future internet

MDPI

Article

Language Cognition and Pronunciation Training Using Applications

Ming Sung Kan * and Atsushi Ito

Utsunomiya University, 7-1-2 Yoto, Utsunomiya, Tochigi 321-8505, Japan; at.ito@is.utsunomiya-u.ac.jp

* Correspondence: s944302@gmail.com; Tel.: +81-080-7806-3131

Received: 10 January 2020; Accepted: 22 February 2020; Published: 25 February 2020

Abstract: In language learning, adults seem to be superior in their ability to memorize knowledge of new languages and have better learning strategies, experiences, and intelligence to be able to integrate new knowledge. However, unless one learns pronunciation in childhood, it is almost impossible to reach a native-level accent. In this research, we take the difficulties of learning tonal pronunciation in Mandarin as an example and analyze the difficulties of tone learning and the deficiencies of general learning methods using the cognitive load theory. With the tasks designed commensurate with the learner's perception ability based on perception experiments and small-step learning, the perception training app is more effective for improving the tone pronunciation ability compared to existing apps with voice analysis function. Furthermore, the learning effect was greatly improved by optimizing the app interface and operation procedures. However, as a result of the combination of pronunciation practice and perception training, pronunciation practice with insufficient feedback could lead to pronunciation errors. Therefore, we also studied pronunciation practice using machine learning and aimed to train the model for the pronunciation task design instead of classification. We used voices designed as training data and trained a model for pronunciation training, and demonstrated that supporting pronunciation practice with machine learning is practicable.

Keywords: language learning; pronunciation practice; perception training; teaching materials design; machine learning

1. Introduction

Language learning may be faster and easier for younger people, and this seems to be due to the differences in effort to acquire native and second languages. However, there is a significant difference in language cognition between native language and second language learning, relative to the learning environment and knowledge of the learner.

According to related works [1,2], the language learning speed of adolescents and adults in the same learning environment is higher than that of children and younger children, who increase their language learning ability as their age increases. However, with the speed of language learning, pronunciation is different from learning words and grammar, or constructing and translating documents [2]. Though older people have more knowledge, learning experience, and memory strategies, it is almost impossible for adult learners of a second language to pronounce new languages with a native-level accent. The existing research was mainly focused on achieving "native-like" fluency or accents in foreign languages [3].

In language learning, we focused on the difference between pronunciation learning and the acquisition of knowledge, such as learning words or grammar, and considered cognitive and pronunciation learning support with the application. This study aims to improve the pronunciation learning efficiency of tones, which is a major issue for foreign language learners, especially those who

do not speak tonal languages. This study was conducted through surveys and evaluation experiments on Japanese learners.

Section 2 describes the related works on Mandarin Chinese tone learning and the cognitive abilities of learners that tend to be neglected in Mandarin Chinese education and theories about human cognition. Section 3 describes the shortcomings of general learning methods and the improvement plans based on cognitive load theory.

In Section 4, the effect of the improved learning method is shown by the evaluation experiment results of the tone learning applications.

Section 5 describes the tasks designed to make pronunciation practice more effective and to give feedback on the pronunciation of learners, provided by machine learning.

Section 6 presents the conclusions and future tasks of this research.

2. Background of This Research

Second language learning refers to interlanguage theory. Systematic errors that occur during language learning are mainly due to negative transfers from the native language. Learning to listen and pronounce a language are different skills from learning to read and write, as the latter involve remembering knowledge (letters, phonetic transcription, words, grammar, phrasing, etc.). The environment in which only the native language is used and age (neurological constraints [3]) make it difficult to cognize elements of pronunciation in a foreign language (phonemes, mora, accent, syllable, tones, etc.) that are not in the learners' mother tongue. To learn the pronunciation of a foreign language, generally, learners need to be able to cognize elements of pronunciation and combine them with knowledge of letters and notation.

A study on language production [4,5] shows that pronunciation of a second language is quite distinct from grammatical encoding, and another study on bilingual language processing [6] shows that bilingual speakers have a knowledge of two phonological systems and can activate both during language processing. However, the difficulty of pronunciation is that there are many learners who cannot pronounce or recognize the second language even if they have knowledge of it. The pronunciation errors of learners are caused by the native-language transfer or part of elements of pronunciation that do not exist in the learners' native language. Though it is important to teach pronunciation properly, it is difficult for all learners to cope with the difficulties and correct pronunciation errors in the second language class, which aims to have learners gain basic conversation skills in a short period of time (one or two semesters). In addition, teachers that instruct language learners may be non-native speakers, and it is possible that the teacher's own pronunciation accuracy and the ability to correct the learner's pronunciation are insufficient.

2.1. Studies about Mandarin Chinese Tone Pronunciation

There are several languages called tonal languages [7] that distinguish the meaning of spoken words based on pitch change patterns included in each syllable. It takes a great deal of effort and time for second language learners to recognize tone correctly and learn to pronounce the words with the right tone, especially for the non-tone native-language speakers. Furthermore, there is a large individual difference in tone learning speed, which places a heavy burden on educators who must help learners deal with tone leaning difficulties and keep learners motivated.

We targeted Mandarin Chinese, which has four main tones, and investigated the study of Mandarin Chinese tone pronunciation. The four main tones used in Mandarin Chinese conversation are the flat-high tone (tone 1), medium-rising tone (tone 2), low-dipping tone (tone 3), and high-falling tone (tone 4), as shown in Figure 1 using Zhao's five-level tone mark [8]. Compared with languages (like Vietnamese) that have six or more tones, the features of Mandarin Chinese tones are more obvious and may seem easier. However, the acquisition of Mandarin Chinese tones is difficult for non-Mandarin Chinese learners, and many studies on Mandarin Chinese tones and tone pronunciation have been conducted to solve this issue.

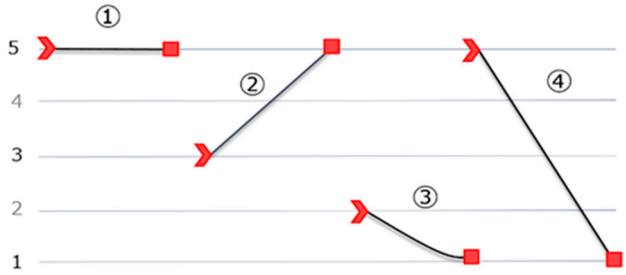

Figure 1. Four tones in Mandarin Chinese using five-level tone marks (the numbers on the left side show the relative pitch: 1. Bottom, 2. Low, 3. Mid, 4. High, 5. Top).

