NITK Kids' Speech Corpus

Conference Paper · September 2019

DOI: 10.21437/Interspeech.2019-2061

6 authors, including:

Sujata Supanekar
National Institute of Technology Karnataka

Pradyoth Hegde
National Institute of Technology Karnataka

Shashidhar Koolagudi
National Institute of Technology Karnataka

INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

NITK Kids' Speech Corpus

Pravin Bhaskar Ramteke[1], Sujata Supanekar[1], Pradyoth Hegde[1], Hanna Nelson[2], Venkataraja Aithal[2] and Shashidhar G. Koolagudi[1]

[1]National Institute of Technology Karnataka, Surathkal, India
[2]School of Allied Health Sciences, Manipal, Karnataka

ramteke0001@gmail.com, sujata.supanekar@gmail.com, pradyothhegde@gmail.com, hannaenelson01@gmail.com, vrajaithal@manipal.edu, koolagudi@nitk.edu.in

Abstract

This paper introduces speech database for analyzing children's speech. The proposed database of children is recorded in Kannada language (one of the South Indian languages) from children between age $2\frac{1}{2}$ to $6\frac{1}{2}$ years. The database is named as National Institute of Technology Karnataka Kids' Speech Corpus (NITK Kids' Speech Corpus). The relevant design considerations for the database collection are discussed in detail. It is divided into four age groups with an interval of 1 year between each age group. The speech corpus includes nearly 10 hours of speech recordings from 160 children. For each age range, the data is recorded from 40 children (20 male and 20 female). Further, the effect of developmental changes on the speech from $2\frac{1}{2}$ to $6\frac{1}{2}$ years are analyzed using pitch and formant analysis. Some of the potential applications, of the NITK Kids' Speech Corpus, such as, systematic study on the language learning ability of children, phonological process analysis and children speech recognition are discussed.

Index Terms: NITK Kids' Speech Corpus, Children speech recognition, Formant, Pitch, Spectral feature, Prosodic feature.

1. Introduction

Most widely used speech interface based applications, such as mobile phones, dictation engine, navigation systems, are mainly designed on adult speech data. Many databases are available for adult speech, whereas, efforts that target children's speech are less common.[1]. In recent years, the databases of childrens' speech are gaining more importance due to its application in diverse areas. Applications such as foreign language learning, computer games and reading tutors have become very common for children, when compared with the adults [2]. Despite of huge research opportunities in the field of children's speech, it is surprising that relatively little research has been focused on the development of speech technologies for children. At present, the available databases of children's speech are significantly less than that of adults' speech. Most applied and relevant databases of children's speech are listed in Table 1 [3]. Many of these databases are recorded in English, from children, aged between 6 to 15 years. It is aimed at automatic speech recognition of children's speech, where it is observed that the speech recognition error rate is very high for children than that for adults [4]. The major contributor to the recognition errors is the intra-speaker and inter-speaker variability [5] and bandwidth-restriction [6]. Performance is improved using vocal tract length normalization (a frequency warping technique), which compensates the effects of frequency shift in children due to shorter vocal tract [7].

Researches have focused more on American English whereas very little work has been done in European and Asian languages. Also, the databases recorded have targeted the age range between 6 to 15 years. This database can only be used for automatic recognition of children's speech, effects of non-native language, identification of errors in reading, etc. The children's speech databases for lower age groups between $2\frac{1}{2}$ to $6\frac{1}{2}$ are rarely available. The children's speech, in this age group, has more research applications compared to the higher age group. Due to underdeveloped vocal tract and lack of neuro-motor control, children face difficulty in pronunciation and substitute the difficult class of sounds with the simple class of sounds. These mispronunciation patterns are known as phonological processes [25]. Speech-Language Pathologists (SLPs) analyze this pattern of disappearance in different age groups and evaluate the language learning ability of children. The order and types of appearance of the phonological processes, vary between languages of different natures. If these phonological processes persist after the age of 8 years, it may result in phonological disorder [26]. Research in this line would yield a better understanding of the neuro-motor limitations on articulatory movement in this age range.

Children speech data in the range of $2\frac{1}{2}$ to $6\frac{1}{2}$ years is rarely available, due to difficulties of recording compared to the recording of adults and school children [7]. Children in this age group are not able to read and have a short attention span which makes the recording difficult. In this paper, the children dataset known as NITK Kids' Speech corpus in Kannada language is proposed along with the basic properties of the data. The protocol and methodology used in the process of development is discussed in detail. As children in the proposed age group are not able to properly read the text, different pictures are used to extract specific phonemes and words. Children are asked to describe the picture and required words are chosen from the description given by the children.

