INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

SingaKids-Mandarin: Speech Corpus of Singaporean Children Speaking Mandarin Chinese

Nancy F. Chen[1], *Rong Tong*[1], *Darren Wee*[2], *Peixuan Lee*[2], *Bin Ma*[1], *Haizhou Li*[1]
[1] Institute for Infocomm Research, Singapore
[2] National University of Singapore, Singapore
{nfychen, tongrong, mabin, hli}@i2r.a-star.edu.sg; {darren.wee, peixuan.lee}@u.nus.edu

Abstract

We present *SingaKids-Mandarin*, a speech corpus of 255 Singaporean children aged 7 to 12 reading Mandarin Chinese, for a total of 125 hours of data (75 hours of speech) and 79,843 utterances. This corpus is phonetically balanced and detailed in human annotations, including phonetic transcriptions, lexical tone markings, and proficiency scoring at the utterance level. The reading scripts span a diverse set of utterance styles, covering syllable-level minimal pairs, words, phrases, sentences, and short stories. We analyze the acoustic properties of Singaporean children. We also observe that while the lack of the neutral tone is the same for Singaporean adults and children, the phonetic pronunciation patterns in these two age groups differ: although Singaporean adults tend to front their retroflex, nasal, and palatal consonants, Singaporean children show both fronting and backing in these consonants. For future work, we plan to develop computer-assisted pronunciation training (CAPT) systems with *SingaKids-Mandarin*.
Index Terms: computer-assisted language learning (CALL)

1. Introduction

While most spoken language resources focus on adult speech, corpora for children have received far less attention. A growing number of children are exposed to advanced technology at a younger age today, yet few studies have examined the necessary resources to develop spoken language technology for children. In this work, we address this issue by surveying existing corpora on children's speech and presenting a new corpus, *SingaKids-Mandarin*, that complements existing resources in terms of size, language, and human annotations.

Primary students in Singapore are at least bilingual [1] due to Singapore's language policy: Although English is used for government administration, law, education, and communication across ethnic groups, every Singaporean student is mandated to learn one of the other three national languages (Malay, Mandarin, or Tamil).[1] At home or in daily affairs, other languages (Arabic, Cantonese, Hakka, Hindu, Hokkien, Punjabi, Teochew) might also be spoken. Due to Singapore's multilingual landscape, it is an interesting linguistic question of whether Mandarin speaking Singaporeans are considered native speakers of Mandarin. These multilingual speakers might possess linguistic characteristics that deviate from native Mandarin speakers who live in a monolingual environment (e.g., China, Taiwan). These speakers could also be viewed as advanced 2nd language learners of Mandarin who are exposed to Mandarin outside of the academic setting. In this work, we present a corpus for analyzing Mandarin spoken by multilingual Singaporean children and for developing spoken language learning technology.

[1] Taking other languages is possible under certain circumstances [2].

2. Existing Corpora on Children's Speech

Table 1 lists existing children's speech corpora and compare them with SingaKids-Mandarin. Among the existing corpora, we see that the majority of them are in English, 1/3 focuses on children before age 4, all have a trade-off between speaker diversity and the total duration and total number of utterances, and none have proficiency ratings. Children's speech has been documented to be more variable in speaking rate and vocal effort [3]. Children's speech is also known to be more prone to mispronunciations, ungrammatical expressions or use words that do not normally appear in adult speech [4]. Due to such a high variance in children's speech, it is desirable for a speech corpus to cover a large number of children speakers, each with sufficiently large amounts of data. In the next section, we delineate the corpus design strategy of *SingaKids-Mandarin*.

3. Corpus Design

3.1. Speakers

A total of 255 Singaporean primary school students of ages 7-12 were recruited. All speakers were bilingual or multilingual, speaking at least Mandarin and English. Students needed to be old enough to have learned *Hanyu Pinyin*, a phonetic system for romanizing Mandarin used in countries like China and Singapore. Hanyu Pinyin prompts were shown alongside with the Chinese characters to assist pronunciation. Since students might start entering puberty around age 12, there was some effort to minimize the number of male subjects in this age group to avoid complications of voice breaking. The speech recordings were recorded in quiet office rooms, sampled at 16 kHz, and encoded in 16 bit pulse-code modulation (PCM).