There are several studies on Mandarin Chinese tonal pronunciation targeting learners whose native language is Vietnamese, which has more kinds of tones [9–11]. Vietnamese speakers, whose native language is a tone language, have the knowledge of a tone language, and they have mastered the Vietnamese tones that seem to be more difficult than Mandarin tones. However, due to the negative transfer of the mother tongue, the Vietnamese tones that learners are used to speaking may cause pronunciation errors. In addition, if the similarity between the native language and the Mandarin Chinese tone is high, specific pronunciation errors are likely to occur frequently, such as that the pitch should be slightly higher or that the falling of the pitch should be more obvious.

The majority of the research on the pronunciation of Mandarin Chinese tones that we investigated focused on error analysis and error corrections by phonetic experiments. Though some research may be conducted without restrictions on the subject's native language and learning experience [12], most of the research is to specify the subject's native language, discuss the differences between the native language and Mandarin Chinese, then discuss the pronunciation errors based on the results of the experiments [13–21]. However, the research on Mandarin Chinese tone pronunciation has little content on tone learning method, and the conclusion of the experiment is mainly the error analysis of tone pronunciation of Mandarin Chinese learners, such as the order of difficulty of each tone and the cause analysis result of pronunciation errors. As a result, almost every study focuses on the negative transfer of the learners' mother tongue.

In addition, according to the related works, the tone pronunciation of subjects with more than two years of learning experience is far from ideal, and there is an experiment result even showing that the tone pronunciation ability of learners with more than six years of learning experience is even lower than those of less than one year [15]. Regarding tone learning, a lack of guidance and learners not being conscious of tones, especially fossilized errors due to the neglect of tone learning in initial education, are the reasons that even advanced learners have such a high tone error rate.

The general method of Mandarin Chinese tone learning is to practice the pronunciation of words, phrases, short sentences, etc. The tone learning method mainly used in previous research and the remedial learning after error analysis is pronunciation practice or practicing pronunciation of two-syllable words focused on the combination of tones and short sentences. The remedial instructions in related works included one-month of intensive instruction by a teacher [21] and imitation pronunciation practice using a computer support system with Mandarin Chinese model voices and a pitch tracking function [10,17]. However, these studies showed that the tone improved with the learning methods, and that they was effective; however, they could not show the superiority of the learning methods performed. Therefore, we aim to improve learners' tone cognition and pronunciation ability of tone by using different learning methods and strategies. For this purpose, we researched theories and experimental results of tone cognition and human processing.

2.2. Tone Cognition and Human Processing

A study on Mandarin Chinese tone perception has shown that even advanced second-language learners who have sufficient knowledge of words have difficulty recognizing tones included in

polysyllables and sentences [22]. The differences between the target language and native languages and the individual cognitive ability determine the difficulty in learning tones. However, educational policies and awareness of tone and learning strategies can be modified, which is a major factor in improving the efficiency of tone acquisition. A study was conducted to compare the learning effects of using monosyllable words with using disyllable words on the initial education of Mandarin Chinese tones [23]. The results show that using disyllable words in the initial education is much more effective in the acquisition of tone perception and pronunciation of monosyllable words and disyllable words (18.75% to 45% improvement). Regarding the tone errors of Mandarin Chinese learners, a study shows that teaching the third tone as a "dipping-rising tone" may cause tone pronunciation errors [24]. Furthermore, there are tone errors due to ineffective learning strategies of the learners, such as learners who are confident or good at recognizing and imitating pitch learning tones as melodies or learners who concentrate on reading and writing without noticing the importance of tone that may lead to fossilized errors.

According to our research [25], although Mandarin Chinese beginners were trained in tone knowledge, most of the learners could not distinguish tones in four-syllable Mandarin Chinese, even when we slowed down the speech speed. It is considered that the pronunciation practice by imitating words and phrases, which is a general learning method used in existing learning materials and related research, is not the best way to acquire tone. Learners with low tone cognitive ability are more likely to practice pronunciation without noticing their pronunciation errors and inappropriate learning strategies, and the pronunciation errors get fossilized over time. Therefore, we focused on tone recognition based on information processing in cognitive psychology [26], which uses computer information processing to approach or explain human information processing. In addition, we used machine learning, which increases in accuracy with large amounts and variations of training data to metaphorize how learners acquire categorical perceptions of Mandarin Chinese tones.

Cognitive psychology aims to elucidate the human information processing (perception, memory, cognition, etc.). Human information processing has stages and includes the following three processes:

1. Discovery—Be aware of information in the environment that was previously overlooked;
2. Memory—Make the information meaningful and store in short-term or long-term memory;
3. Modify—Modify the long-term memory (re-learn).

We assume that tone learning is recognizing audio information based on the existing knowledge of learners [27]. The knowledge of tones, especially the visual symbols as shown in Figure 1, can greatly help learners to discover and memorize tones. Furthermore, it is considered that recognizing tones requires considering the shortness of sensory memory (about 1 second or 0.25 to 4 seconds) among the three stages of human memory: sensory memory, short-term memory, and long-term memory.

In the theory of cognitive psychology and education, based on the fact that the mental resources that can be used for human information processing are quite finite, cognitive load theory [28] for effective use of human cognitive resources has been developed. Cognitive load can be classified into the following three categories in relation to learning.

1. Intrinsic cognitive load—Related to the learning content (difficulty of learning content);
2. Extraneous cognitive load—Related to how to present the learning content (interface design, etc.);
3. Germane cognitive load—Related to the learning effects (learning tasks, function design, etc.).

Even learners who are unable to recognize the tones are likely to be able to pronounce tones correctly in pronunciation practice through the imitation of words and phrases due to the help of sensory memory. Imitation pronunciation practice seems to be effective; however, we consider that for learners who are unable to recognize the tone, pronunciation practice by imitation improves speech memory and imitation ability rather than the acquisition of tone. In addition, to avoid fossilized errors caused by pronunciation practice without correction, it is important to support learners' language recognition with different learning contents and methods or voice analysis and machine learning.

3. Research Method

Based on the related research and theory in the previous section, we considered the intrinsic cognitive load (Difficulty of speech recognition, complexity of information received by learners), extraneous cognitive load (Information that is not expected to be effective in learning tones, teaching material interface, and operation design, etc.), and germane cognitive load (The instructional design and feedback to promote tone understanding) of the general learning method. The research procedure is shown in Figure 2.