Rest of the paper is organized as follows. Section 2 discusses the acoustic-phonetic structure of Kannada language. Section 3 gives details of design of the corpora, along with the equipments used for recording and procedure adopted to record the speech. Spectral and prosodic analysis of children speech is given in Section 4. The details of the applications, where the database can be used is given in Section 5. Section 6 presents the conclusion and describes the perspectives of further research projects.

2. Acoustic-phonetic Features of Kannada

Kannada is one of the 22 scheduled languages of India according to the constitution of India. It belongs to the South Dravidian language family), mainly spoken in the state of Karnataka. Different phones used in Kannada along with their IPA symbol and ASCII representation are given in Table 2. The first sec-

http://dx.doi.org/10.21437/Interspeech.2019-2061

Table 1: *List of available children speech datasets*

Speech Corpus	Language	Recording Type	No. of Kids	Age (year)
CID children's speech corpus [8]	American English	read	436	5-17
CMU Kid's speech corpus [9]	American English	read	76	6-11
CU Kid's Prompted and Read Speech corpus [10]	American English	read	663	4-11
CU Kid's Read and Summarized Story corpus [11]	American English	spontaneous	326	6-11
OGI Kid's speech corpus [12]	English	read	1100	5-15
ChIMP corpus [13]	American English	spontaneous	160	8-14
Tball corpus [7]	English (native Spanish)	–	256	5-8
TIDIGITS corpus [14]	English	–	101	6-15
PF-STAR corpus [15]	English, German, Swedish, Italian	read, spontaneous, emotion	491	4-15
ChildIt corpus [16]	Italian	–	171	7-13
Tgr-child corpus [16]	Italian	–	30	8-12
SponIt corpus [16]	Italian	–	21	8-12
NICE corpus [17]	Swedish	–	75	8-15
PIXIE corpus [18]	Swedish	–	2885	-NA-
Rafael.0 telephone corpus [2]	Danish	–	306	8-18
CHOREC corpus [19]	Dutch	read	400	6-12
SPECO corpus [20]	Hungarian	read	72	5-10
Takemaru-kun corpus [21]	Japanese	–	17392	-NA-
VoiceClass Database [22]	German	free speech	170	7-14
Deutsche Telekom telephone speech corpus [22]	German	free speech, prompt	106	7-14
Lesetest corpus [23]	German	read	62	10-12
SpeeCon corpus [24]	20 languages	–	50/language	8-15

tion of Table 2, lists the vowels ($swaras$); the second section lists the $yogavaahakas$ and third section lists the consonants ($vyanjanas$) namely; plosives, affricates and nasals, whose production involves complete closure of oral tract. Fourth section lists the semivowels and fricatives [27]. The acoustic-phonetic profile of Kannada language significantly differs from the European languages. Unlike English, aspiration is one of such phone level feature observed in Kannada. There are total eight aspirated plosives and two aspirated fricatives, along with their unaspirated counterparts. Retroflex sounds hold a special place in Kannada phonemic system. These characteristics are taken into account while choosing the representative speech samples for speech database recording from children.

3. The Design of Corpora

The proposed database is recorded from 160 native Kannada language (one of the important South Indian languages) children in the age group of $2\frac{1}{2}$ to $6\frac{1}{2}$ years. Children are divided in to 4 age groups, $2\frac{1}{2}$-$3\frac{1}{2}$, $3\frac{1}{2}$-$4\frac{1}{2}$, $4\frac{1}{2}$-$5\frac{1}{2}$ and $5\frac{1}{2}$-$6\frac{1}{2}$ years respectively. In each age group, data is recorded from 20 girls and 20 boys. Representative speech samples are those that reflect a child's typical use of speech sounds during everyday activities. For each of these speech sound units, three words which have that sound unit in initial, medial and final positions are selected. This gives a recommended sample size of 112 words (representative speech samples) [28]. Comparison of an isolated word pronunciation, pronounced in a the specific context, concludes that the context-based speech samples better describe the articulatory behavior in children [29, 30]. To get the contextual information, pictures representing the representative Kannada words are selected. Figure 1 shows some of the pictures and corresponding representative words used for speech recording.

