International Conference on Research in Education and Science
www.icres.net May 18-21, 2023 Cappadocia, Turkiye www.istes.org

Dataset of Chinese Language Beginning Learners Reading Speech and Text-to-Speech

Yoke Lian Lau
Universiti Malaysia Sabah: Kota Kinabalu, Sabah, Malaysia, https://orcid.org/0000-0002-7949-7683

Abstract: Three categories of audio recordings, a web service platform, and commercial software were involved in this analysis process. The three categories of audio recordings included eight filtered student recordings, an audio recording provided by a Mandarin instructor from University Malaysia Sabah, and an audio recording generated by Text-to-Speech MP3, available at https://ttsmp3.com/. All ten recordings will be uploaded into the NCH WavePad Sound Editor to build a spectrum image. This will allow us to compare and identify differences in tone reading, emotion reading, etc. You can download the NCH WavePad Sound Editor from https://www.nch.com.au/WavePad/index.html. This information is beneficial for all Mandarin language beginners and their teachers, as it can be used for comparison purposes. With this knowledge, they can further undertake a pilot study. This study provides a foundation for investigating differences in human and machine reading from multiple perspectives. Furthermore, this dataset can be used to analyze reading speed. It serves as a crucial starting point for spectrum, voice wave, and reader behavior research.

Keywords: NCH Wave Pad Sound Editor, teaching, learning, foreign language, spectrum

Citation: Lau, Y.L. (2023). Dataset of Chinese Language Beginning Learners Reading Speech and Text-to-Speech. In M. Koc, O. T. Ozturk & M. L. Ciddi (Eds.), *Proceedings of ICRES 2023-- International Conference on Research in Education and Science* (pp. 625-639), Cappadocia, Turkiye. ISTES Organization.

Introduction

The data collection process can be divided into three distinct phases: the preparation period, the data collection period, and the data validation period. The planning stage begins in early January 2020, with a team meeting to confirm the research plan and timeline. Team members will recommend six proficient Mandarin instructors who will provide three Mandarin sentences from the University of Malaysia Sabah Mandarin Class textbook. These sentences will be compiled in a Google Drive document. The selected Mandarin instructors will have approximately five months to determine the appropriate sentences.

The team will inform Level 1 Mandarin students at the University of Malaysia Sabah about the study and encourage their participation. They will discuss the research and ensure the students are prepared for the recording process. After May 2020, the team will reconvene to select the ten most relevant sentences for reading speech. A research assistant will prepare for data collection and invite interested individuals to join a WhatsApp group.

The data collection period commences in June 2020 and concludes in July 2020. The research assistant will brief the participants before they begin recording their readings. Participants will be required to carefully read the Participant Consent Form, sign it, and submit it to the research assistant. Furthermore, the research assistant will distribute honorariums to each participant prior to data collection. Participants will then record themselves reading the ten Mandarin sentences listed in section C of the participant authorization form and send the audio file to the research assistant via email or WhatsApp.

During the data validation phase, if there are any errors in the recordings, the participant will be requested to resubmit them.

In total, 16 audio recordings in MP3, MP4, M4A, and MPEG file formats were obtained voluntarily from 14 participating students who used their smartphones to record their readings. These were then sent to the researcher via email or WhatsApp. The NCH WavePad Sound Editor, available at https://www.nch.com.au/WavePad/index.html, was used to capture 10 spectrum images in PDF format. The software was downloaded onto a Windows 10 personal computer. Six experienced Mandarin language teachers at the University of Malaysia Sabah each proposed three classic, culture-based Mandarin sentences. These were added to a Word document containing ten Mandarin sentences and stored in a Google Drive document. Finally, ten of the eighteen sentences were selected for reading speech. All sentences were chosen from the University Malaysia Sabah Mandarin Class textbook.

Participant Selection Criteria

Students from the Mandarin Level 2 class volunteered for this research. The criteria for participation were as follows:

1. Participants must have joined the research from Mandarin Level 1.
2. Participants must be able to read and understand the provided sentences.
3. Participants must not have any disabilities that would prevent them from performing the tones and rhythms of the language.

.

Reading Criteria:

Recordings were obtained from students, a Mandarin language lecturer, and Text-to-Speech MP3 recordings at https://ttsmp3.com/. All recordings had to meet these criteria:

1. The recording must be done in a quiet room.
2. The recording must be audibly clear, pronunciation must be correct, and the speed and rhythm of speech must be natural.