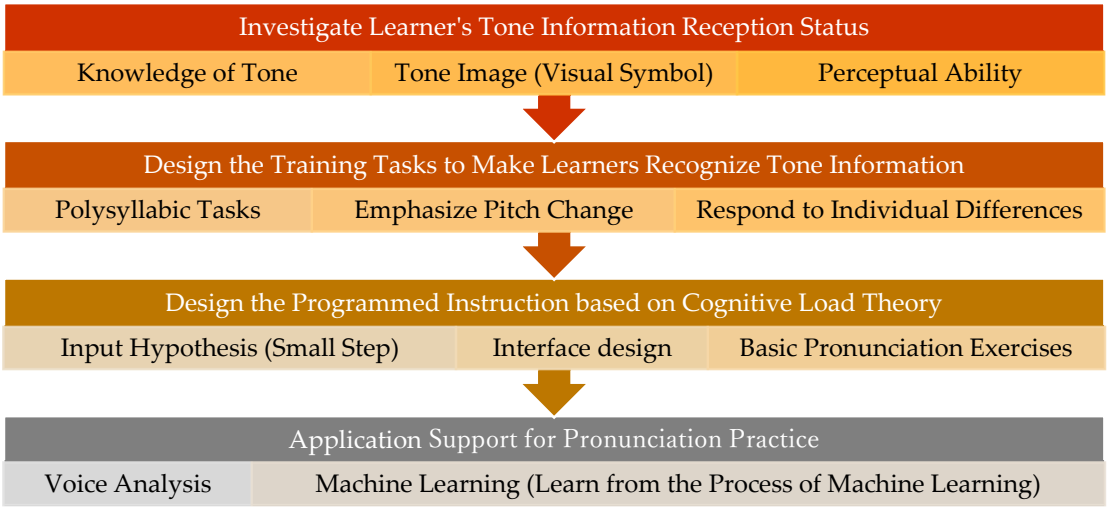

Figure 2. Our research method.

3.1. Intrinsic Cognitive Load (Contents of Learning)

The model voices used for imitation pronunciation practice are mainly of words and phrases. However, a Chinese word or phrase is composed of many elements, such as phonemes, tones, phonetic notations, characters, and meaning, and practicing the pronunciation of words and phrases is accompanied by changes of these elements. It is assumed that the intrinsic cognitive load appears to be very high for most Chinese beginners.

First, we focused on the first stage of human information processing, discovery, and aimed to make learners cognizant of the tone. An experiment was conducted to investigate the relationship between knowledge and tone perception using visual symbols of tone and tone perception tests [25]. The experiment was conducted with native speakers of Japanese that use Chinese characters as well as Mandarin Chinese as subjects and investigated how beginners can recognize tones, and how designed perception tests can help improve cognitive abilities and individual learner differences in tonal cognitive abilities. We have created a collection of perceptual training questions that are unproblematic, emphasizing the comparison of each tone, and with many variations of model voice (multiple pitch ranges, sounds, phonemes, and speakers) to make learners recognize the differences of tone in relative pitch patterns.

3.2. Extraneous Cognitive Load (Perceptual, Visual, and Operation Load)

The intrinsic cognitive load of phrase pronunciation practice seems to be too high for Chinese beginners as described in the previous section, although it can be reduced by having preliminary knowledge. However, information unrelated to tone would lower the learners' concentration on tone and affect the learning effectiveness. In addition, the pitch tracking function often used in teaching materials and pronunciation practice applications takes time to record, analyze, and confirm, which may reduce the pace of learning and increase the cognitive load. The results of audio analysis could not faithfully reproduce what humans feel and may not be effective for learning.

Using teaching materials and applications for learning can be further classified into three categories: perceptual load, visual load, and operation load. Regarding the perceptual load, we set the learning target only to tone recognition, and the audio task was designed based on the learner's cognitive ability. The audio task was provided as a self-learning material in the order of task difficulty. To minimize the visual load of learning contents, only the images of the four tones (Figure 1) that show relative pitch patterns of Mandarin tones were used as visual symbols in the perceptual training application. The operation load of using the application is manly related to how to select the perceptual training question sets and answer the perception questions.

It is assumed that the application interface and operation design are related to the perceptual load, visual load, and movement load during learning and affect the learning effectiveness. To investigate the effects of interface and operation design, we focused on the burden and efficiency of learning and optimized the interface design, the loading time, and the operation method of the perceptual training app.

3.3. Germane Cognitive Load (Instructional Design and Feedback for Learning)

The general teaching materials are a provision of a model voice (mainly words, phrases, and sentences) and provide visual feedback using a pitch tracking function for learners to check their pronunciation. However, a lack of cognitive ability, especially for beginners of Chinese, means that categorization perception of tone has not been advanced, thus it is unlikely that learners can improve or evaluate their own pronunciation correctly by listening or checking the results of pitch tracking at this stage. Furthermore, the results of pitch tracking are different from human perception, and acoustic analysis results vary for visual symbols designed based on the perception of native Chinese speakers.

We considered that the task design of tone learning aims at improving the cognitive ability, which is the basic ability for tone acquisition, and that the speech information provided should focus on promoting the categorization perception of tones. Therefore, referring to the input hypothesis [29], we made the perceptual training questions collection, composed of various training tasks, that led learners to start from the simplest perception questions and gradually change the pitch ranges, sound phoneme, or speaker as the germane cognitive load. Furthermore, showing the answer status of each question and task as learning feedback, the users could improve their perception ability and select perceptual training tasks that matched their abilities according to the feedback.

Regarding pronunciation practice, we investigated the effects of pronunciation practice before and after answering perceptual training questions. The content of the pronunciation practice was pronunciation with tone combinations and phoneme changes.

4. The Improvement of Tone Pronunciation Ability with Applications

We gathered 30 Japanese college students who had never learned Chinese and divided them into groups (depends on their age and gender) to measure the learning effect of the applications. Each group used the application with different learning methods or application designs. We considered that practice every day in a short period is more effective for acquisition of tone, and the procedure of the experiment we designed was as follows:

1. Acquisition of knowledge (about 20 minutes);
2. Training with application (3 days, 30 minutes per day);
3. Pronunciation test (the first test);
4. Again—Training with application (3 days, 30 minutes per day);
5. Again—Pronunciation test (the second test);
6. Questionnaire.