Recordings are done from different regions of Karnataka, in a quiet room using a single microphone; without any ob-

Table 2: *IPA symbol and ASCII representation of the Kannada phonemes*

Vowels

a	$a:$	i	$i:$	u	$u:$	$\underset{\cdot}{r}$	e	$e:$	ai	o	$o:$	au
a	A	i	I	u	U	ru	e	E	ai	o	O	au

Yogavaahaka

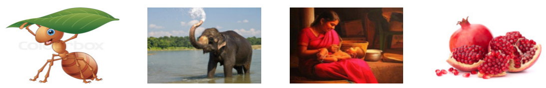
$a\dot{m}$	$a\dot{h}$
aṁ	aḣ

Consonants

k	k^h	g	g^h	ŋ
k	kh	g	gh	ng
tʃ	tʃʰ	ʤ	ʤʰ	ɲ
c	ch	j	jh	jn
ʈ	ʈʰ	ɖ	ɖʰ	ɳ
T	Th	D	Dh	N
t	t^h	d	d^h	n
t	th	d	dh	n
p	p^h	b	b^h	m
p	ph	b	bh	m

Semivowels and Fricatives

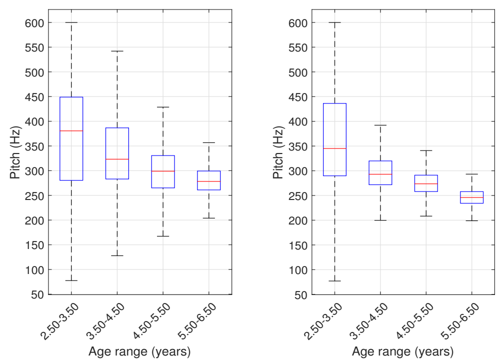
j	r	l	ω	ʃ	ʂ	s	h	$\underset{\cdot}{l}$
y	r	l	w	sˆ	S	s	h	l



(a) (b) (c) (d)

Figure 1: *Some of the representative speech samples for Kannada phonemes: (a)* ”$iruve$” *(ant) and* ”ele” *(leaf) (b)* ”Ane” *(elephant) and* ”$snAna$” *(bath) (c)* ”$amma$” *(mother) (d)* ”$dALimbe$” *(pomegranate)*

stacles in the recording path. Blue Yeti USB Microphone has been used for the dataset recording. It has polar patterns of cardioid, bidirectional, omnidirectional and stereo, with a frequency response of 20 Hz-20 kHz. Polarity is set to cardioid for the recording. Cardioid represents the ’heart-shaped’ pick-up pattern, where it emphasizes sounds from the direction mic is pointed. It is good at rejecting sounds from other directions. Audio data is recorded at the sampling rate of 48 kHz, with a bit rate of 16-bits per sample. Microphone is connected to a laptop computer to record speech using WaveSurfer (an open source tool for sound recording, visualization, annotation/transcription and manipulation). Children are made to sit in a comfortable position in front of a computer and microphone. They are asked to describe the picture representing words displayed on a PowerPoint slide on the computer screen. Children get bored easily, hence sufficient break is given between the recording to maintain proper response during the session. Each recording is analyzed by qualified phonetician and SLPs, which are rerecorded, if suggested by them.



(a) (b)

Figure 2: *Analysis of variation in pitch over age range 2.50 to 6.50 years: (a) Pitch variation in female children (b) Pitch variation in male children*

4. Spectral and Prosodic Analysis of Children’s Speech

Characteristics of the children’s speech can be seen at the source level (characteristics of excitation signal and shape of the glottal pulse), system level (shape of the vocal tract and nature of movements of different articulators) and prosodic level [31].