All recordings had to meet these criteria: I. The recording must be done in a quiet room. II. The recording must be audibly clear, the pronunciation must be correct, and the speed and rhythm of speech must be natural.

Selection of Sentences for Reading

1. The sentences are in classic Mandarin and may be in dialogue form.
2. The sentences used were suggested by Mandarin language teachers who chose them based on their reflection of one or more of the following: a. Describing Mandarin culture b. Reflecting the Mandarin lifestyle c. Common sentences often used in the Mandarin language.

Before pre-recording, the researchers explained the meaning of each sentence to each participant and also provided translations in Malay for each sentence. Every participant was thoroughly briefed before the recording took place.

Limitation

The study's subjects were limited to students of the University of Malaysia Sabah. Additionally, the recordings were limited to sentences appearing in the Mandarin language textbook of the Mandarin language class.

The research took place at:

- Institution: University Malaysia Sabah
- City/Town/Region: Kota Kinabalu, Sabah
- Country: Malaysia
- Latitude: 6° 02' 7.20" N
- Longitude: 116° 07' 4.20" E

Methodology

Tools Used

Every participant, including native Mandarin language speakers, used a smartphone app or recorder to create the recording. The Text-to-Speech feature from the Text-to-Speech MP3 at https://ttsmp3.com/ provided a free text-to-speech recording via a web service.

Method

The NCH WavePad Sound Editor, which can be downloaded from https://www.nch.com.au/WavePad/index.html, was primarily used to analyze and observe the spectrum of the recording. All participants were encouraged to record their readings using smartphone audio recording applications. Each participant, including a native Mandarin instructor, utilized a smartphone application to generate the recording. The Text-to-Speech MP3 provided a free text-to-speech recording through a web service. The NCH WavePad Sound Editor was mainly used to analyze and monitor the recording's spectrum.

Analysis

The NCH WavePad Sound Editor was used to perform a spectrum comparison of the recordings. To simplify the comparison, the spectrums of the recordings were segmented into similar-length, fixed-size sections.

Materials

Several essential materials were required to complete the investigation. These included a Google Drive spreadsheet, a Mandarin textbook (Liu, 2010), a Respondent Consent Form (Lau, 2022b), Spectra (Lau, 2022b), and Audio Recordings (Lau, 2022a). The research team recommended six proficient Mandarin instructors to contribute three Mandarin sentences from the University Malaysia Sabah Mandarin Class textbook (Liu, 2010). These sentences were compiled into a Google Drive document sheet. The selected Mandarin instructors were given approximately five months to finalize the appropriate sentences.
For audio recording collection, the research assistant briefed the participants before they started recording. Participants were required to carefully read the Participant Consent Form, sign it, and return it to the research assistant.

Results

This article references three types of data (Lau, 2022a, 2022b) that are available on Mendeley Data (Mendeley Data, 2022) via a specified URL. The initial contents of the "Audio SGA0028-2019" folder (Lau, 2022a) include 16 audio recordings. The "Spectrums and Respondent Consent Form" PDF file (Lau, 2022b) contains "Spectrums" on pages 1 to 3, and a "Respondent Consent Form" (Lau, 2022b) on pages 4 and 5.

The "Audio SGA0028-2019" folder (Lau, 2022a) contains 16 MP4-formatted audio files. They are labelled as Audio A, Audio A+1, Audio B, Audio B+1, Audio C, Audio D, Audio E, Audio F, Audio H, Audio I, Audio J, Audio K, Audio L, Audio M, and Audio N. Additionally, there is a file titled "Spectrum and Respondent Consent Form" (Lau, 2022b). This file contains 10 voice wave displays. These include eight sound waves from student audio readings, one sound wave from a native Mandarin speaker, and one sound wave from the Text-to-Speech MP3 (2022). These sound waves are represented in spectrum form on pages 1 to 3. The respondent consent form is presented on pages 4 and 5.