We designed two basic pronunciation tasks: reordering of four tones (use the same phoneme and all four kinds of tones in four syllables and changing the order of the four tones, for a total of 24 patterns; 96 syllables) and two-syllable phoneme combination changes (using two phonemes and disyllable

tone combinations, for a total of 15 patterns and two tasks; 60 syllables) and a phrase pronunciation task. Before testing the basic pronunciation tasks and phrase pronunciation task, participants may have to practice their pronunciation of the tone templates twice (Figure 1) and monosyllable tones (eight syllables), or phrase imitation pronunciation as a break-in.

4.1. Learning Method

The target of the perceptual training application consisted of the perceptual training questions designed in Sections 3.1 and 3.2 to compare with the "Chinese pronunciation learning tool provided by NHK (Japan Broadcasting Corporation)", which includes a function to display model voices of Chinese words, phrases, and sentences. The users of the NHK's tone leaning application can record and play their pronunciation, track their voice pitch, and compare to the model voice. The participants can practice pronunciation by imitation for a range of more than 160 model voices.

After use of the application for a total of three hours, users of the perceptual training application that focused on cognitive ability of tone had an average correct answer rate 20% and were 16% higher in the two basic tone pronunciation tasks. In the results of the paired sample t-test, there is a statistically significant difference in the correct answer rate of the two basic tone pronunciation tasks from learning with the NHK application and the perceptual training application ($t(9) = 2.65$, $p = 0.026$, Cohen's $d = 0.839$; $t(9) = 3.68$, $p = 0.005$, Cohen's $d = 1.16$) [30]. We assume that the simplification of the learning content reduces the intrinsic cognitive load, and the instructional design of the perception questions causes the germane cognitive load for learners. However, regarding the performance of the phrase pronunciation tests after the imitation practice, the scores of users of the perceptual training application were about 2% and 3% lower than the users of the NHK application. The participants who practiced by phrase imitation had a slightly higher ability to imitate and memorize phrases.

To confirm the learning process and phrase pronunciation ability, we conducted an experiment with a pronunciation pre-test. The pronunciation pre-test was performed after providing the tone templates model voices to participants three times. Carrying out the pre-test made the performance of participants improve by about 20% after training with the application. The performance of the perceptual training application users were 17% and 7% higher in the basic pronunciation tasks than the users of the imitation pronunciation practice application. The correct answer rate improvement of the perceptual training application users before and after training were more than double the improvement of the imitation pronunciation practice application users. Furthermore, canceling the imitation practice before the phrase pronunciation test made the imitation pronunciation practice application users perform worse in the phrase pronunciation test. However, the perceptual training application users continued to improve their performance in the phrase pronunciation tests. The experimental results of the phrase pronunciation pre-test, first test, and second test (canceling the imitation practice before the second test) are shown in Figure 3.

4.2. Application Interface and Operation Design

Although the perceptual training application was shown to be more effective than the imitation pronunciation practice application, according to the questionnaire results after the experiment, the users of the NHK application focused more on learning and found it harder to learn tones. In addition, the NHK app is more interesting for tone learning, and the perceptual training app users felt that it was more troublesome.

Therefore, we optimized the interface design, operation design, and loading time, aiming to minimize the extraneous cognitive load without changing the intrinsic cognitive load (contents of learning). The most important optimization for tone recognition is the effort required to answer the perceptual training questions. We eliminated the loading time (0.8 to 1 s in the prototype) required to move to each question, task, and confirmation screen of answer status. The operations of the optimized perceptual training application required to answer the questions are: 1. Click the button to play the sound of the perception question, 2. Click the button to answer the question, 3. Swipe to move to

another question. With the perceptual training app after optimization, it was possible to move to the next question within the length of sensory memory (about 1 second) and challenge the next question.

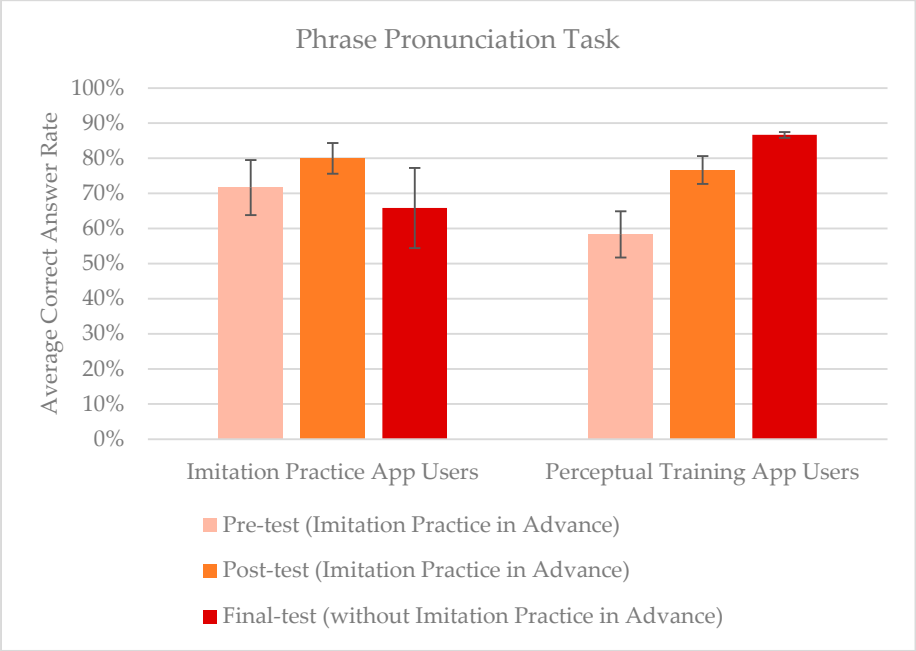

Figure 3. The average correct answer rate and variance of application users on the phrase pronunciation task.

The results of the experiment showed that the users of the optimized perceptual training app performed 10.6% and 20.1% higher on the two basic pronunciation tasks than they did with the prototype, and in the phrase pronunciation tests, the performances of the optimized perceptual training app users were improved by 12.5%. In the results of the paired sample t-test, there is a statistically significant difference in the correct answer rate of the pronunciation tasks from learning with the prototype and the optimized perceptual training application ($t(9) = 2.36$, $p = 0.043$, Cohen's $d = 0.745$; $t(9) = 2.61$, $p = 0.028$, Cohen's $d = 0.827$; $t(9) = 2.45$, $p = 0.037$, Cohen's $d = 0.766$) The results of the questionnaire after the experiment (Table 1) show that the appraisal of the app after optimization improved considerably; however, it is slightly lower than the NHK's app.

