Large-Scale Characterization of Mandarin Pronunciation Errors Made by Native Speakers of European Languages

Nancy F. Chen[†], *Vivaek Shivakumar*[*], *Mahesh Harikumar*[§], *Bin Ma*[†], *Haizhou Li*[†]

[†]Institute for Infocomm Research, Singapore [§]Nanyang Technological University, Singapore
[*]Massachusetts Institute of Technology, Cambridge, MA, USA

{nfychen, mabin, hli}@i2r.a-star.edu.sg, vivaek@mit.edu, mahesh004@e.ntu.edu.sg

Abstract

In this work, we quantify common tonal and phonetic errors made by second language learners of Mandarin Chinese. Pronunciation patterns of 300 native speakers of European languages are analyzed. Tonal errors (30.56%) are found to be more prevalent than phonetic ones (8.71%). Common errors include overemphasis of Tone 3 and inadequate aspiration of affricate consonants. Decision tree clustering was used to further characterize these error patterns with their tonal and phonetic context. Our findings are potentially useful in second language education and in computer assisted language learning.

Index Terms: computer-aided pronunciation training, second language (L2) acquisition, , lexical tones, aspiration, affricate, Mandarin Chinese

1. Introduction

An increasing number of people are learning Mandarin Chinese as a second language (L2), yet limited research has focused on quantitatively characterizing common mispronunciations on a large-scale basis. In this work, we attempt to fill in this gap.

It is often qualitatively stated as conventional wisdom that L2 learners of Mandarin have difficulty acquiring affricate and fricative consonants [1, 2], which are peculiar to many European languages (except some Slavic languages like Polish which have retroflex and palatal fricatives [3]). However, few research studies have quantitatively characterized the mispronunciation patterns of these consonants.

In contrast, many L2 Mandarin studies focus on non-native lexical tone productions (e.g., [4, 5]), since most resource-rich languages like English lack lexical tones. Reference [4] evaluated how perceptual training helped improve English speakers' tone production. Reference [5] analyzed tone errors and its relationship with prosodic phrasing in L2 Mandarin. These studies provide a foundation of how to help L2 learners acquire Mandarin Chinese.

Our work extends past studies by (1) examining a larger set of speakers (an order of magnitude more than [4, 5]), and (2) characterizing L2 Mandarin mispronunciations of both tones and phones using tonal and phonetic context. The latter extension is based on our knowledge that native tone productions in tonal languages are affected by tonal context [6], and that dialect studies have shown that acoustic implementations of phonemes can often be characterized using phonetic context [7]. Our analyses help predict when and where non-native mispronunciations are more likely to occur. These findings can potentially help refine computer-assisted language learning software systems (e.g., [8]) or classroom exercises.

2. Experiment Design

2.1. Mandarin Phonology Background

Each Chinese character is spoken as one syllable, which consists of an *initial* and a *final*, or merely just a final. The initial is a consonant; the final can consist of vowel(s) or vowel(s) followed by a nasal. Each syllable is also encoded by a *tone*.

Tone is the use of pitch in speech. Mandarin Chinese uses *lexical tones* to encode semantics; i.e., a change in tone can change the meaning of a lexical term (e.g., character, word).

We elaborate on lexical tones in Mandarin below: Tone 1 (hight-level): a steady high pitch, as if it were being sung instead of spoken; Tone 2 (high-rising): a high-rising pitch, like the utterance-final intonation of a question in English (e.g., What?!); Tone 3 (dipping): the pitch lowers and then rises within the same syllable; Tone 4 (falling): a short tone with a sharp fall in pitch, similar to curt commands in English (e.g., Stop!). Tone 5 (or the zeroth tone) is a neutral tone, or viewed as *lack of tone*; it is analogous to an unstressed syllable.

2.2. Phonetic Symbols

Pinyin is a phonetic system used to transcribe Chinese characters into Latin script. In Pinyin, syllables are separated by white space, and the tone of a syllable is appended after the final. For example, the word *afternoon* is phonetically represented as 'xia4 wu3', where 'x' and 'w' are the initials, 'ia' and 'u' are the finals, 4 and 3 are the tones. In this paper, Pinyin symbols are encased in single quotes (e.g., 'z' is the voiceless, unaspirated, alveolar affricate in Mandarin); International phonetic alphabet (IPA) symbols are encased in "/ /" or "[ ]" (e.g., /z/ is the voiced alveolar fricative).