Table 1. Details of 16 audio files

Audio Nama	Read by	Item Type	Size	Length(time)	Note
Audio A	Male	MP3	5.89MB	00:02:34	Respondent A provided two recordings
Audio A+1	Male	MP4	1.17MB	00:01:15	
Audio B	Male	MP4	1.82MB	00:01:57	Respondent B provided two recordings
Audio B+1	Male	MP4	1.58MB	00:01:42	

Audio C	Female	MP3	1.34MB	00:01:46	
Audio D	Female	MP4	1.29MB	00:02:40	
Audio E	Female	MP4	1.63MB	00:01:45	
Audio F	Female	MPEG	3.32MB	00:01:27	
Audio G	Male	MP3	4.83MB	00:02:06	
Audio H	Male	M4A	836KB	00:01:39	
Audio I	Female	MP4	1.12MB	00:01:12	
Audio J	Male	MP4	659KB	00:01:20	
Audio K	Female	M4A	839KB	00:01:43	
Audio L	Female	M4A	854KB	00:01:41	
Audio M	Female	MPEG	1.10MB	00:01:12	
Audio N	Female	MP3	4MB	00:01:44	

Table 1 provides information about the sixteen audio files. The table lists the 16 audio file names, the gender of the person who read each file, and whether they were saved in MP3, MP4, M4A, or MPEG format. It also includes information about their size, duration, and notation. The majority of the recordings are between 1MB and 2MB in size, and between 1 and 2 minutes in duration.

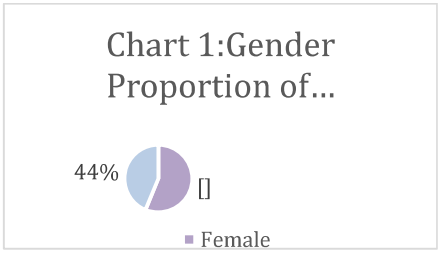

Chart 1. Gender proportion of recordings readers

Chart 1 presents the gender distribution of the readers who participated in the recordings, represented as percentages of the sixteen recordings obtained from University Malaysia Sabah, seven were performed by male students and nine by female students.

Recordings	F %
MP3	25
MP4	43,75
M4A	18,75
MPEG	12,5

Chart 2. Percentage of file types for audio recordings

Chart 2 illustrates the distribution of file types for audio recordings, expressed as percentages. Files are saved in either MP3, MP4, M4A, or MPEG formats. Specifically, 4 files were saved in MP3 format, 7 in MP4, 3 in M4A, and 2 in MPEG.

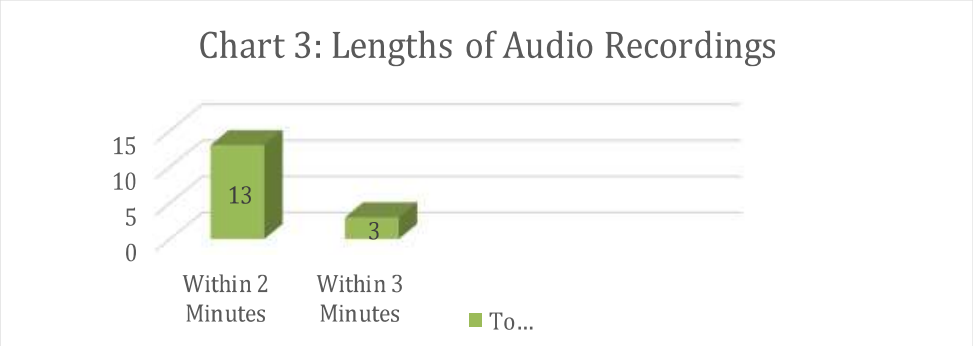

Chart 3. Lengths of Audio Recordings

Chart 3 depicts the total count of audio recordings that are two to three minutes long.

The "Spectrums" (Lau, 2022b) section contains 10 spectrum photos. These include spectra from Subjects 1, 2, 3, 4, 5, 6, 7, and 8. There's also a spectrum from the Text-to-Speech MP3 (2022) reading and one from the Mandarin instructor's reading. All these reading speech recordings were converted into spectrum files using the NCH WavePad Sound Editor, which can be downloaded at http://www.nch.com.au/WavePad.

Table 2:.Categories of Spectrums

Spectrums	Machine reading	Human reading	Explanation
Subject 1		v	
Subject 2		v	
Subject 3		v	
Subject 4		v	
Subject 5		v	
Subject 6		v	
Subject 7		v	
Subject 8		v	
Subject 9		v	
TTSMP3 Reading	v		TTSMP3 refers to Text-to-Speech MP3
Reading of Mandarin Language Lecturer		v	

Table 2 presents the categories of spectrums (Lau, 2022b). Subjects 1 through 8 were randomly selected from the 16 student-submitted recordings and then converted into spectrums for the comparative study. The category "TTSMP3(2022)" refers to the spectrums converted from the Text-to-Speech MP3 web service (Free Text-To-Speech for US English language and MP3 Download | ttsMP3.com, 2022). The "Reading of Mandarin Language Lecturer" category, on the other hand, refers to the spectrums converted from the recording made by the Mandarin instructor at the University Malaysia Sabah.