Table 1. The questionnaire result of application users (6-point Likert scale). Japan Broadcasting Corporation (NHK).

Average Score/Item	NHK App	Perceptual Training App	
		Prototype	Optimized
I focus on learning the tone	5	4.2	4.8
(At first use) I use the application thoroughly	5.4	4	5.2
(At second use) I use the application thoroughly	5.4	4.4	5.2
This application helps me to understand tone	5.4	5	5
The pronunciation test helps to train tone ability	5.2	5	6
Pronunciation tests are interesting	5.4	5.2	4.8
Tone learning is interesting with this application	5.2	4.4	4.8
The learning method of this application is troublesome	2.4	2.8	2.4
Tone perception is difficult	4.4	5	4.2
Tone pronunciation is difficult	4.2	5	3.8

4.3. Practicing Pronunciation in the Perceptual Training Application

As outlined in Section 4.1, we conducted the pronunciation pre-test before training with the app, and the correct answer rate after training with the application was significantly improved by about 20%. Therefore, we added a screen for instructing the tone pronunciation pre-exercise in the application. In the pronunciation training screen, light practicing content (12 syllables) was used to support the perceptual training after enabling the app, and in addition, the basic pronunciation tasks and the model voice of the four tones were presented at the end of the perceptual training questions for users to practice pronunciation.

As a result, the implementation of the tone pre-exercise screen did not help perceptual training app users to improve their tone pronunciation ability. On the other hand, the number of the perception questions answered by users was reduced due to the pronunciation practice, which may adversely affect the tone learning effects of the perceptual training app. Furthermore, a participant who used the perception questions was only in the first stage of the training with the app, and in the second stage they were only practicing the pronunciation, making the performances of two basic pronunciation tasks and the phrase pronunciation decrease by 4.2% to 14.6%.

5. Discussion

There are studies that have conducted perceptual training to improve the tone ability in Chinese perception and production; however, the content of the learning is mainly monosyllabic and disyllabic. A study that described the effects of perceptual training showed that after two weeks of perceptual training using disyllable words, the tone perceptual ability improved by 21% and lasted for over six months [31]. Studies have also divided participants into groups to measure the learning effect of the perceptual training and pronunciation practice, showing that when using monosyllabic words to perform perceptual and pronunciation training, perceptual training has a significant effect on improving perceptual ability, but the improvement in pronunciation ability is much less than that achieved by pronunciation practice [32]. Similarly, pronunciation practice is effective at improving monosyllable tone pronunciation; however, the effect on improving monosyllable tone perception is lower than that of perceptual training. However, using the application to practice disyllable word perception and pronunciation by imitating a model voice with pitch tracking shows that it is effective to improve pronunciation ability only with perceptual training. Training only in perception is sufficient for improving the tone perception and pronunciation in disyllabic words [33].

There are various factors that can affect the results of the tone pronunciation experiment, such as the training time, individual differences among learners, learning experience of participants, etc. However, considering the reasons why the two studies that compared perception and pronunciation training described above led to different conclusions, we speculate that using monosyllable or disyllable words and training with an application or a teacher is key. Training with a teacher can give the learners a proper opinion and help them to avoid pronunciation errors when practicing pronunciation. As shown in our experimental results, pronunciation practice with only the application may lead to fossilized errors. Furthermore, using monosyllable words to perform perceptual training does not focus on the relative pitch and categorization perception of tones. Although the participants in the study could generally distinguish four tones in monosyllables, recognizing and imaging the Mandarin Chinese tones correctly was still difficult for them.

Based on cognitive psychology theories, the perceptual training app we developed examines the learning content, conducts experiments, and optimizes the perceptual training tasks and application interfaces that help users in learning Mandarin tones. The results of the evaluation experiment showed that the perceptual training application user's cognitive ability was improved, and the tone pronunciation errors were reduced. However, a participant who had a high correct answer rate (93.33%) in the disyllabic pronunciation task after training with perception questions performed worse (83.33) in the disyllabic pronunciation task after practicing their pronunciation instead of perception. Therefore, to support pronunciation practice to improve tone production, the task design and feedback

of pronunciation practice with the application should be based on considering the tone understanding and categorization, error correction, vocal cord exercises, etc.

Learning Mandarin Tone with Machine Learning

In many cases, artificial intelligence (AI) refers to machine learning or neural networks that refer to human cognitive structures. For foreign language pronunciation learning, there are some cases that recommend using AI, such as speech recognition systems, which can be used for practicing conversation in foreign languages. Speech recognition can grade learners' pronunciation through speech analysis of the acoustic features, etc. However, the answers calculated using these technologies and AI may be wrong, or the accuracy rate may be significantly reduced depending on the purpose of the system. Using a speech recognition conversational system to practice tone pronunciation, it is likely that the correct tones and tonal errors will be ignored to make a conversation. In the studies of Chinese tone recognition using machine learning, the tone of monosyllable words was used as training data, and the method and accuracy of tone recognition were examined.

The highest accuracy (95.53%) was obtained using the convolutional neural network (CNN) and the mel-frequency cepstral coefficients (MFCCs) for tone pronunciation of 4500 syllables by 125 children aged 3 to 10. This study [34] showed that tone recognition by machine learning is possible; however, there are some shortcomings in learning tones with the tone recognition system. This study used children's speech, which had limitations on the pitch range, and the spectrogram and MFCCs of pronounced monosyllabic words in the paper showed that the third tone was a dipping-rising tone (Figure 4). In related works, as mentioned in Section 2.2, training with disyllabic words is more effective than training with monosyllabic words [23], and teaching the dipping-rising tone may cause pronunciation errors [24]. Furthermore, as the tones are made up of differences in relative pitch patterns, we considered that it is necessary to design pronunciation tasks that promote tone categorization of learners.