2.3. iCALL (I^2R Computer Assisted Language Learning) Corpus

Three hundred beginner learners of Mandarin Chinese were asked to read 300 Pinyin prompts, including 200 words (each word is at least two characters) and 100 sentences. Each speaker received a different set of utterances. Their non-native speech was sampled at 16 kHz, encoded in 16 bit pulse-code modulation (PCM), recorded in quiet office rooms, and transcribed in Pinyin through perceptual listening tasks. These transcriptions represent the *surface pronunciation*, while the Pinyin prompts are defined as the *underlying ground-truth pronunciation*.

The corpus was split into the training, development and test sets; these sets contained 180, 60, and 60 speakers respectively. Gender and age were balanced across all three sets. The speakers' native languages were of European origin: 52% Germanic

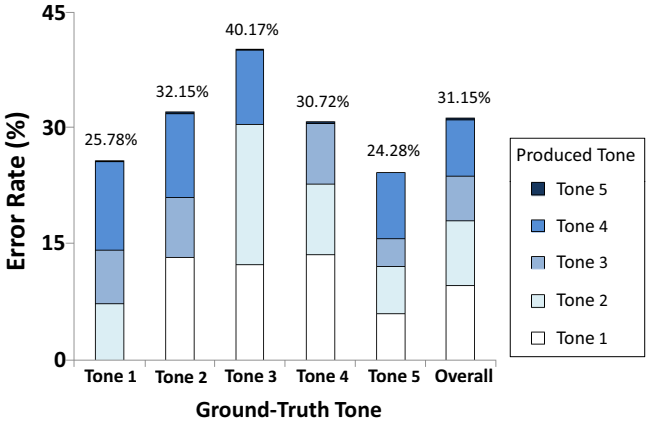

Figure 1: *Lexical tone error distribution.*

Table 1: *Non-distinctive feature attributes used in Figure 2 and Figure 5. L2 denotes second language.*

Attribute	Description
UttInitial	The syllable is at an utterance initial position.
UttFinal	The syllable is at an utterance final position.
SononrantInitial	The initial syllable is a sonorant (i.e., 'w', 'y', 'l', 'r', 'm', 'n' in Pinyin).
NasalFinal	The final is a nasal (i.e., 'n', 'ng' in Pinyin).
FinalVelarNasal	The final is a velar nasal (i.e., 'ng' in Pinyin).
VowelFinal	The final is a vowel.
Short	Syllable length ≤ 3 characters in Pinyin.
1*	The L2 learner produced Tone 1.
3*	The L2 learner produced Tone 3.

(e.g., English, German); 32% Romance (e.g., French, Spanish, Italian); 15% Slavic (e.g., Russian). Among individual languages, English has the largest number of speakers (119 speakers; 40% of total data).

Preliminary analysis showed that the general trends of the error patterns were insensitive to the speaker's native language (exceptions of phonetic errors are discussed in Section 3.2.3). Therefore we collapsed the speaker groups of native languages to increase the sample size of our study.

2.4. Decision Tree Clustering

Decision tree clustering [9] is a customary approach in automatic speech recognition to group phones with similar acoustic properties together. We chose to use decision tree clustering because it provides intuitive results that are linguistically informative to second language learners and educators.

For each underlying tone, a set of attributes were iteratively chosen through decision tree clustering to characterize its corresponding non-native surface tones. Attributes include distinctive features [10]. Attributes not derived from standard distinctive features were also used, including utterance position of syllable in question, neighboring syllables' underlying tones, and neighboring syllables' surface tones, which were derived from observations at Mandarin classes at MIT. In Table 1 we only list non-distinctive-feature attributes determined from decision tree clustering results in Figures 2 and 5. Similar to characterizing tones, we used decision tree clustering to determine attributes characterizing surface pronunciations for each underlying initial or final.

3. Experiments

3.1. Lexical Tone Errors

3.1.1. Context-Independent Error Patterns

We first analyze the tone errors by aligning the underlying ground-truth tone sequences with the transcribed surface tone sequences. The mispronunciation rates for each tone are listed in Figure 1. When collapsing data from all lexical tones, nearly 1 out of every 3 syllables (error rate: 30.56%) are produced incorrectly. Among the 5 tones, Tone 3 is the most challenging for Mandarin learners (40.17% error rate), while Tone 1 and Tone 5 are the easiest (25.78% and 24.28% error rate, respectively). This trend is similar to that reported in [4]. Among the Tone 3 errors, the majority (46%) are mispronounced as Tone 2, which corresponds to [4]. Over 40% of the Tone 2 and Tone 4 errors are mispronounced as Tone 1; 45% of Tone 1 errors and 36% of Tone 5 errors are mispronounced as Tone 4.

