Multilingual Aphasia Speech Analysis with Machine Learning

Rong Tong[1], Shih-Cheng Yen[2], Arthur Tay[2], Yiting Emily Guo[1, 3]

[1] Singapore Institute of Technology
[2] National University of Singapore
[3] National University Hospital Singapore
tong.rong@singaporetech.edu.sg, shihcheng@nus.edu.sg, arthurtay@nus.edu.sg, emily.guo@singaporetech.edu.sg, emily_yt_guo@nuhs.edu.sg

Abstract

Aphasia is an acquired language disorder that occurs after brain injury such as stroke, head trauma or tumor. People with aphasia (PWA) may have trouble speaking or understanding speech. If diagnosed early, aphasia is often treatable, and the communication can be improved with speech therapy. Early detection and evaluation of aphasia is crucial for the treatment and recovery. This paper reports a preliminary study of multilingual aphasia speech evaluation. In this study, the characteristics of speech from PWA and healthy controls are compared from both acoustic and linguistic perspectives. Multiple acoustic features are extracted from aphasic and normal speech to build a language independent aphasic speech detection model. The model achieved good aphasic speech detection performance on both English and Mandarin test sets.

Introduction

Aphasia is an acquired language disorder that occurs after brain injury such as stroke, head trauma or tumour. People with Aphasia (PWA) experience different difficulties with communication such as understanding spoken and written words and sentences, finding the correct words and stringing sentences together. Aphasia has a profoundly negative impact on quality of life both personally and socially (Moeller and Carpenter 2013) even more so than cancer or dementia in some research (Lam 2010).

Speech therapy has been found to be beneficial for language recovery in aphasia, especially if provided at a high dosage and intensity with targeted home practice (Brady 2021). During a therapy session, the speech therapist makes use of customized training strategies and may incorporate pictures or reading stimulus to train language skills. However, the dosage and intensity of usual speech therapy services is usually inadequate both locally and overseas (Yiting Emily Guo 2014) (Rose et al. 2014) .

An automatic speech evaluation system may help in more efficient diagnosis of aphasia as analysis and transcription of speech samples is often time-consuming (Prins and Bastiaanse 2004). Such a system may also be used to provide self-paced speech therapy to PWA to increase the intensity of practice. Many researchers are working with speech therapists and neurologists to study how to use technology in helping the recovery of PWA. A survey paper (Jothi and Mamatha 2020) summarizes the machine learning based automatic speech assessment systems. Various speech and non-speech features are explored for aphasic speech. A set of paralinguistic features (Kohlschein et al. 2017) are explored in automatic aphasia evaluation system. Statistics of non-speech, clear-Speech and vague-speech are explored (Le and Provost 2014) in aphasic speech quality assessment. Posteriorgrams (Qin, Lee, and Hin Kong 2019) feature derived from automatic speech recognition is also used for the prediction of the degree of aphasia severity.

Although the majority of studies in aphasic speech are focused on English speaking patients (Herath et al. 2022), aphasic speech assessments in other languages are also explored. A deep learning based speech assessment is performed on time frequency distribution of Mandarin aphasic speech (Kumar et al. 2022). A fully automated system is built for the assessment of Cantonese (Qin, Lee, and Kong 2020) speaking PWA. With the trend of globalization, bilingual aphasia (Abutalebi et al. 2009) (Kuzmina et al. 2019) has drawn much research attention.

This paper studies the aphasic speech of English and Mandarin speakers. The characteristics of normal speech and PWA speech are compared, the distinctive acoustic features for aphasia speech detection are studied. A multilayer perceptron classifier is trained from the 180-dimensional acoustic features for aphasia detection.

Aphasia Speech Evaluation

This work is a preliminary study of an automatic aphasic (Lam 2010) speech evaluation system on Singapore speakers. In the first phase, the proposed aphasia speech evaluation system is targeted to detect aphasic speech in two languages: English, Mandarin. In the second phase, the system will be able to analysis the severity of the aphasic speech, report the potential errors and language usage feedback to help PWA. In this paper, we present the findings in the first phase of development.