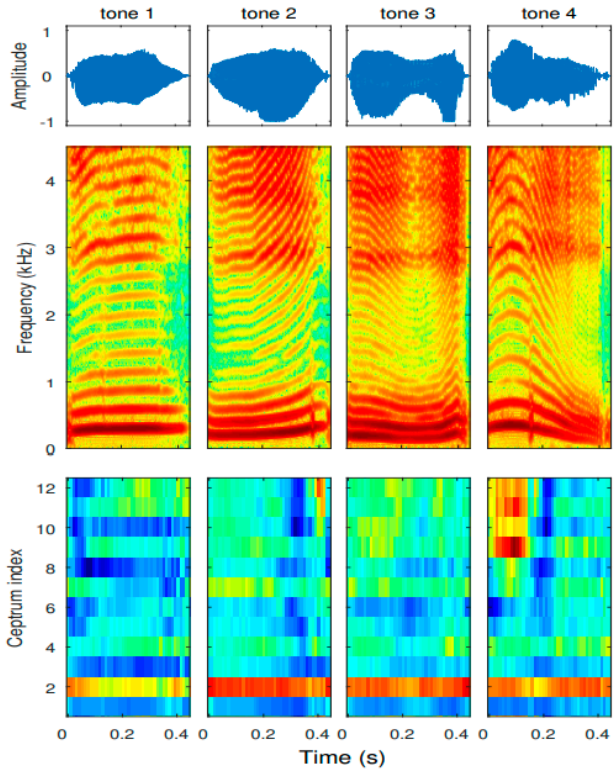

Figure 4. Machine learning data from previous studies on tone recognition [34]. Top: Time waveforms; Middle: Spectrograms; Bottom: Mel-frequency cepstral coefficients (MFCCs).

A study [35] that used the training data fetched and selected from the pronunciation of 590 short sentences by two pairs of men and women obtained an accuracy of 87.6% using a CNN. The accuracy of tone recognition is strongly related to the data used for machine learning, and we aimed to design pronunciation tasks that improve the abilities required for tone pronunciation rather than correctly recognizing tones. At the first stage of learning tone, we consider that learners should acquire the cognitive ability of tones, then learners should practice pronunciation with vocal cords and their tonal cognition. We used the designed voices as training data and made a pronunciation exercise using a CNN. The designed task for the pronunciation exercise was a disyllabic task that requested the pronunciation of tones 1, 2, 3, and 4 in the first syllable combined with tone 1 in the second syllable within 1.1 seconds and the length of the first syllable in less than 0.55 seconds. The data used for machine learning (CNN) were pronunciation data (a total of 547 items of audio, 60% for training and 40% for validation) of four men and three women who followed the rules of the pronunciation task for the training and test set. The classification accuracy was 98.18%, which was high due to the data pattern selection by rules and the features of the two-syllable tone. Figure 5 shows a part of the data used for machine learning.

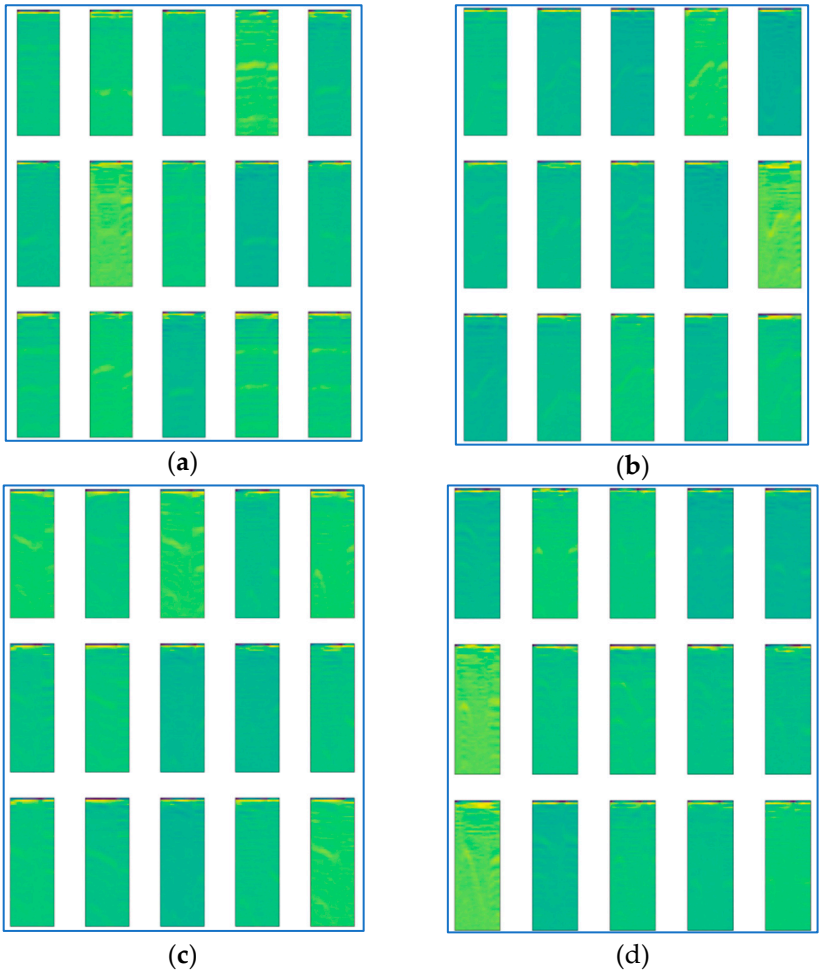

Figure 5. Training data for the pronunciation task designed in this study. (**a**) Combination of tone 1 and tone 1; (**b**) Combination of tone 2 and tone 1; (**c**) Combination of tone 3 and tone 1; (**d**) Combination of tone 4 and tone 1.

We considered that we should make tone learners from the machine learning aware of the processes of how a CNN classifies different tone pairs and the results of the classification by percentage. Learners can obtain key information to acquire tones from the process of machine learning, learning

with the training data (pronunciations follow rules of pronunciation task) visually (Figure 5) and through hearing and practicing with the visual form of their pronunciations and the classification results by percentage by using the CNN.

The design of the pronunciation task aimed to enable learners to be able to pronounce tones 1, 2, 3, and 4 combined with tone 1 correctly and individually with the speed of ordinary conversation (about 1 second). Even for native speakers, to get a high percentage of classification by a CNN with a pronunciation following the rules that are strictly judged requires practicing a little. It is assumed that practicing with the designed tone pronunciation task using a CNN helps to promote the tone understanding and categorization of learners and makes learners focus on controlling their vocal cords.