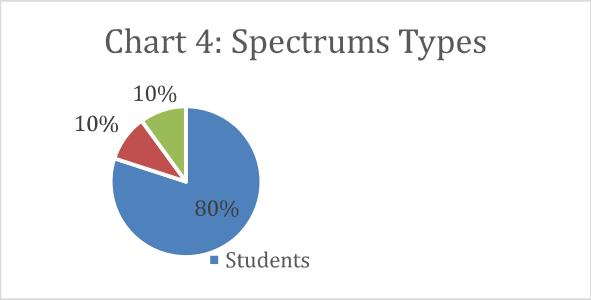

Chart 4. Spectrums Type

Chart 4 represents the types of spectrums as a pie chart. Of these, 80% are derived from student audio recordings, 10% from machine readings, and 10% from a Mandarin instructor's readings. The researcher purchased all spectrums from the NCH WavePad Sound Editor online, thereby gaining access to all functionalities of the web service. Each analysis was conducted in accordance with the website's, web services', and software's open rules and standards (NCH Software legal terms, 2022). The "Respondent Consent Form" (Lau, 2022b) provided pertinent instructions.

Table 3: Components of the Respondent Consent Form and their duties

Part	Components	Purposes	Duties	Action was taken by
A	Name	Check the credentials of respondents.	Ensure that the actual name is written.	Research Assistant
	Gender		Ensure that the correct gender is written.	Research Assistant
	Age		Ensure that they are between 18 and 22 years old.	Research Assistant
	Course Code		Ensure that the correct Mandarin course code is written.	Research Assistant
	Religion		Ensure that the respective religion is written.	Research Assistant
	Which language are you studying as a foreign language?		Ensure that “Mandarin” is written	Research Assistant
	to what extent?		Ensure Level 2 is documented	Research Assistant
	From University Malaysia Sabah Main campus/ Sandakan Campus or Labuan Campus:	Area restrictions of the respondent.	Ensure that one of these campuses is indicated.	Research Assistant

	How to pay you? By way of BOOST? GRAB PAY? TNGO: (Please supply your telephone number or account number.)	Honorarium payments.	Ensure that the payment received details are recorded.	Research Assistant
B	Please record yourself reading the 10 sentences stated below in **Part C.**	Describe the requirements of the recording.	Ensure that the respondents comprehend the recording's terms and conditions.	Mandarin language lecturer
	Please ensure you receive your honorarium payment prior to beginning the recording.	Describe the requirements of the recording.	Ensure that the respondents comprehend the recording's terms and conditions.	Mandarin language lecturer
	All of these recordings will be used as research data to determine the volume/volume of tones/frequency/spectrum of reading speech, etc.	Describe the requirements of the recording.	Ensure that the respondents comprehend the recording's terms and conditions.	Mandarin language lecturer
	As a respondent, your participation in the study signifies that you consent to the use of your data for analysis, sharing, and research purposes and that the researchers will keep it secure.	Describe the requirements of the recording.	Ensure that the respondents comprehend the recording's terms and conditions.	Mandarin language lecturer
	Please use a recorder to create a recording, and please make the recording clear, read it out loud (and before you read it, please read the translation words first), and record what you read properly, correctly, and neatly. After that, please WhatsApp or email your recording to yokelian@ums.edu.my. Thank you for your assistance; the file name should be Mandarin +your name.	Describe the requirements of the recording.	Ensure that the respondents comprehend the recording's terms and conditions.	Mandarin language lecturer

C	宋华一九八二年十月二十七日出生，属狗。 sòng huá yī jiǔ bā èr nián shí yuè èr shí qī rì chū shēng， shǔ gǒu . Song Hua's birthday is October 27th, 1982, and his zodiac sign is the dog.	Pinyin assists respondents in reading the text, while English translations aid respondents in comprehending the text's content.	Ensure that the content is read out loud and correctly.	Respondent
	A: 我很好。你和爸爸身体怎么样? wǒ hěn hǎo 。nǐ hé bà bà shēn tǐ zěn me yàng ? B: 我的身体很好，你爸爸也很好。 wǒ de shēn tǐ hěn hǎo， nǐ bà bà yě hěn hǎo. A: I'm fine, how about your father? B: Your father and I are both in good health.	Pinyin assists respondents in reading the text, while English translations aid respondents in comprehending the text's content.	Ensure that the content is read out loud and correctly.	Respondent
	一斤香蕉多少钱? yī jīn xiāng jiāo duō shǎo qián ? How much does a 500g banana cost?	Pinyin assists respondents in reading the text, while English translations aid respondents in comprehending the text's content.	Ensure that the content is read out loud and correctly.	Respondent
	我买两瓶红葡萄酒。 wǒ mǎi liǎng píng hóng pú táo jiǔ. I purchase two bottles of red wine.	Pinyin assists respondents in reading the text, while English translations aid respondents in comprehending the text's content.	Ensure that the content is read out loud and correctly.	Respondent