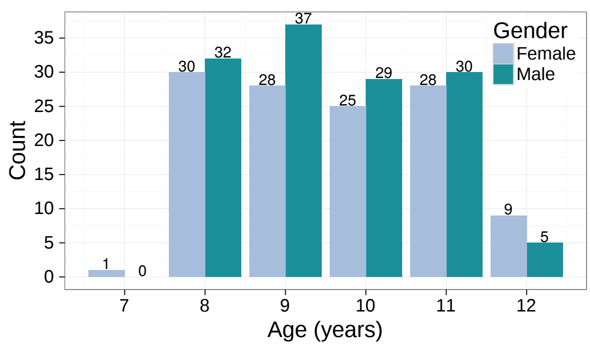

Figure 1: Demographic distribution of speakers in SingaKids-Mandarin by age and gender.

http://dx.doi.org/10.21437/Interspeech.2016-139

3.2. Lexical Content

The reading prompts were designed to have complete phonetic and syllable coverage, with special focus on commonly made phonetic and tonal errors, and a diverse range of utterance lengths ranging from single syllables to sentences and stories.

The recording was rolled out in three phases. The first phase focused on short stories and sentences that embedded commonly mispronounced words. The second phase focused on phonetic minimal pairs at the syllable level, tonal minimal pairs at the word level, and tone sandhi at the word and sentence level. The third phase focused on phonetic minimal pairs at the word level, reduplicated syllables, heteronyms (words with more than one pronunciations), and frequent mispronunciations from analyzing data from the first phase.

SingaKids-Mandarin is phonetically rich in that its phonetic frequency not only matches that of the natural phonetic distribution in Mandarin [19], but also include rare syllables to ensure complete coverage.

3.3. Human Annotations

Phonetic transcriptions were done in Hanyu Pinyin while tones were transcribed using numbers. The fluency scoring protocol was developed by two native Mandarin speakers who adapted their protocol from [20]. There are three levels of fluency scores: fluent (4), good (3), average (2), poor (1). Utterances that were read in a more expressive manner were also marked.

4. Acoustic Analysis

4.1. Experimental Setup

I^{2}R's in-house Mandarin speech recognizer (trained on adult speech) was used to obtain time boundaries of each phone through forced alignment of the phonetic transcriptions. Manual inspection was done to ensure the precision of the time boundaries. The estimates of pitch ($F0$) and formant frequencies ($F1$ and $F2$) were extracted at the mid point of the corner vowels /i/, /a/, and /u/ using Praat [21]. The Hamming window was set to 32 ms. Vowels that preceded and followed nasals were excluded from analysis to avoid the complication of nasal resonances as is customary in phonetic analysis.

4.2. Pitch and Formant Estimates

Table 2 lists the mean pitch and formant frequency estimates for the corner vowels /i/, /a/, and /u/ for each age group and for males and females respectively. The 7 year old age group is excluded as there is only one female speaker present in that age group. In general, the frequency estimates decrease as age increases, though there are some upward fluctuations. These trends and fluctuations correspond with the staggered nature of pubertal voice development in female and male adolescents [22, 23]. The upward fluctuations are more obvious in female speakers, possible because adolescent voice development can start as early as age 8 for females and spans a much longer duration than males [22]. To better visualize the high individual variation, we also show the mean formant frequency estimates on a per-speaker basis in Figure 2.

[2]Only 101 speakers are children.

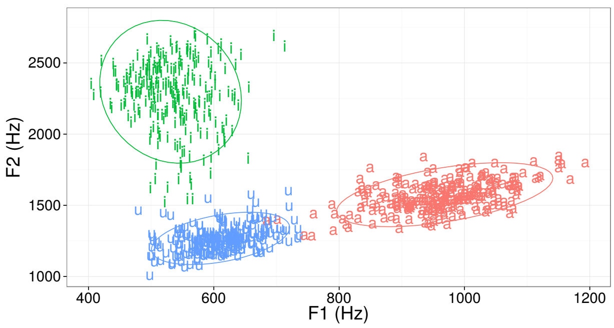

Figure 2: Average values of $F1$ and $F2$ for each speaker in SingaKids-Mandarin for the vowels /i/, /a/ and /u/. The ellipses represent the 95% confidence interval for each phone. Each text label represents a single speaker's mean $F1$ and $F2$.