6. Conclusion and Future Tasks

This paper discusses the specialty of pronunciation learning in language learning and emphasizes the importance of auditory cognitive ability in the second language. It takes Chinese tone learning as an example and considers the cognitive load and the shortage of the general learning method's pronunciation practice by imitating the model voices of words, phrases, or sentences. Against this background, we focused on polysyllable perceptual training, emphasized the comparison of each tone, and designed tasks commensurate with the learner's cognitive ability using the small step strategy. Furthermore, we considered the shortness of sensory memory and optimized the interface and operation design of the perceptual training application, compared with the imitation practice app. As a result, the tone pronunciation correct answer rate of users was 30% to 36% higher in two basic pronunciation tasks and 12% higher in the phrase pronunciation task after use for a total of three hours. In the results of the paired sample t-test, there was a statistically significant difference in the correct answer rate of the pronunciation tasks mentioned above from learning with the imitation practice app and the perceptual training application ($t(9) = 5.78$, $p < 0.001$, Cohen's $d = 1.83$, $t(9) = 6.22$, $p < 0.001$, Cohen's $d = 1.97$ and $t(9) = 3.33$, $p = 0.009$, Cohen's $d = 1.05$).

However, the implementation of the tone pronunciation practice function did not help perceptual training app users to improve their tone pronunciation ability. Furthermore, practicing pronunciation could be the cause of tonal pronunciation errors. Therefore, we considered using machine learning to support tone pronunciation practice. We aimed to use machine learning to design pronunciation tasks rather than recognize tones correctly, as is the case in existing research. Furthermore, in addition to the prediction results of the trained model, the percentage of classification, the training data set, and the visualized data were presented to the learner as feedback for learning. We concluded that machine-learning-assisted tone pronunciation practice is feasible.

In the future, it is necessary to consider the instructional design of pronunciation practice based on the difficulty of tone combination, error patterns, phonemes, etc. Furthermore, it is also necessary to optimize the interfaces and operation design or to use cloud computing to minimize the computing time and extraneous cognitive load. On the other hand, there are studies showing that the individual differences between people and the data of language processing (in Mandarin Chinese) can be used to predict people's ability, which helps in understanding human cognition and creates navigation by machine learning [36,37]. Classifying learners automatically according to their phonological ability and providing navigation considering their individual differences using machine learning are also future issues.

Author Contributions: Conceptualization, M.S.K. and A.I.; methodology, M.S.K. and A.I.; software, M.S.K.; validation, M.S.K.; formal analysis, M.S.K. and A.I.; investigation, M.S.K. and A.I.; resources, A.I.; data curation, M.S.K. and A.I.; writing—original draft preparation, M.S.K.; writing—review and editing, A.I.; visualization, M.S.K.; supervision, A.I.; project administration, A.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: I would like to express my gratitude to the students who participated in the experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Snow, C.E.; Hoefnagel-Hoehle, M. The critical period for language acquisition: Evidence from second language learning. *Child Dev.* **1978**, *49*, 1114–1118. Available online: https://www.jstor.org/stable/1128751?read-now=1&seq=1 (accessed on 25 December 2019). [CrossRef]
2. McLaughlin, B. Myths and Misconceptions About Second Language Learning: What Every Teacher Needs to Unlearn. *Educ. Pract. Rep.* **1992**, *5*, 1–18.
3. Keeley, T.D. Is a Native-like Accent in a Foreign Language Achievable? Examining Neurological, Sociological, Psychological, and Attitudinal Factors. *KEIEIGAKURONSHU* **2016**, *26*, 59–92.
4. Bock, K.; Levelt, W.J.M. *Handbook of Psycholinguistics*; Academic Press: San Diego, CA, USA, 1994; pp. 945–984. Available online: https://pdfs.semanticscholar.org/c03e/bb859d84446fb88e5f35da16e67b235374da.pdf?_ga=2.244264455.322676749.1581925134-1162886882.1568556778 (accessed on 16 February 2020).
5. Griffin, Z.M.; Ferreira, V.S. Properties of Spoken Language Production. 2006. Available online: https://doi.org/10.1016/B978-012369374-7/50003-1 (accessed on 17 February 2020).
6. Altenberg, E.P.; Cains, H.S. The effects of phonotactic constraints on lexical processing in bilingual and monolingual subjects. *J. Verbal Learn. Verbal Behav.* **1983**, *22*, 174–188. Available online: https://vdocuments.mx/the-effects-of-phonotactic-constraints-on-lexical-processing-in-bilingual-and.html (accessed on 17 February 2020). [CrossRef]
7. Deutsch, D. Speaking in Tones. Scientific American Mind. 2007. Available online: https://www.researchgate.net/publication/44884565_Speaking_in_Tones (accessed on 16 February 2020).
8. Chao, Y.R. *Language Problems*; The commercial Press: Beijing, China, 1980; pp. 59–81.
9. Chen, H.Y. The Study of Mandarin Chinese Phonetic Errors and Teaching Strategies of Vietnamese Spouses in Taiwan. 2007. Available online: http://paperupload.nttu.edu.tw/CA4BD71DBC8AB41/e8f07f5da1cb3336.pdf (accessed on 10 October 2018).
10. Tran Thi, K.L. Error Analysis of Mandarin Tones from Vietnamese Learners. 2005. Available online: http://etds.lib.ntnu.edu.tw/cgi-bin/gs32/gsweb.cgi?o=dstdcdr&s=id=%22G0069124031%22.&#XXXX (accessed on 10 October 2018).
11. Nguyen Thi, N.T. A Case Study of Mandarin Mandarin Chinese Tone Training: Error Corrections in Pronunciation for Beginning Vietnamese Learners. 2016. Available online: https://hdl.handle.net/11296/7bk39j (accessed on 11 February 2017).
12. Cheng, L.Y. The Perception and Production of Mandarin Tones by L2 Learners. 2014. Available online: https://hdl.handle.net/11296/fgsjr3 (accessed on 12 October 2017).
13. Zhang, K.J.; Chen, L.M. Tonal Errors of Japanese Students Learning Mandarin Chinese: A Study of Disyllabic Words. 2015. Available online: https://aclanthology.info/papers/O05-1009/o05-1009 (accessed on 20 July 2017).
14. Wu, H.W. A Study of Mandarin Chinese Tone Production by Polish Speakers. 2011. Available online: https://hdl.handle.net/11296/e9nbkb (accessed on 10 October 2018).
15. Kim, S. Analysis and Teaching Research on Mandarin Chinese Tone Errors of Korean Students. 2005. Available online: http://etds.lib.ntnu.edu.tw/cgi-bin/gs32/gsweb.cgi?o=dstdcdr&s=id=%22N2005000097%22.&searchmode=basic (accessed on 10 October 2018).
16. Lin, L. An Analysis of the Tone Errors of Hong Kong Students Who Are Studying in Taiwan. 2017. Available online: http://etds.lib.ntnu.edu.tw/cgi-bin/gs32/gsweb.cgi?o=dstdcdr&s=id=%22G060285014I%22.&searchmode=basic (accessed on 10 October 2018).
17. Takamatsu, K. A Case Study of Mandarin Mandarin Chinese Tone Combination Training: Error Corrections in Pronunciation for An Aged Japanese Learner. 2018. Available online: https://hdl.handle.net/11296/4seuaz (accessed on 10 October 2018).
18. Lin, H.Y. Mandarin Tonal Acquisition of Novice Japanese Learners. 2007. Available online: http://etds.lib.ntnu.edu.tw/cgi-bin/gs32/gsweb.cgi?o=dstdcdr&s=id=%22GN0693240217%22.&searchmode=basic (accessed on 10 October 2018).
19. Jiang, R.H. A Study on Mandarin Disyllabic Tones of German Learners. 2012. Available online: http://etds.lib.ntnu.edu.tw/cgi-bin/gs32/gsweb.cgi?o=dstdcdr&s=id=%22GN0697800194%22.&searchmode=basic (accessed on 10 October 2018).