3.1.2. Context-Dependent Error Patterns

In this section, we further investigate whether tonal and phonetic context can help obtain more refined characterizations of the mispronunciations.

A decision tree was grown for each of the five tones. All syllables that belong to a particular underlying tone are the samples at the root node. The samples are recursively split into leaf nodes to reduce entropy. The surface tones are the classification labels. The stop criteria for splitting were based on minimal split size and minimal leaf node size. The optimal values of these parameters leading to the lowest classification error were tuned on the development set.

Figure 2 lists the top pronunciation error patterns on the test set. The descriptions of the attributes used are listed in Table 1. Below we elaborate on some general trends.

Confusion between Tone 1 and Tone 4. Underlying Tone 1 and Tone 4 are most likely mispronounced as Tone 4 and Tone 1 respectively, regardless of phonetic/tonal context, as seen in the previous section. From the clustering results, we were able to identify that there are certain exception cases: Error Patterns T1a and T1b specify cases where Tone 1 is most likely mispronounced as Tone 3; in all other conditions, Tone 1 is mostly likely mispronounced as Tone 4. Error Pattern T4a specifies when Tone 4 is most likely mispronounced as Tone 2; in all other conditions, Tone 4 is most likely misproduced as Tone 1.

Over emphasizing Tone 3. In spoken Mandarin, when a Tone 3 syllable is not at the end of an utterance, Tone 3 is often realized as *half-third*: the pitch does not rise again after falling. At the end of an utterance, however, Tone 3 is usually acoustically implemented as a full third tone, which ends with a rising pitch. Beginner learners might therefore over exaggerate the rising portion of Tone 3 when it is at the end of an utterance, making Tone 3 sound like Tone 2 (see Error Patterns T3a and T3b in Figure 2). This exaggeration occurs less if the preceding tone was already produced as Tone 3, since consecutive Tone 3's are difficult to pronounce. In these cases, the latter Tone 3 is more likely to be produced as Tone 1, despite being in the final position of an utterance (see Error Pattern T3c in Figure 2).

It has been hypothesized that the reason Tone 3 is difficult to acquire is due to the novelty of the required pitch manipulation [4]. Even after learners are able to perceive Tone 3, they still might not be able to produce it correctly [4]. The difficulty of acquiring Tone 3 is not a proprietary to non-native speakers. Tone 2 and Tone 3 have been found to be confusing in both first

No.	Error Pattern	Probability (No. of samples)	Examples
T1a	[+1] → [+3] / [−3˚][ ][+1˚]	0.403 (2381)	莫斯科 Moscow mo4 si1 ke1 → mo4 si3 ke1
T1b	[+1] → [+3] / [+*UttInitial*][+1˚]	0.519 (822)	八億 eight hundred million ba1 yi4 → ba3 yi1
T1c	Default (all contexts other than T1a, T1b): [+1] → [+4]	0.48 (11670)	喝了 drink/drank he1 le5 → he4 le5
T2a	[+2] → [+1] / [ ][+*NasalFinal*][ ]	0.5 (6775)	嚴肅 serious yan2 su4 → yan1 su4
T2b	[+2] → [+3] / [ ][+*SonorantInitial* −*NasalFinal*][+1˚]	0.521(654)	離開 leave li2 kai1 → li3 kai1
T3a	[+3] → [+2] / [−3˚][+*SonorantInitial* +*UttFinal*]	0.562 (4245)	新領域 new areas xin1 ling3 yu4 → xin1 ling2 yu4
T3b	[+3] → [+2] / [−3˚][+*UttFinal*]	0.642 (2960)	下午 afternoon xia4 wu3 → xia4 wu2
T3c	[+3] → [+1] / [+3˚]___	0.428 (2996)	辦法 method ban4 fa3 →ban3 fa1
T4a	[+4] → [+2] / [ ][+*SonorantInitial* +*UttFinal*][ ]	0.544 (1490)	八日 the eighth (of the month) ba1 ri4 → ba1 ri2
T4b	Default (all contexts other than T4a) : [+4] → [+1]	0.45 (25048)	不可 cannot bu4 ke3 →bu1 ke3
T5a	[+5] → [+2] / [ ][+*NasalInitial* −*VowelFinal*][ ]	0.570 (509)	他們的 their ta1 men5 de5 → ta1 men2 de5
T5b	[+5] → [+4] / [+3˚][+*VowelFinal*][ ]	0.501 (589)	喝了 drink/drank he1 le5 → he3 le4