4.1. Prosodic Analysis

Among the features from different levels for analysis of speech in different age groups, prosodic features have attributed to a major role by the existing literature [32]. Therefore, this paper shows the importance of prosodic parameters using pitch extracted from the speech database. Pitch is the rate of vocal fold vibration, where thinner the vocal folds, higher the pitch and vice versa. The pitch values of each utterance are obtained from the autocorrelation of the Hilbert envelope of the linear prediction residual [33]. Speech of five male and five female children in each age group is considered for illustrating the statistics of prosodic parameters. Male and female children from the age groups of 3.00-3.50, 4.00-4.50, 5.00-5.50, 6.00-6.50 years are considered to show the change in the pitch with an increase in age. Pitch values are determined frame-wise at word level. Children have thinner vocal folds, hence have a high pitch when compared to adults. Pitch drops down according to the increase in age, due to increase in the thickness of the vocal fold in both male and female child. The variation in the pitch of female and male children in the proposed age range of $2\frac{1}{2}$ to $6\frac{1}{2}$ years is shown in Fig. 2 (a) and Fig. 2 (b) respectively. The systematic decrease in the pitch over the specified range indicates the development in the size of the vocal folds. Here, it is interesting to note that, there is a significant difference in the pitch of male and female children in each age group and male pitch is low compared to the female pitch. Median value of pitch for age 3 to $3\frac{1}{2}$ years in female is around 380Hz whereas for male it is 340Hz. For age range 4 years to $4\frac{1}{2}$ years in female is around 340Hz whereas for male it is 300Hz. Female pitch between age 5 years to $5\frac{1}{2}$ years is observed to be 300Hz whereas male pitch is observed to be around 280Hz. The last age group shows the difference of 30Hz between male and female median pitch. It shows that, within the same age range, the male children vocal folds are thick when compared to the female vocal folds.

4.2. Spectral Analysis

Spectral and temporal properties of children’s speech are greatly affected by the growth and other developmental changes [34]. These variations are characterized by anatomical and morphological development in the vocal-tract geometry and control over the articulators. It is also reported that, the children speech has higher variability in speaking rate, vocal effort and degree of spontaneity [34]. The detailed analysis of variation in age-dependent behavior of the measurements of spectral and temporal parameters has been based on American English vowels [8]. The analysis has shown an orderly decrease in the values of the acoustic correlates, such as, formants, pitch and duration, with increase in age. It is observed that the formant values of the vowels decrease with increase in age, representing the lengthening of the vocal tract with the growth of the children. Also, it is noticed that with increase in age, there is a decrease in the dynamic range of the formant values [8]. In this paper, formant frequencies are estimated using filter coefficients obtained through LPC analysis of speech [35]. The number of kids considered in each age group is same as in pitch analysis (refer Section 4.1). For the analysis, four formants (F1, F2, F3 & F4) are extracted from the vowels namely $/a/$, $/A/$, $/i/$, $/u/$, $/e/$, $/o/$. Figure 3 (a)-(d) shows the scatter plot of the first formant (F1) vs second formant (F2) in each age group. It is observed that there is no significant change in the formant frequencies of different vowels in consecutive age groups. When the formant frequencies of the vowels in the age range 3 years to $3\frac{1}{2}$ years and 6 years to $6\frac{1}{2}$ years are observed, there is a systematic decrease in the formant values of vowel sounds. For example, in the age range 3 years to $3\frac{1}{2}$ years, for vowel $/a/$ the mean value of F1 is 975.00Hz, F2 is 2356.10Hz, F3 is 4144.00Hz

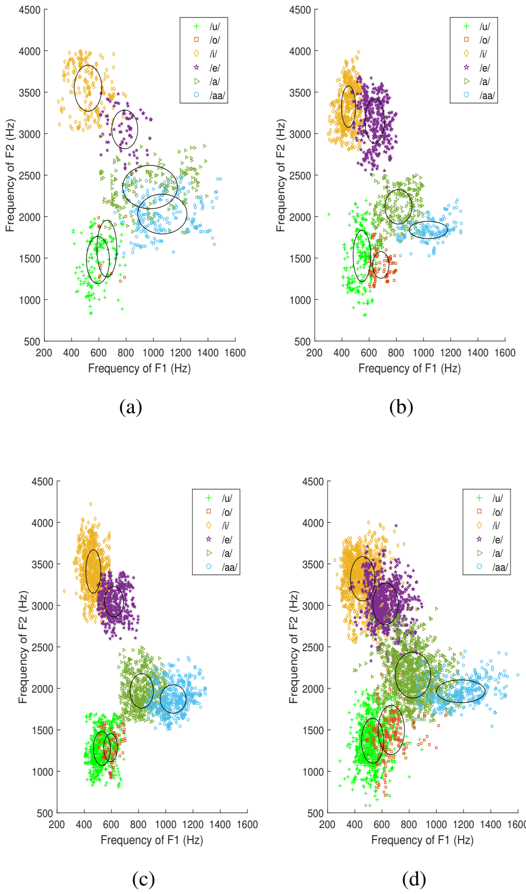


Figure 3: *Analysis of variation in formants over age range 2.50 to 6.50 years for vowels /a/, /A/, /i/, /u/, /e/, /o/ using scatter plot of F1 vs F2: (a) Age between 3.00 to 3.50 years (b) Age between 4.00 to 4.50 years (c) Age between 5.00 to 5.50 years (d) Age between 6.00 to 6.50 years.*