	喝点儿啤酒。 hē diǎnr pí jiǔ。 Consume a small amount of beer.	Pinyin assists respondents in reading the text, while English translations aid respondents in comprehending the text's content.	Ensure that the content is read out loud and correctly.	Respondent
	可以吸烟吗? kě yǐ xī yān ma? Permission to smoke?	Pinyin assists respondents in reading the text, while English translations aid respondents in comprehending the text's content.	Ensure that the content is read out loud and correctly.	Respondent
	我们吃寿面。 wǒ men chī shòu miàn . We eat longevity noodles.	Pinyin assists respondents in reading the text, while English translations aid respondents in comprehending the text's content.	Ensure that the content is read out loud and correctly.	Respondent
	我想租一套房子。 wǒ xiǎng zū yī tào fáng zǐ 。 I wish to lease a home.	Pinyin assists respondents in reading the text, while English translations aid respondents in comprehending the text's content.	Ensure that the content is read out loud and correctly.	Respondent
	你愿意吃中药还是愿意吃西药? B:	Pinyin assists respondents in	Ensure that the content is read	Respondent

	我愿意吃中药。 nǐ yuàn yì chī zhōng yào hái shì yuàn yì chī xī yào？ B: wǒ yuàn yì chī zhōng yào . A: Do you favour traditional Chinese medicine or Western medicine? B: I prefer Chinese medicine.	reading the text, while English translations aid respondents in comprehending the text's content.	out loud and correctly.	
	今年是马年，你属什么? jīn nián shì mǎ nián， nǐ shǔ shí me？ This is the year of the horse; what is your zodiac sign?	Pinyin assists respondents in reading the text, while English translations aid respondents in comprehending the text's content.	Ensure that the content is read out loud and correctly.	Respondent
D	I agree to join the research, and I permit for the researchers to use all the given data for analysis for research grant SGA 0028-2019 purpose.	Respondents must check the "v" box.	Make sure respondents check the "v" box.	Research Assistant
	Signature	The respondent must sign the form.	Verify that responders have signed the form.	Research Assistant
	Identity Card Number	Respondents must enter their identification card number.	Ensure that respondents gave the right number for their identification card.	Research Assistant

Table 3 consists of four components. Part A requests the respondent's information to verify that the respondent matches the intended participant. Respondents are asked to provide their E-wallet number, linked to their mobile phone number. They will be informed about their compensation prior to the recording session. Part B outlines the recording criteria. The recording must meet these specified standards, or a retake will be necessary. Part C

comprises 10 sentences from a Mandarin textbook (Liu, 2010). Each sentence is accompanied by its Pinyin and English translation to ensure that each respondent comprehends the meaning and pronunciation of each Mandarin term. The final section, Part D, requires respondents to sign their names, record their identification numbers, and check a box to indicate their agreement with the terms and conditions governing the researchers' use of their data.

Discussion

This research involves three parties. University Malaysia Sabah provides an appropriate platform, locations, and resources for conducting language research, cooperating with the research team and suggesting a proficient Mandarin instructor for the project. The University also assists in the development of each section of the consent form, rules, and conditions related to the audio recording criteria, and verifies the accuracy of all data. The research assistant aids in the payment of honorariums to all respondents, maintains contact with respondents, and examines all acquired data.

The architecture of this study integrates an online machine reading web service, online software, and human reading. The research team selected Text-to-Speech MP3 (2022) to construct machine speech reading audio from text-to-MP3 format due to its user-friendly nature.

Text-to-Speech MP3 (2022) is a free web service accessible to all users. Upon enabling cookies from the homepage, users can access all features of the website for free. This web service has a character limit of 3,000, and the device's speakers must be activated. The researchers selected Chinese Mandarin, the fourteenth option in the drop-down menu, as their target language. They then inserted the ten Mandarin sentences from section C of the "Respondent Consent Form" (Lau, 2022b) into the empty column, clicked 'read', and downloaded the MP3 file of the speech reading. The machine-reading audio recording was thus generated.