Figure 2: *Context-dependent error patterns for tones. Format is adopted from phonological rules [15, 16]: [underlying ground-truth speech segment] → [surface speech segment] / [left context description] [segment of interest description] [right context description], where speech segment refers to a syllable, an initial, or a final. Probability denotes the conditional probability of the substitution error (in the test set) given the context listed in the Error Pattern column.*

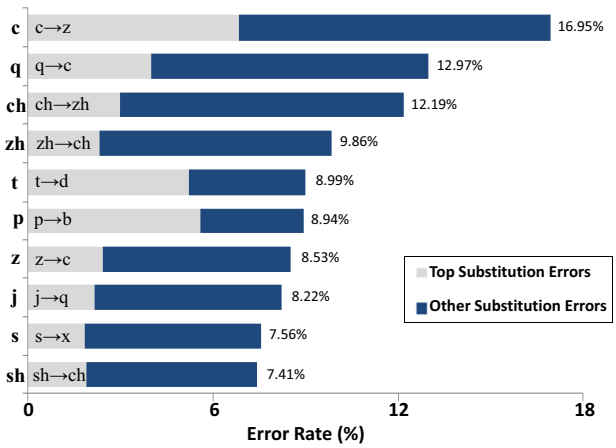

Figure 3: *Most mispronounced initials and their corresponding top substitution errors are lightly shaded bars (when disregarding phonetic context). Format of error: underlying ground-truth initial → surface initial.*

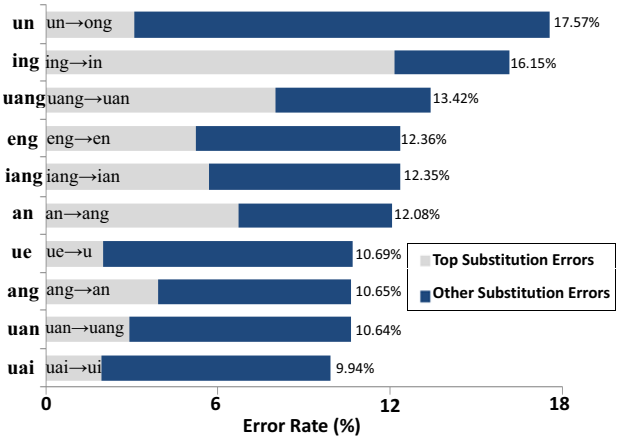

Figure 4: *Most mispronounced finals and their corresponding top substitution errors are lightly shaded bars (when disregarding phonetic context). Format of error: underlying ground-truth final → surface final.*

language acquisition [11, 12] and second language acquisition [13, 14].

Rising tone at utterance final. When non-native speakers lack confidence in their pronunciation, which is not uncommon, there is a tendency to raise their pitch at the end of utterances, leading to the production of Tone 2 for the last syllable of the utterance. (See Error Patterns T4a, T3b in Figure 2.)

Tone 5 elongation. Most syllables specified in Error Pattern T5a in Figure 2 are 'men5', the plurality particle in Mandarin Chinese. The tendency to produce 'men5' as 'men2' could be due to the longer syllable length caused by the presence of a sonorant initial and sonorant final. (A syllable like 'men5' is longer in duration than other syllables because most Mandarin syllables do not include nasals in the final, and initials are optional in Mandarin syllable structures.)

Common tone-pair mapping. The underlying tone pair (2,4) to is often misproduced to surface tone pair (3,1). Error Patterns T2b and T4b in Figure 2 could partially explain this trend. In addition, if Tone 2's pitch level is too low, it is hard to produce the following Tone 4 with an even lower pitch. Maintaining a similar pitch level causes the perception of Tone 1.

3.2. Phonetic Errors

3.2.1. Context-Independent Error Patterns

The phonetic error rate is 8.71%. Phonetic error rate is defined as the total number of initial or final errors over the total number of syllables/characters. If we break down phonetic errors, 6.51% initials and 7.69% finals were mispronounced in the total 292,756 syllables in the training set.