Singapore is a multiracial multicultural society. There are four major languages spoken by people in Singapore: English, Mandarin, Malay and Tamil. The multilingual environment has a great impact on people's oral language. For example, the English spoken in Singapore is often referred as Singlish, as it has its unique pronunciation and vocabu-

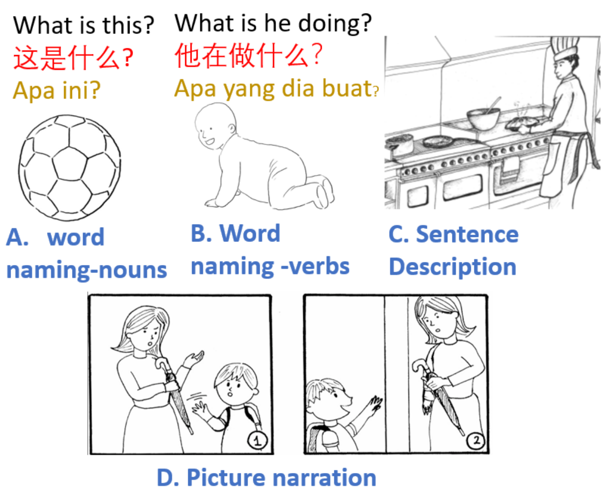

Figure 1: Example of stimulus material in four sessions

lary. The PWA's speech is inevitably influenced by the multilingual environment. This paper reports the preliminary findings towards the characteristics of Singapore aphasia speech and building of a language independent aphasia speech detection system. We believe our study will contribute to the research on accented and multilingual aphasic speech.

Data Collection

The data collection is conducted using a customized tablet application. Both healthy people and PWA are invited in the data collection. During data collection, four sessions with different stimuli were provided for each speaker:

- A. word naming for nouns
- B. word naming for verbs (Ping 2011)
- C. sentence description with customized drawings
- D. picture narration using the Refused Umbrella picture (MacWhinney et al. 2011)

Figure 1 shows an example stimulus for each session. During data collection, a series of such pictures will be shown to the speaker. The same set of stimulus material is used for multilingual speakers, hence verbal and written instructions in multiple languages will be provided as to speakers.

The speech data are recorded in 22k sampling rate and stored in m4a format, the speech data are organized according to speaker and stimulus task. In general, 164 audio files are recorded for each speaker, in which 70 for word naming-nouns, 72 for word naming-verbs, 21 for sentence description, and 1 long audio file for picture narration. After the recording, the speech content of each recording is transcribed into text by research assistants. This study is based on the processed data of 5 PWA, where 3 English speaking (2 male, 1 female) and 2 Mandarin speaking (1 male, 1 female); 129 normal speakers (68 English speaking, 61 Mandarin speaking) with balanced gender distribution.

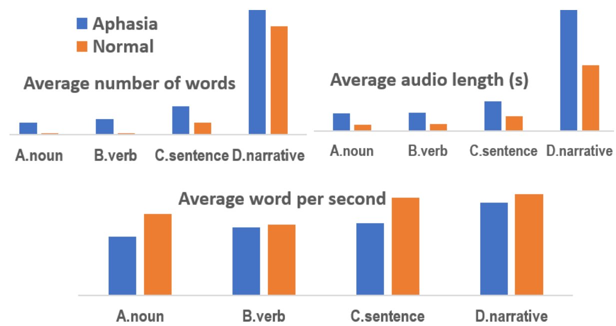

Figure 2: Statistics of aphasic and normal speech

Aphasia Speech Analysis

Speech from PWA and healthy controls were analyzed and compared. Figure 2 illustrates the average speech length and number of words in aphasia speech and normal speech. The breakdown statistics of four sessions are plotted. Compared to normal speech, aphasic speech is longer in length and had a higher number of words across all of the four sessions. Subsequently PWAs had lower speaking rate, which is measured by the number of words per second. The statistics support the observation that PWAs usually need more time to figure out how to describe the pictures in proper words.