5. Pronunciation Error Pattern Analysis

In this section we analyze the pronunciation patterns from the perspective of lexical tones and phonetic units. For the reader's convenience, we provide both Hanyu Pinyin with numerical tone markers in italics and IPA enclosed in square brackets or slashes.

5.1. Lexical Tone Errors

Tone is the use of pitch in speech. Tones only express paralinguistic information (e.g., emotion) in languages like English. In contrast, languages such as Mandarin Chinese uses *lexical tones* to encode semantics; i.e., a change in tone changes the meaning of a word. For example, ma1 [ma˥] and ma2 [ma˧˥] are phonetically the same but differ in tone, resulting in different meanings: mom vs. hemp. In Table 3, we introduce the lexical tones in Mandarin. Note that there is no equivalent in English for Tone 3, and Tone 5 is a neutral tone or *lack of tone*.

The overall tone error rate is 14.59%. Table 4 shows the accuracy breakdown in a confusion matrix. We see that Tone 5 has one of the lowest accuracies, achieving only 83.28% correct production given the reference is Tone 5. Below we elaborate more on the lack of neutral tone in Mandarin spoken in Singapore and examine the prevalence of tone sandhi.

5.1.1. *The Neutral Tone*

It is reported that neutral tone (Tone 5) is rarely used in daily speech in Singapore [24]. One potential reason is that the neutral tone is not known to exist in Southern Chinese languages like Min Nan, Teochew, and Cantonese, which are sometimes spoken in the households of these children. Our results show that the mapping of the neutral tone to other lexical tones (5 $\rightarrow$ {1,2,3,4}) occurs at a rate of 16.57%, while the opposite mapping ({1,2,3,4} $\rightarrow$ 5) is only 2.26%.

5.1.2. *Tone Sandhi*

Tone sandhi is a phonological change where the tones assigned to individual words change according to the pronunciation of adjacent words [25]. The most well-known example of tone sandhi in Mandarin is that when two consecutive Tone 3's occur, native speakers will instead produce a Tone 2 followed by a Tone 3: i.e. (3,3) $\rightarrow$ (2,3). We did not find literature discussing how tone sandhi is implemented in Mandarin spoken in Singapore. Tone sandhi was not found to be prevalent among Singaporean Chinese school children: only 22.25% of canonical (3,3) pairs

Table 1: Comparison of children's speech corpora. This table is sorted by the lower and upper bounds for age. An interactive version of this table is also available at `https://en.wikipedia.org/wiki/List_of_children%27s_speech_corpora`. Blanks indicate unavailable information. Word Trans.: Word transcription. Phon. Trans.: Phonetic transcription. Spon. Speech: Spontaneous speech. Prof. Rating: Proficiency rating. "P" indicates partial availability.

Corpus	Languages	# Spkrs	#Utt.	Dur.	Age Range	Date	Word Trans.	Phon. Trans.	Spon. Speech	Prof. Rating
Providence Corpus [5]	English	6		363h	1 - 3	2006	Y	Y	Y	N
Lyon Corpus [6]	French	4		185h	1 - 3	2004	Y	Y	Y	N
CASS_CHILD [7]	Mandarin	23		631h	1 - 4	2012	P	P	Y	N
Demuth Sesotho Corpus [8]	Sesotho	4	13250	98h	2 - 4	1992	N	Y	Y	N
CHIEDE [9]	Spanish	59	15,444	∼8h	3 - 6	2008	Y	Y	Y	N
PF-STAR Children's Speech Corpus [10, 11]	English	158		14.5h	4 - 14	2006	Y	N	N	N
TBALL [12]	English	256	30,000	40h	5 - 8	2005		P	Y	N
CMU Kids Corpus [13]	English	76	5180		6 - 11	1997	P	P	N	N
CU Children's Read and Prompted Speech Corpus [14]	English	663	∼66,300		K - G5	2001	Y	N	N	N
TIDIGITS [15]	English	326[2]			6 - 15	1993	Y	N	N	N
CSLU Kids' Speech Corpus [16]	English	1100	1,017		K - G10	2007	Y	N	P	N
Swedish NICE Corpus [17]	Swedish	75	5,580		8 - 15	2005	Y	N	Y	N
CU Story Corpus [14]	English	106	5,000	40h	G3 - G5	2003	P	P	P	N
FAU Aibo Emotion Corpus [18]	German	51	13,642	9h	10 - 13	2002	Y	N	Y	N
SingaKids-Mandarin	**Mandarin**	**255**	**79,843**	**125h**	**7 - 12**	**2016**	**Y**	**Y**	**N**	**P**