Figure 3 shows the most mispronounced initials and their corresponding top substitution errors. We see that 7 out of 10 of these top substitution errors are related to aspiration. This trend is possibly due to the fact that while aspiration is an important feature defining many phonetic differences in Mandarin, it is not necessarily so in European languages.

Figure 4 shows the most mispronounced finals and their corresponding top substitution errors. We see that 8 out of 10 of these top substitution errors are related to nasals, and among these errors more than 60% are fronting of velar nasals, implying that velar nasals are more challenging to produce correctly.

No.	Error Pattern	Probability (No. of samples)	Example
Z1	$\begin{bmatrix}-aspiration\\+alveolar\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}-aspiration\\+retroflex\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}-diphthong\end{bmatrix}$	0.343 (510)	zu1 → zhu1 祖 → 猪 rent → pig
Z2	$\begin{bmatrix}-aspiration\\+alveolar\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}+aspiration\\+retroflex\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}+diphthong\end{bmatrix}$	0.361 (319)	zai4→cai4 在→菜 at→vegetables
Z3	$\begin{bmatrix}-aspiration\\+alveolar\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}+aleveolar\\+fricative\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}+diphthong\end{bmatrix}$	0.354 (319)	zai4 → sai4 在 → 赛 At → game
J1	$\begin{bmatrix}-aspiration\\+palatal\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}+aspiration\\+palatal\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}-round\end{bmatrix}$	0.456 (528)	ji1 →qi1 鷄 →妻 chicken → wife
J2	$\begin{bmatrix}-aspiration\\+palatal\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}-aspiration\\+retroflex\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}+round\\-glide\end{bmatrix}$	0.481 (283)	ju4 → zhu4 聚 → 住 gather → live
J3	$\begin{bmatrix}-aspiration\\+palatal\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}+aspiration\\+palatal\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}+round\\+glide\end{bmatrix}$	0.303 (188)	jiao1 → qiao1 交 → 敲 pay → knock
J4	$\begin{bmatrix}-aspiration\\+palatal\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}-aspiration\\+retroflex\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}+round\\+glide\end{bmatrix}$	0.229 (188)	juan1 → zhuan1 捐 → 專 donate → special
J5	$\begin{bmatrix}-aspiration\\+palatal\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}\ \end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}+round\\+glide\end{bmatrix}$	0.202 (188)	jiao1 → yao1 交 → 腰 pay → waist
Q1	$\begin{bmatrix}+aspiration\\+palatal\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}-aspiration\\+palatal\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}-round\end{bmatrix}$	0.561 (467)	qi1 →ji1 妻 →鷄 wife → chicken
Q2	$\begin{bmatrix}+aspiration\\+palatal\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}+aspiration\\+retroflex\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}+round\\-triphthong\end{bmatrix}$	0.663 (528)	qun2 → chun2 裙 → 純 skirt → pure
Q3	$\begin{bmatrix}+aspiration\\+palatal\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}-aspiration\\+palatal\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}+round\\+triphthong\end{bmatrix}$	0.646 (79)	qiao1 → jiao1 敲 → 交 knock → pay
Ch1	$\begin{bmatrix}+aspiration\\+retroflex\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}+aspiration\\+palatal\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}-FinalVelarNasal\\+short\end{bmatrix}$	0.398 (364)	chi1 → qi1 吃 →七 eat → seven
Ch2	$\begin{bmatrix}+aspiration\\+retroflex\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}-aspiration\\+retroflex\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}+FinalVelarNasal\end{bmatrix}$	0.479 (259)	chong1 → zhong1 衝 → 中 rush → middle
Sh1	$\begin{bmatrix}+retroflex\\+fricative\end{bmatrix} \rightarrow \begin{bmatrix}+palatal\\+fricative\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}+short\\-round\end{bmatrix}$	0.506 (405)	sha1 → xia1 沙 → 蝦 sand → shrimp
Sh2	$\begin{bmatrix}+retroflex\\+fricative\end{bmatrix} \rightarrow \begin{bmatrix}+aspiration\\+retroflex\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}-short\end{bmatrix}$	0.398 (352)	shuai4 → chuai4 帥 → 踹 handsome → kick
Zh1	$\begin{bmatrix}-aspiration\\+retroflex\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}+aspiration\\+retroflex\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}-short\\-FinalVowel\end{bmatrix}$	0.486 (434)	zhuang1 → chuang1 裝 → 窗 pretend → window
Zh2	$\begin{bmatrix}-aspiration\\+retroflex\\+affricate\end{bmatrix} \rightarrow \begin{bmatrix}-aspiration\\+alveolar\\+affricate\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}-short\\+FinalVowel\end{bmatrix}$	0.367 (139)	zhao1 → zao1 招 → 糟 recruit → spoil
X1	$\begin{bmatrix}+palatal\\+fricative\end{bmatrix} \rightarrow \begin{bmatrix}+retroflex\\+fricative\end{bmatrix} / \_\_\_\_\_ \begin{bmatrix}+short\\+round\end{bmatrix}$	0.563 (286)	xu1 → shu1 需 → 書 need → book