20. Tsai, P.L. Perception and Production of Mandarin Chinese Lexical Tone by Adult English Speaking Learners. 2008. Available online: http://etd.lib.nctu.edu.tw/cgi-bin/gs32/hugsweb.cgi?o=dnthucdr&s=id=%22GH000934709%22.&searchmode=basic (accessed on 10 October 2018).
21. Li, C.C. Error Analysis and Remedial Instruction of Mandarin Chinese Tones—A Study on American Learners of Mandarin. 2010. Available online: http://hdl.handle.net/11296/3wh27p (accessed on 12 April 2017).
22. Pelzl, E.; Lau, E.F.; Guo, T.; DeKeyser, R. Advanced second language learners' perception of lexical tone contrasts. *Stud. Second Lang. Acquis.* **2019**, *41*, 59–86. Available online: https://doi.org/10.1017/S0272263117000444 (accessed on 12 October 2019). [CrossRef]
23. Wu, N.H. Initiating Foreign Students into Mandarin Mandarin Chinese Tone Pairs and Its Effects on Tone Acquisition. 2015. Available online: https://hdl.handle.net/11296/3b65pc (accessed on 20 February 2017).
24. Zhang, P.S. An Analysis of Tone Errors Made by Japanese Learners of Mandarin in Disyllabic Words. 2013. Available online: https://hdl.handle.net/11296/f34wqa (accessed on 11 February 2017).
25. Kan, M.S.; Ito, A. A Study on Learning Method to Improve Mandarin Chinese Pronunciation for Native Japanese Speaker, 2017. *IEICE Tech. Rep.* **2017**, *117*, 19–24.
26. Umemoto, T. *Course Psychology 7 Memory*; University of Tokyo Press: Tokyo, Japan, 1969; pp. 165–170.
27. Yin, L.; Li, W.; Chen, X.; Anderson, R.; Zhang, J.; Shu, H.; Jiang, W. The role of tone awareness and pinyin knowledge in Chinese reading. *Writ. Syst. Res.* **2011**, *3*, 59–68. Available online: https://www.researchgate.net/publication/239788061_The_role_of_tone_awareness_and_pinyin_knowledge_in_Chinese_reading (accessed on 18 February 2020). [CrossRef]
28. Sweller, J. Cognitive Load During Problem Solving: Effects on Learning. *Cogn. Sci.* **1988**, *12*, 257–285. [CrossRef]
29. Krashen, S. We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Mod. Lang. J.* **1989**, *73*, 440–464. [CrossRef]
30. Kan, M.S.; Ito, A.; Hatano, H. Development and Evaluation of A Mandarin Chinese Tone Perceptual Training App. In Proceedings of the IEEE Conference on e-Learning, e-Management and e-Services, Langkawi, Malaysia, 20–22 November 2018; pp. 40–45.
31. Wang, Y.; Spence, M.M.; Jongman, A.; Sereno, J. Training American listeners to perceive Mandarin tone. *J. Acoust. Soc. Am.* **1999**, *106*, 3649–3658. [CrossRef] [PubMed]
32. Li, M.; DeKeyser, R. Perception Practice, Production Practice, and Musical Ability in L2 Mandarin Tone-word Learning. *Stud. Second Lang. Acquis.* **2017**, *39*, 593–620. Available online: https://doi.org/10.1017/S0272263116000358 (accessed on 10 October 2019). [CrossRef]
33. Hsia, A.W. Training Non-Tonal Speakers in the Perception and Production of Mandarin Tones in Disyllabic Words. 2010. Available online: https://hdl.handle.net/11296/8xez8w/ (accessed on 12 October 2017).
34. Chen, C.; Razvan, B.; Li, X.; Liu, C. Tone Classification in Mandarin Mandarin Chinese Using Convolutional Neural Networks [C]//Conference of the International Speech Communication Association. 2016. Available online: https://pdfs.semanticscholar.org/c36a/d2e471625046adfe797876c873e4818b5e1b.pdf?_ga=2.181750500.296250888.1576671934-1162886882.1568556778 (accessed on 14 December 2019).
35. Shen, L.J.; Wang, W. Fusion Feature Based Automatic Mandarin Chinese Short Tone Classification. *Tech. Acoust.* **2018**, *37*, 167–174. Available online: http://sxjs.cnjournals.cn/ch/reader/create_pdf.aspx?file_no=20180213&flag=1&journal_id=sxjs&year_id=2018 (accessed on 20 December 2019).
36. Stella, M.; Kenett, Y.N. Viability in Multiplex Lexical Networks and Machine Learning Characterizes Human Creativity. 2019. Available online: https://pdfs.semanticscholar.org/1659/e2f39263337c8d49e75ded719f00b3a5aeeb.pdf?_ga=2.248849157.322676749.1581925134-1162886882.1568556778 (accessed on 20 February 2020).
37. Neergaard, K.D.; Huang, C.R. Constructing the Mandarin Phonological Network: Novel Syllable Inventory Used to Identify Schematic Segmentation. *Complexity* **2019**, *2019*, 1–21. Available online: https://www.researchgate.net/publication/332596249_Constructing_the_Mandarin_Phonological_Network_Novel_Syllable_Inventory_Used_to_Identify_Schematic_Segmentation (accessed on 20 February 2020). [CrossRef]