Table 2: Average fundamental and formant frequency estimates of vowels from male and female children in SingaKids-Mandarin by age (in years). All measurements are in Hz. The standard deviation is shown in parantheses.

		Male					Female				
Age (years)		8	9	10	11	12	8	9	10	11	12
/i/	$F0$	254 (58)	239 (56)	241 (54)	234 (52)	166 (40)	249 (61)	253 (68)	236 (53)	253 (61)	229 (53)
	$F1$	552 (168)	523 (155)	527 (167)	522 (168)	433 (180)	576 (168)	548 (172)	529 (175)	543 (162)	490 (157)
	$F2$	2262 (592)	2334 (556)	2402 (485)	2373 (490)	2322 (318)	2150 (646)	2152 (634)	2324 (504)	2333 (520)	2326 (449)
/a/	$F0$	244 (58)	230 (55)	234 (53)	224 (52)	159 (36)	241 (63)	243 (66)	229 (54)	244 (62)	218 (56)
	$F1$	1001 (135)	967 (126)	933 (117)	903 (113)	826 (140)	1045 (138)	1008 (156)	963 (128)	953 (127)	894 (130)
	$F2$	1596 (218)	1545 (224)	1559 (213)	1498 (233)	1409 (211)	1644 (277)	1596 (274)	1646 (262)	1632 (248)	1572 (252)
/u/	$F0$	253 (62)	238 (59)	243 (57)	233 (56)	163 (46)	250 (65)	250 (69)	239 (56)	254 (62)	224 (56)
	$F1$	629 (152)	596 (141)	594 (141)	594 (143)	534 (148)	651 (159)	628 (151)	601 (164)	614 (149)	576 (157)
	$F2$	1273 (344)	1238 (371)	1247 (335)	1230 (353)	1229 (486)	1318 (350)	1301 (374)	1323 (362)	1291 (331)	1286 (405)

Table 3: Lexical Tones in Mandarin.

Tone	Pitch Contour	English Equivalent
1	High-level	Singing
2	High-rising	Question-final intonation; e.g., What?!
3	Dipping	*No equivalent*
4	Falling	Curt commands; e.g., Stop!
5	*Undefined*	Unstressed syllable

were acoustically implemented as (2,3).

Table 4: *Confusion matrix of lexical tones (%).*

		Child's Production				
		Tone 1	Tone 2	Tone 3	Tone 4	Tone 5
Canonical	Tone 1	**86.55**	4.17	3.01	5.77	0.51
	Tone 2	4.27	**86.02**	2.54	6.17	1.00
	Tone 3	4.68	8.26	**81.90**	4.92	0.24
	Tone 4	4.89	3.97	3.49	**87.13**	0.52
	Tone 5	2.41	5.06	3.02	6.23	**83.28**

5.2. Phonetic Pronunciation Patterns

We found the overall phone error rate to be 5.48% by comparing the manually transcribed phone sequences with the canonical reference pronunciations of the reading prompts. As there is no literature characterizing Mandarin pronunciation patterns for Singaporean children, we compared our data against that of Singaporean adults [26]. In Table 5, we summarize how the pronunciation patterns of Singaporean children differ from those reported for Singaporean adults. We discuss the pronunciation patterns in more detail below.