We also compared the linguistic characteristics of aphasic and normal speech. Specifically, the part-of-speech (POS) ratio of the following broad categories are calculated:

- NN: single, plural and proper noun
- VB: all forms of verbs
- ADJ: adjective and superlative
- ADV: adverbs
- CON: preposition and conjunction

For English data, NLTK tool (NLTK 2001) is used for part of speech tag detection. For Mandarin data, (Jieba 2012) Mandarin text segmentation tool is used for word segmentation and POS tag extraction. The part-of-speech ratio is calculated by using the number of part-of-speech words divided by the total number of words in the sentence.

Table 1 reports the average part-of-speech ratio of aphasic and normal speech on four sessions. For session A, which is noun word naming task, healthy speakers obtained 0.96 noun ratio on average. That means 96% of words spoken by healthy speakers were nouns. However, the average noun ratio in session A for PWAs was 0.77, which means PWAs only produced 77% percent of noun words. Similarly for session B, verb naming task, healthy speakers achieved an average verb ratio of 0.97 while PWAs only had 0.51 verb ratio. For sessions C and D, where long sentences and paragraphs are expected from speakers, PWAs tended to have a higher ratio of adjectives and adverbs.

Figure 3 lists the top 10 most frequently used words for both PWAs and healthy speakers of English and Mandarin. Compared to healthy speakers, less nouns and verbs were by PWAs and this is consistent with the findings in Table 1. Mandarin PWAs used filler words (ah, uh) frequently, and

session /POS ratio	NN	VB	ADJ	ADV	CON
aphasia A	0.77	0.08	0.05	0.01	0.03
aphasia B	0.27	0.51	0.04	0.01	0.05
aphasia C	0.36	0.24	0.06	0.02	0.06
aphasia D	0.22	0.23	0.07	0.07	0.10
normal A	0.96	0.02	0.02	0.00	0.00
normal B	0.02	0.97	0.00	0.00	0.00
normal C	0.31	0.31	0.02	0.00	0.03
normal D	0.21	0.22	0.06	0.06	0.14

Table 1: POS ratio of aphasia and normal speech

Mandarin aphasia	ah, 在, don't, know, 了, uh, 这个, 什么,那个,他
Mandarin normal	在, 啊, 他, 的, uh, 一个, 这个, 那个, 雨伞, 做
English aphasia	body, people, uh, tree, kruk, krek, kricking, bog, inside, body
English normal	uh, umbrella, boy, driving, mother, chef, delivering, playing, rain, back

Figure 3: Top 10 most frequent words

they were observed to switch to English (I don't know) to express difficulty with the task. Such code-switching behavior is commonly observed in a multilingual society like Singapore. English aphasia speakers produced more unknown words (kruk, krek, kricking,bog) as they had problems in retrieving the words correctly.

Aphasia Speech Detection

The task of aphasia speech detection is to detect if the input speech is produced by PWA or normal speaker.

Acoustic Features

Three types of acoustic features are extracted from aphasic and normal speech to train an aphasic speech detection model. Specifically, they are MFCC feature, log melspectogram feature and chroma features. Figure 4 compares the three acoustic features extracted from an audio file of male aphasia speaker and an audio file of male normal speaker.

MFCC stands for Mel-frequency cepstral coefficients. It is one of the most popular features adopted in many speech processing applications, such as automatic speech recognition and audio classification. MFCC features measure the rate of changes in the different spectrum bands. It represents texture or timbre of sound. Compare the MFCC feature of aphasic speech in Figure 3(a) with the MFCC feature of normal speech in Figure 3(b), the coefficients of the aphasic speech have less dynamics.