Figure 5: *Context-dependent error patterns for initials. The representation used is adopted from phonological rules [15, 16]: [underlying ground-truth initial] → [surface initial] / [left context description] _____ [right context description]*

3.2.2. Context-Dependent Error Patterns

The decision tree clustering implementation details are similar to those in Section 3.1.2. After decision tree clustering, we found that the error patterns of finals were context independent, while those of initials are context-dependent. Common initial errors obtained through decision tree clustering are listed in Figure 5. We discuss these mispronunciation trends in more detail below.

Aspiration/De-aspiration: see Error Pattern No. Z2, J1, J3, Q1, Q3, Ch2, Zh1 in Figure 5. Among these 18 mispronunciation patterns, 1/3 are related to improper aspiration. Many of the native languages of the speakers have no aspiration distinction among different phonemes. Some languages do not even have affricate consonants at all (e.g., French and Spanish). Therefore, these speakers might have difficulty controlling aspiration appropriately. This aspiration error also occurs frequently among stops (/p/, /t/, /k/), especially for native speakers of Romance languages. (See Section 3.2.3.)

Frication: see Error Pattern No. Z3 in Figure 5. The Pinyin symbol 'z' represents the voiceless, unaspirated, alveolar affricate in Mandarin, while it represents the voiced alveolar fricative in many other Romanization schemes (e.g., English or IPA). This potential confusion across phonetic systems might make speakers pronounce the alveolar affricate ('z' in Pinyin) as an alveolar fricative ('s' in Pinyin) instead.

Backing: see Error Pattern No. J2, J4, Q2, X1 in Figure 5. Rounded vowels prompt the place of articulation to move backwards, which could turn a palatal affricate to its retroflex counterpart. It has also been reported that rounding might be an enhancing gesture for the retroflex feature [17], which potentially explains why palatal affricates before a rounded vowel becomes retroflex.

For Error Pattern No. Q2, the four syllables that fit this context are 'qu', 'que', 'quan', and 'qun'. The 'u' in these syllables is the close/high front rounded vowel /y/ in IPA as opposed to the close/high back rounded vowel/u/ in IPA. If this front vowel /y/ is mispronounced, the aspirated palatal affricate consonant ('q' in Pinyin) easily becomes backed to the aspirated retroflex affricate ('ch' in Pinyin). Similar to its affricate counterparts, the palatal fricative ('x' in Pinyin), could also become retroflex ('sh' in Pinyin) due to the rounding of its following vowel.

Gliding: see Error Pattern No. J5 in Fg. 5. The unaspirated palatal affricate before the high-front-vowel/palatal-glide approaches the palatal approximant, when given sufficient frication or lack of clear articulation.

Fronting: see Error Patterns No. Ch1, Sh1 in Figure 5. The Pinyin symbol 'i' usually refers to the high front vowel /i/ in IPA. When the Pinyin symbol 'i' is followed by retroflex fricatives and affricates, 'i' is acoustically implemented as a mid-back vowel. However, L2 learners of Mandarin might still pronounce this vowel as the high front vowel /i/ in IPA, making the retroflex consonant before the vowel more fronted, which becomes a palatal consonant instead.

3.2.3. Native Language Dependent Errors

The native languages of the speakers generally did not affect the overall mispronunciation patterns. However, this was not the case for the de-aspiration of stop initials. Note that 'p' in Pinyin corresponds to the aspirated, voiceless, labial stop [p^h] in IPA, while 'b' in Pinyin corresponds to the unaspirated, voiceless, labial stop [p] in IPA. Aspirated stops ('p, t, k' in Pinyin) produced as unaspirated counterparts ('b, d, g' in Pinyin) were more than twice as likely from native speakers of Romance languages (7.00% error rate) than from those of other languages (2.96%). This discrepancy might be because there are only unaspirated stops in Romance languages [18], while there are both aspirated and unaspirated allophones for voiceless stops in Germanic languages [19]. For example, in American English /p/ is aspirated in words like *pray*, but unaspirated when preceded by an obstruent consonant as in *spray* [20]. It is interesting to note that though Slavic languages only have unaspirated stops [21], their native speakers made as few mistakes as those of Germanic languages.