5.2.1. Retroflex Sibilant Consonants: zh, ch, sh

Retroflex refers to the tip of the tongue turning back toward the hard palate as the place of articulation. *Sibilance* is a manner of articulation of fricative and affricate consonants, where air is directed towards the sharp edge of the teeth; the narrow constriction causes a hissing sound. Retroflex sibilant consonants are an important distinguishing feature of Mandarin Chinese, which are often the most challenging phonemes for second language

Table 5: Pronunciation patterns (represented in phonological rules) of Singaporean children. Phonological rules and their corresponding occurrence rates are in bold if they are *opposite* from the adult literature or *non-existent* in the adult literature [24, 26]. The last column indicates whether our findings in children's data has positive correlations with the adult literature (+), or negative correlations with the adult literature (-), or if the finding was never reported in the adult literature (N/A). IPA is provided in addition to Hanyu Pinyin.

Observed Patterns for Singaporean Children	SingaKids-Mandarin Phonological Rule	Frequency (%)	Patterns Reported for Singaporean Adults	Correlate with Adults?
Confusion of Retroflex Sibilants	zh /ʈʂ/ → z [ts]	10.74	Fronting of Retroflex Sibilant Consonants	+
	z /ts/ → zh [ʈʂ]	**6.36**		−
	ch /ʈʂʰ/ → c [tsʰ]	8.40		+
	c /tsʰ/ → ch [ʈʂʰ]	**8.12**		−
	sh /ʂ/ → s [s]	4.77		+
	s /s/ → sh [ʂ]	**8.18**		−
Collapse of Alveolar/Velar Nasal Finals	n /n/ → ng [ŋ]	2.71	Fronting of Velar Nasal Finals	+
	ng /ŋ/ → n [n]	2.26		+
Minimal Dentalization of Palatals	j /tɕ/ → z [ts]	1.12	Dentalization of Palatals	+
	z /ts/ → j [tɕ]	**1.08**		−
	q /tɕʰ/ → c [tsʰ]	0.68		+
	c /tsʰ/ → q [tɕʰ]	**1.51**		−
	x /ɕ/ → s [s]	0.57		+
	s /s/ → x [ɕ]	**2.52**		−
Deaspiration of Labial Stops	**p /pʰ/ → b [p]**	**6.03**	*None reported*	N/A
Monophthongization	**iu /i̯oʊ̯/ → u [u]**	**5.30**	iu[3]/i̯oʊ̯/ → iu [iu]	−
	io /i̯ɔ/ → o [o]	12.75		
Lack of Neutral Tone (Tone 5)	5 → {1, 2, 3, 4}	16.57	Lack of Neutral Tone (Tone 5)	+
	{1, 2, 3, 4} → 5	**2.26**		−

learners [27, 20]. Fronting of retroflex sibilant consonants (*zh* /ʈʂ/ → *z* [ts]; *ch* /ʈʂʰ/ → *c* [tsʰ]; *sh* /ʂ/ → *s* [s]) were reported for Singaporean adults (the generation of parents or grandparents of the subjects in SingaKids-Mandarin). However, we see that this is only true for the unaspirated retroflex affricate *zh* [ʈʂ] from Table 5. For the aspirated affricates, both fronting (*ch* /ʈʂʰ/ → *c* [tsʰ]) and backing (*c* /tsʰ/ → *ch* [ʈʂʰ]) are equally likely. For the fricatives *s* [s] and *sh* [ʂ], backing is actually more common.

5.2.2. *Alveolar and Velar Nasals: n, ng*

The confusions between velar and alveolar nasals is less frequent than that of sibilant consonants, and both directions of fronting and backing are equally likely. A similar trend is noted in Singaporean adults [26]. The alveolar nasal *n* [n] and velar nasal *ng* [ŋ] confusion is common even for native monolingual Mandarin speakers in Southern China and Taiwan [28].

