robust algorithm for ptch tracking. Speech Coding and Synthesis. 1995

19. Wambaugh, Julie L., Doyle, Patrick J., Kalinyak, Michelene M., West, Joan E. A critical review of acoustic analyses of aphasic and/or apraxic speech. Clinical Aphasiology. 1996; 24:35–63.

20. Young, Steve, Evermann, Gunnar, Gales, Mark, Hain, Thomas, Kershaw, Dan, Liu, Xunying, Moore, Gareth, Odell, Julian, Ollason, Dave, Povey, Dan, et al. The HTK book. Vol. 2. Entropic Cambridge Research Laboratory Cambridge; 1997.

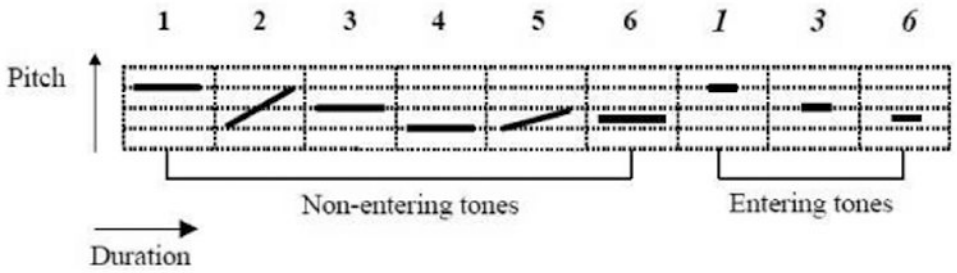

Fig. 1. Cantonese tones

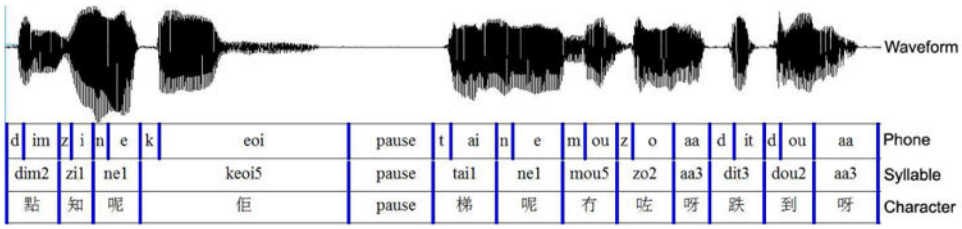

Fig. 2. Forced alignment result on an aphasic speech utterance

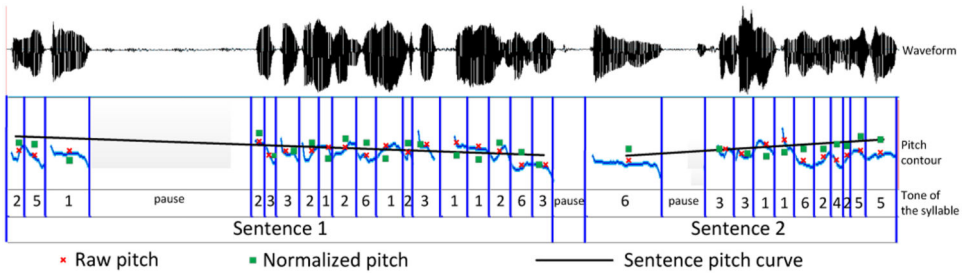

Fig. 3. Pitch curve of two sentences

Table I

Recording tasks used in this study

Task	Recording	Description
Single picture description	CatRe	Black and white drawing of a cat on a tree being rescued
	Flood	A colour photo showing a fireman rescuing a girl
Sequential picture description	BroWn	Black and white drawing of a boy accidentally break a window
	RefUm	Black and white drawing of a boy refusing an umbrella from his mother
Procedure description	EggHm	Procedures of preparing a sandwich with egg, ham and bread
Story telling	CryWf	Telling a story from a picture book “The boy who cried wolf”
	TorHa	Telling a story from a picture book “The tortoise and the hare”
Personal monologue	ImpEv	Description of an important event in life

Table II

Pitch conversion of the tones

Tone	1	2	3	4	5	6
Ratio	0.80	1.07	1.02	1.32	1.13	1.13

Table III

Percentages of intonation patterns with positive slopes in spontaneous oral discourses by aphasia and unimpaired speakers

	sentence length	≤ 8	9-12	≥ 13	overall
Aphasia	no. of +ve slope	52	35	21	108
	total number	179	128	112	419
	% of +ve slope	29.05%	27.34%	18.75%	**25.78%**
Control	no. of +ve slope	14	26	29	69
	total number	107	232	314	653
	% of +ve slope	13.08%	11.21%	9.24%	**10.57%**

Table IV

Comparison between story-telling and descriptive discourses

	Task type	Story	Descriptive
Aphasia	no. of +ve slope	68	39
	total number	226	185
	% of +ve slope	30.09%	21.08%
Control	no. of +ve slope	37	29
	total number	377	267
	% of +ve slope	9.81%	10.86%