4. Conclusions

We quantified Mandarin mispronunciation patterns of 300 native speakers of European languages. These speakers made more tonal errors (30.56%) than phonetic ones (8.71%). We characterized these errors using contextual information such as the distinctive features of neighboring syllables. Most error patterns (e.g., overemphasis of Tone 3, inadequate aspiration) were similar across native language groups, but not the de-aspiration of stop initials, which were more prevalent among Romance language speakers. To foster research in L2 acquisition and computer assisted language learning, we plan to further refine the iCALL corpus used in this work and publicly release it afterwards.

5. References

[1] Special Sounds in Mandarin Chinese: http://www.bbc.co.uk/schools/primarylanguages/mandarin/sounds/, last accessed, February 26, 2013.

[2] Chiu, C.-Y., Lia, Y.-F., Kulls, D., Mixdorff, H., Chen, S.-L. "A Preliminary Study on Corpus Design for Computer-Assisted German and Mandarin Language Learning", Speech Database and Assessments, Oriental COCOSDA, 2009.

[3] Jassem, W., "Polish", Journal of the International Phonetic Association 33 (1): pp. 103-107, 2003.

[4] Wang, Y., Jongman, A. and Sereno, J. A., "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training", Journal of Acoustical Society of America, Volume 113, Issue 2, pp. 1033–1043, 2003.

[5] Yang, C., "The Acquisition of Mandarin Prosody by American L2 Learners of Chinese as a Foreign Language", Ph.D. Thesis, the Ohio State University, 2011.

[6] Xu, Y., "Contextual Tonal Variations in Mandarin", Journal of Phonetics 25, pp. 61-83, 1997.

[7] Wells, J. C., "Accents of English", Cambridge University Press, 1982.

[8] Peabody M. and Seneff, S., "Towards Automatic Tone Correction in Non-native Mandarin," Proc. 5th International Symposium on Chinese Spoken Language Processing (ISCSLP), Kent Ridge, Singapore, December 2006.

[9] Quilan, J. R., "Induction of Decision Trees", Machine Learning 1, no. 1, pp. 81-106, 1986.

[10] Chomsky, N. and Halle, M., "The Sound Pattern of English", New York: Harper and Row, 1968.

[11] Li, C. N. and Thompson, S., "The acquisition of tone in Mandarin speaking children", Journal of Child Language 4. pp. 185–199, 1977.

[12] Clumeck, H., "The acquisition of tone", in G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Freguson, Child Phonology, Academic Press, New York, Vol I, 1980.

[13] Leather, J. "Speaker normalization in perception of lexical tone", Journal of Phonetics 11, pp. 373–382, 1983.

[14] Miracle, W. C., "Tone production of American students of Chinese: A preliminary acoustic study", Journal of Chinese Language Teachers Association 24, pp. 49–65, 1989.

[15] Goldsmith, J. A, "Phonological Theory", In John A. Goldsmith, The Handbook of Phonological Theory, Blackwell Handbooks in Linguistics. Blackwell Publishers, 1995.

[16] Hayes, B. "Introductory Phonology", Blackwell Textbooks in Linguistics, Wiley-Blackwell, 2009.

[17] Stevens, K.N., Li, Z., Lee, C.-Y., and Keyser, J. (2004), "A note on Mandarin fricatives and enhancement", In G. Fant, H. Fujisaki, J. Cao, and Y. Xu (eds.), From Traditional Phonology to Modern Speech Processing (pp. 393-404), Beijing: Foreign Language Teaching and Research Press, 2004.

[18] Tranel, B. "The sounds of French: an introduction (3rd ed.)", New York: Cambridge University Press. pp. 129130, 1987.

[19] Jessen, M., and Ringen, C. "Laryngeal features in German", Phonology 19.2, pp.189-218, Cambridge University Press, 2002.

[20] Roach, P., "English Phonetics and Phonology: A Practical Course, 4th Ed.", Cambridge: Cambridge University Press, 2009.

[21] Petrova, O., Plapp, R., Ringen, C. and Szentgyrgyi, S., "Voice and Aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish", The Linguistic Review, 2006.