5.2.3. *Palatal Sibilant Consonants: j, q, x*

Dentalization of palatal sibilant consonants (*j* /tɕ/ → *z* [ts], *q* /tɕʰ/ → *c* [tsʰ], *x* /ɕ/ → *s* [s]) has been mentioned in previous literature [26] and is an anecdotal observation often commented on forums and blog posts discussing Singaporean Mandarin. In our data, this trend was not observed to occur often (less than 1%). The opposite trend (*z* /ts/ → *j* [tɕ], *c* /tsʰ/ → *q* [tɕʰ], *s* /s/ → *x* [ɕ]) was at least equally frequent, if not more.

5.2.4. *Labial Stop Consonants: p, b*

In this work, we observed that Singaporean children only have deaspiration trends for the labial stop, a trend that has not been observed in adults. For Mandarin learners whose L1 is European, we have found that deaspiration/aspiration of sibilants and stops to be a major source of error, especially for native Romance language speakers (e.g., French, Italian, Spanish) [20].

5.2.5. *Monophthongization*

We see that P(*io* /i̯ɔ/ → *o* [o]) = 12.75% for Chinese Singaporean children. We find that Singaporean children tend to undergo monophthongization when a diphthong is encountered. This may be related to the anecdotally reported difficulty of acquiring diphthongs in Singaporean children. This appears contrary to literature where diphthongs are *easier* to acquire in other languages such as Cantonese [29].

6. Discussion

Reported speech data on children has been scarce, and this scarcity is particularly true for Mandarin. To the best of our knowledge, we have not found other reported work documenting the acoustic variation in Mandarin spoken by children. Our acoustic analysis of *SingaKids-Mandarin* corresponds with voice development literature, and also indicates that individual variations among each child is high, implying there is much research potential for developing robust spoken language technology for children. In terms of Mandarin pronunciation patterns of Singaporeans, we observed that while the lack of neutral tone usage is the same for adults and children, the phonetic pronunciation patterns in these two age groups differ: while adults tend to show fronting in retroflex, nasal, and palatal consonants, children show both fronting and backing in these consonants.

For future work, we plan to develop a language learning platform tailored for Singaporean children (speaking Mandarin and English), with an emphasis on prosody.

We are currently in the process of making *SingaKids-Mandarin* publicly available to the research community. For readers interested in the corpus, please contact `nfychen@i2r.a-star.edu.sg`.

[3]There are some inconsistencies in Pinyin causing the IPA representations [i̯oʊ̯] and [iu] to be the same in Pinyin.

7. References

[1] Peter Siemund, Monika Schulz, and Martin Schweinberger, "How multilingual are singaporeans really? a sociological and sociolinguistic analysis of the linguistic ecology of singapore," in *International Symposium on Bilingualism*, 2013.

[2] "Mother Tongue Language Policy in Singapore: http://www.moe.gov.sg/education/admissions/returning-singaporeans/mother-tongueSingapore'," .

[3] Alexandros Potamianos, Shrikanth Narayanan, and Sungbok Lee, "Automatic speech recognition for children.," *Eurospeech*, vol. 97, pp. 2371–2374, 1997.

[4] Mohamed Benzeghiba, Renato De Mori, Olivier Deroo, Stephane Dupont, Teodora Erbes, Denis Jouvet, Luciano Fissore, Pietro Laface, Alfred Mertins, Christophe Ris, et al., "Automatic speech recognition and speech variability: A review," *Speech Communication*, vol. 49, no. 10, pp. 763–786, 2007.

[5] Katherine Demuth, Jennifer Culbertson, and Jennifer Alter, "Word-minimality, Epenthesis and Coda Licensing in the Early Acquisition of English," *Language and Speech*, vol. 49, no. 2, pp. 137–173, 2006.

[6] Katherine Demuth and Annie Tremblay, "Prosodically-conditioned Variability in Children's Production of French Determiners," *Journal of child language*, vol. 35, no. 01, pp. 99–127, 2008.

[7] Jun Gao, Aijun Li, and Ziyu Xiong, "Mandarin Multimedia Child Speech Corpus: CASS_CHILD," *International Conference on Speech Database and Assessments*, 2012.

[8] Katherine Demuth, "Acquisition of Sesotho," *The Cross-Linguistic Study of Language Acquisition*, 1992.

[9] Marta Garrote, *CHIEDE: A Spontaneous Child Language Corpus of Spanish*, Ph.D. thesis, Universidad Autonoma de Madrid, 2008.