The time by frequency representation of the audio wave form is called a spectrogram. That spectrogram is then mapped to the Mel-scales to obtain Mel-spectrogram. As human perception of sound intensity is logarithmic in nature, the Mel-spectrogram is then converted into log form, which is the log Mel-spectrogram. Figure 3 (c) and (d) are the illustrations of 128-dimension log Mel spectrograms of aphasia

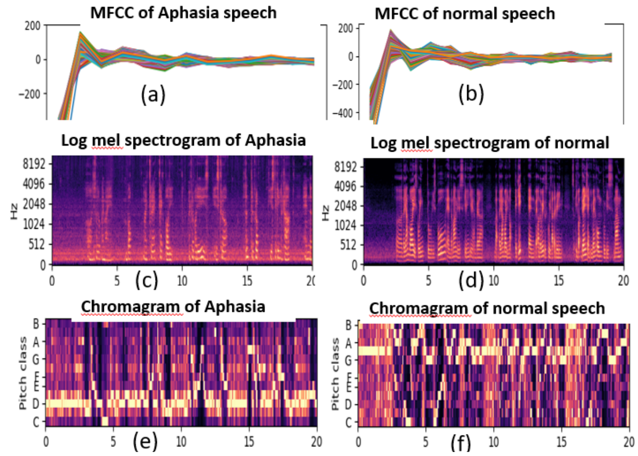

Figure 4: Acoustic features of aphasic and normal speech from two male speakers

speech and normal speech. In this example, aphasic speech is relatively weak in high frequency domains.

Chroma features are often used in music study, it classifies the pitches into 12 distinct classes or pitch profiles. If we compare the chroma feature of aphasic speech and normal speech as shown in Figure 3 (e) and (f), they have significant differences. The aphasic speech is mostly in lower pitch classes. This indicates the chroma feature has a good potential in distinguishing aphasic from normal speech.

Experiment Setup

The three types of acoustic features show there's significant difference between aphasic and normal speech. We are motivated to build an aphasia speech detection model using these features. During feature extraction, three types of acoustic features are concatenated to form 180 dimensional features (40 MFCC+128 log melspectrogram+12 chroma).

During model training, both aphasic and normal speech data are incorporated in training data. Since there are only 5 PWAs' data available, to address the data imbalance issue, we randomly selected a small number of normal speakers as the training data. To explore the possibility of building a multilingual aphasia detection model with only one language, data from two English aphasia speakers are used to present aphasia speech in training. The rest of the aphasia speech is kept for testing. A multilayer perceptron classifier is trained from the 180-dimensional features, the model has hidden layer size of 300, batch size 256. During training, Adam optimizer is used in weight optimization.

Table 2 summarizes the number of audio files and the speaker distribution in the train and test sets. To validate the performance on multilingual speech data, two test sets are formed. The first one is a pure English test set (Test1) and the second one is a pure Mandarin test set (Test2). There is no speaker overlap among the train and test sets. The last row of Table 3 reports the aphasia detection performance. The performance is measured by accuracy, which is the number of correctly detected recordings divided by the total number of test recordings. The experimental result shows the model

Data	Train	Test1	Test2
Audio	984	328	492
Speaker	English 1 A(m) 1 A(s) 4 N(2m,2f)	English 1 A(m) 1 A(s)	Mandarin 1 A(m) 1 A(s) 1 N(m)
Accuracy	NA	100	98.8

Table 2: Data and performance of aphasia detection, where A denotes aphasia speaker, N denotes normal speaker

can detect all English speech correctly for Test1, for Mandarin test set, 98.8% accuracy is obtained, only 6 tests are misclassified, 4 aphasia and 2 normal recordings.

The experiment results show the potential of multilingual aphasia speech detection using multiple acoustic features.

Conclusions and Future Works

This paper presents aphasic speech analysis on English and Mandarin. The aphasic and normal speech of Singapore speakers are analyzed from both acoustic and linguistic perspective. A language independent aphasia speech detection model is built by using three types of acoustic features. Our experiment results show the model has good performance on both English and Mandarin aphasia detection. We would like to extend the study to the other two most popular languages in Singapore: Malay and Tamil.

For the future works, we will study the automatic analysis on the aphasia speech content. Natural language understanding techniques will be explored to characterize the linguistic information in the aphasia speech.

To continue this work, more samples from PWA in multiple languages will be collected. With the support of more data, we hope to develop more efficient aphasia assessment systems and increase therapy dosage and intensity for PWA.

References

Abutalebi, J.; Della Rosa, P. A.; Tettamanti, M.; Green, D. W.; and Cappa, S. F. 2009. Bilingual aphasia and language control: a follow-up fMRI and intrinsic connectivity study. *Brain and language*, 109(2-3): 141–156.