To complete the Mandarin-Level-2 learners' speech-reading audio recordings, the research team engaged a research assistant to gather the human speech-reading audio recordings. As the potential respondent was unfamiliar with the recording equipment, they were notified about the research seven months in advance. A member of the research team who also served as a Mandarin instructor at University Malaysia Sabah performed the instructor's speech reading.

After collecting all audio recordings, the research team converted the files to spectrums using the NCH WavePad Sound Editor (2022). As this software wasn't free, the research team paid for it using provided research funds. Ten identically-sized spectrums (Lau, 2022b) were laid out for visual analysis. The analysis focused on reading tone, reading mood, loudness level, sensitivity to key words, and the relationship between word-by-word reading and reading tone speed (Yoke et al., 2021).

Conclusion

Text-to-Speech MP3 (2022) offers a human-like reading format, recording syllable-by-syllable and word-by-word. In the process of translating text to speech, the system combines syllables and words into sentences and then vocalizes them. One of its limitations is the inability of text-to-speech technology to convey human emotions such as happiness, grief, and frustration during the reading process. The data (Lau, 2022a, 2022b) provide an overview of the progression of technology and the extent to which it can substitute humans in education. These data are valuable for a variety of quantitative and qualitative research approaches.

This information can be beneficial to all beginners of the Mandarin language and their teachers. Both instructors and beginner-level Mandarin language learners could utilize the data (Lau, 2022a, 2022b) for comparative purposes. The reading characteristics or style of Malaysian Mandarin may be of interest to researchers. With this understanding, they can conduct a pilot study. The data (Lau, 2022a, 2022b) provide several angles to examine human and machine reading from various perspectives. This dataset (Lau, 2022a, 2022b) can also be used to analyze reading speed, and it illustrates how some students read in a highly formal manner while others do not. It serves as a critical starting point for spectrum, voice wave, and reader behavioral studies. The audio recordings (Lau, 2022a, 2022b) are easy to access and analyze. They can be accessed anywhere and analyzed using any relevant software. Given the typical difficulty of obtaining audio recording type data from respondents, this ready-to-use data is particularly valuable.

Those wishing to learn culture-based Mandarin sentences from the perspective of a native Malaysian Chinese speaker could use the selected sentences from a Mandarin textbook (Liu, 2010) — which are believed to convey cultural traits — as a guide. Mandarin instructors can use it as a data gathering model, while Mandarin learners can use it as a means of self-improvement.

The audio recordings are ideal data for analyzing the differences between human and machine readings. Converting audio recordings into spectrums allows for a clear and vivid visual comparison of the spectrums generated by the three groups of respondents. This provides a straightforward analysis to meet the objectives of the research.

The findings contribute to the achievement of the three goals of the original study. Firstly, they determine the effectiveness of Text-to-Speech in reading Mandarin, the limitations of Text-to-Speech, and the differences between machine and human readings. Secondly, the study suggests that while Text-To-Speech technology is sufficiently advanced at present, there are still challenges to be addressed in certain areas such as conveying human nuances, incorporating more elements of human nature, and expressing human emotions in future reading processes. Lastly, the study can broaden the research scope regarding innovative language teaching and learning technologies.

Acknowledgements

This work was supported by the University Malaysia Sabah SGA0028-2019

References

Yoke, L. L., Osman, W. H., Tobi, B., & Othman, I. W. (2021). A case study on Mandarin text-to-speech. *International Journal of Education, Psychology and Counselling*, *6*(41), 183-198. https://doi.org/10.35631/ijepc.641014

Free Text-To-Speech for US English language and MP3 Download | ttsMP3.com. (2022). https://ttsmp3.com/

Lau, Yoke Lian. (2022a). Audio SGA0028-2019.*Mendeley Data*, V1. doi: 10.17632/ymw548hg6v.1

Lau, Yoke Lian. (2022b). Spectrum and respondent consent form. *Mendeley Data*, V1. doi: 10.17632/dkskjr8zjc.1

Liu, Xun. (2010). *New practical Chinese textbook*. Beijing Language & Culture University Press.

Mendeley Data. (2022)https://data.mendeley.com/

NCH Software - Free Software Downloads and Installs. (2022). https://www.nch.com.au/index.html

NCH Software legal terms. NCH Software - Free Software Downloads and Installs. (2022).https://www.nch.com.au/general/legal.html