[10] Martin Rusell, "The PF-STAR British English Children's Speech Corpus," *The Speech Ark Limited*, 2006.

[11] Anton Batliner and Mats Blomberg and Shona D'Arcy and Daniel Elenius and Diego Giuliani and Matteo Gerosa and Christian Hacker and Martin Russell and Stefan Steidl and Michael Wong, "The pf-star children's speech corpus," *Proceedings of Interspeech*, 2005.

[12] Abe Kazemzadeh, Hong You, Markus Iseli, Barbara Jones, Xiaodong Cui, Margaret Heritage, Patti Price, Elaine Anderson, Shrikanth Narayanan, and Abeer Alwan, "TBALL Data Collection: The Making of a Young Children's Speech Corpus," *Proceedings of Interspeech*, 2005.

[13] Maxine Eskenazi, Jack Mostow, and David Graff, "The CMU Kids Corpus," *Linguistic Data Consortium*, 1997.

[14] Andreas Hagen, Bryan Pellom, and Ronald Cole, "Children's Speech Recognition with Application to Interactive Books and Tutors," *IEEE Workshop on Automatic Speech Recognition and Understanding*, 2003.

[15] R. Gary Leonard and George Doddington, "TIDIGITS," *Linguistic Data Consortium*, 1993.

[16] Khaldoun Shobaki, John-Paul Hosom, and Ronald Cole, "CSLU: Kids' Speech Version 1.1," *Linguistic Data Consortium*, 2007.

[17] Linda Bell, Johan Boyce, Joakim Gustafson, Mattias Heldner, Anders Lindstrom, and Mats Wiren, "The Swedish NICE Corpus - Spoken Dialogues between Children and Embodied Characters in a Computer Game Scenario," *Proceedings of Eurospeech*, 2005.

[18] Stefan Steidl, *Automatic Classification of Emotion-Related User States in Spontaneous Children's SPeech*, Ph.D. thesis, Friedrich-Alexander University Erlangen-Nrnberg, 2009.

[19] "Chinese Text Computing: http://lingua.mtsu.edu/chinese-computing/," last accessed, March 17, 2015.

[20] Nancy F. Chen, Rong Tong, Darren Wee, Peixuan Lee, Bin Ma, and Haizhou Li, "iCALL Corpus: Mandarin Chinese Spoken by Non-Native Speakers of European Descent," in *INTERSPEECH*, 2015.

[21] Paul Boersma, "Praat, a system for doing phonetics by computer.," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[22] Lynne Gackle, "The Adolescent Female Voice Characteristics of Change and Stages of Development," *Choral Journal*, vol. 31, no. 8, pp. 17, 1991.

[23] Meredydd Harries, Sarah Hawkins, Jeremy Hacking, and Ieuan Hughes, "Changes in the male voice at puberty: vocal fold length and its relationship to the fundamental frequency of the voice," *The Journal of Laryngology & Otology*, vol. 112, no. 05, pp. 451–454, 1998.

[24] Guowen Shang and Shouhui Zhao, "Singapore mandarin: Its positioning, internal structure and corpus planning," in *SEALS 22*, 2012.

[25] Moira Yip, *Tone*, Cambridge University Press, 2002.

[26] Chung-Yu Chen, *Segmental Salient Features in the Mandarin Spoken in Singapore*, Chinese Language & Research Centre, National University of Singapore, 1982.

[27] Nancy F. Chen, Vivaek Shivakumar, Mahesh Harikumar, Bin Ma, and Haizhou Li, "Large-scale characterization of Mandarin pronunciation errors made by native speakers of European languages," in *Interspeech*, 2013, pp. 803–806.

[28] Eric Zee, "Sound Change in Syllable Final Nasal Consonants in Chinese," *Journal of Chinese Linguistics*, vol. 13, pp. 291–330, 1985.

[29] Stephanie F. Stokes and I. Man Wong, "Vowel and diphthong development in Cantonese-speaking children," *Clinical Linguistics & Phonetics*, vol. 16, pp. 597–617, 2002.