Brady, M. C. 2021. Dosage, Intensity, and Frequency of Language Therapy for Aphasia: A Systematic Review-Based, Individual Participant Data Network Meta-Analysis. *Stroke*, 53(3): 956–967.

Herath, H. M. D. P. M.; Weraniyagoda, W. A. S. A.; Rajapaksha, R. T. M.; Wijesekara, P. A. D. S. N.; Sudheera, K. L. K.; and Chong, P. H. J. 2022. Automatic Assessment of Aphasic Speech Sensed by Audio Sensors for Classification into Aphasia Severity Levels to Recommend Speech Therapies. *Sensors*, 22(18).

Jieba. 2012. Jieba Chinese text segmentation. https://github.com/fxsjy/jieba. Accessed: 2023-06-11.

Jothi, K.; and Mamatha, V. 2020. A systematic review of machine learning based automatic speech assessment system to evaluate speech impairment. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 175–185. IEEE.

Kohlschein, C.; Schmitt, M.; Schüller, B.; Jeschke, S.; and Werner, C. J. 2017. A machine learning based system for the automatic evaluation of aphasia speech. In *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 1–6. IEEE.

Kumar, A.; Mahmoud, S. S.; Wang, Y.; Faisal, S.; and Fang, Q. 2022. A Comparison of Time-Frequency Distributions for Deep Learning-Based Speech Assessment of Aphasic Patients. In *2022 15th International Conference on Human System Interaction (HSI)*, 1–5. IEEE.

Kuzmina, E.; Goral, M.; Norvik, M.; and Weekes, B. S. 2019. What influences language impairment in bilingual aphasia? A meta-analytic review. *Frontiers in psychology*, 10: 445.

Lam, W. P., Jonathan M. C. ; Wodchis. 2010. The Relationship of 60 Disease Diagnoses and 15 Conditions to Preference-Based Health-Related Quality of Life in Ontario Hospital-Based Long-Term Care Residents. *Medical Care*, 48: 380–387.

Le, D.; and Provost, E. M. 2014. Modeling pronunciation, rhythm, and intonation for automatic assessment of speech quality in aphasia rehabilitation. In *Fifteenth Annual Conference of the International Speech Communication Association*.

MacWhinney, B.; Fromm, D.; Forbes, M.; and Holland, A. 2011. AphasiaBank: Methods for studying discourse. *Aphasiology*, 25(11): 1286–1307.

Moeller, D.; and Carpenter, C. 2013. Factors affecting quality of life for people who have experienced a stroke. *International Journal of Therapy and Rehabilitation*, 20: 207–216.

NLTK. 2001. Natural Language Toolkit (NLTK). https://github.com/nltk/nltk. Accessed: 2023-06-01.

Ping, S. W. 2011. The Development of the Bilingual Language Assessment Battery. https://scholarbank.nus.edu.sg/handle/10635/48475. Accessed: 2023-07-01.

Prins, R.; and Bastiaanse, R. 2004. Analyzing the spontaneous speech of aphasic speakers. *Aphasiology*.

Qin, Y.; Lee, T.; and Hin Kong, A. P. 2019. Combining Phone Posteriorgrams from Strong and Weak Recognizers for Automatic Speech Assessment of People with Aphasia. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6420–6424.

Qin, Y.; Lee, T.; and Kong, A. P. H. 2020. Automatic Assessment of Speech Impairment in Cantonese-Speaking People with Aphasia. *IEEE Journal of Selected Topics in Signal Processing*, 14(2): 331–345.

Rose, M.; Ferguson, A.; Power, E.; Togher, L.; and Worrall, L. 2014. Aphasia rehabilitation in Australia: Current practices, challenges and future directions. *International Journal of Speech-Language Pathology*, 16(2): 169–180.

Yiting Emily Guo, L. T. . E. P. 2014. Speech pathology services for people with aphasia: what is the current practice in Singapore? *Disability and Rehabilitation*, 36(8): 691–704.