and F4 is 5608.60Hz. Whereas, in the age range 6 years to $6\frac{1}{2}$ years, the mean values of F1 is 826.80Hz, F2 is 2160.50Hz, F3 is 4068.20Hz and F4 is 5384.20Hz. Similarly, from the analysis of the scatter plots of 3 years to $3\frac{1}{2}$ years and 6 years to $6\frac{1}{2}$ years it is observed that, for all the vowels there is a systematic decrease in the formant frequencies of the vowel with the growth and development of vocal-tract geometry and control over the articulators.

5. Applications of the Database

This children speech database known as NITK Kids' Speech corpus may be used in realizing the following tasks.

5.1. Children Speech Recognition

Spectral and temporal variability is induced in children speech due to developmental changes in vocal tract [36]. Such variabilities pose challenges for robust automatic recognition of children speech. From the analysis of age-based characteristics of children's speech, such as frequency scaling of spectral envelope, vocal tract length normalization (VTLN), pitch normalization etc., are commonly performed for automatic children speech recognition [36]. Many of these implementations have used speech databases of children five years and above. Speech recognition in the age less than 6 years is more challenging due to higher variability in speech production parameters. This database can be used to build a robust children speech recognition for lower age groups.

5.2. Identification of speaker, age-group and gender

Speech signals carry paralinguistic information such as speaker's identity, age, gender, emotion, etc [32]. In adult's speech, this information is proved to be very evident due to proper mental and physiological development. Children in lower age groups have higher variability in speech production parameters and lack of command over language [32]. This makes the task of identification of speaker, age-group and gender in age groups $2\frac{1}{2}$ to $6\frac{1}{2}$ more complicated when compared to the adults. This database can be used to explore the paralinguistic information for the above mentioned tasks.

5.3. Analysis of Language Learning Ability and Mispronunciation Detection

Mispronunciations are very common in children due to underdeveloped vocal tract and lack of neuro-motor control [32]. They face difficulty in pronouncing the speech sounds which are complex in nature, hence substitute those sounds with the simpler ones (refer Section 1) [25]. This is one of the interesting areas in children speech processing, where SLPs analyze mispronunciation patterns in different age groups and evaluate the language learning ability of children. Persistence of these mispronunciation patterns beyond 8 years indicate higher chances of phonological disorder [26]. This analysis is generally performed manually. Hence, the proposed database can be explored for automatic identification of phonological processes/pronunciation errors and measure language learning ability of children.

6. Summary and Conclusion

In this paper, we have proposed a NITK Kids' speech corpus recorded from 160 children of age range $2\frac{1}{2}$ to $6\frac{1}{2}$ years in Kannada language. The corpus includes spontaneous and imitated speech recordings from $2\frac{1}{2}$ to $3\frac{1}{2}$ year old children and spontaneous speech recordings of children in the age range of $3\frac{1}{2}$ to $6\frac{1}{2}$ years. In total, the corpus includes nearly 10 hours of recordings of speech from 160 children. The proposed database has wide variety of characteristics in terms of recordings of children from lower ages, different age groups, speakers and representation samples (images). NITK Kids' speech corpus can be further exploited for analysis and characterization of childrens' speech using features extracted from the vocal tract, prosodic and excitation source. A systematic study on the language learning ability of the children, phonological process analysis, speech language pathology, children speech recognition, practice producing a variety of speech sounds can be performed.

7. Acknowledgements

The authors would like to thank the Cognitive Science Research Initiative (CSRI), Department of Science & Technology, Government of India, Grant no. SR/CSRI/49/2015, for its financial support to this work.

8. References

[1] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren, "DARPA TIMIT acoustic phonetic continuous speech corpus (vol. ldc93s1)," *Philadelphia: Linguistic Data Consortium*, 1993.

[2] J. G. Wilpon and C. N. Jacobsen, "A study of speech recognition for children and the elderly," in *icassp*. IEEE, 1996, pp. 349–352.

[3] F. Claus, H. G. Rosales, R. Petrick, H. U. Hain, and R. Hoffmann, "A survey about databases of children's speech." in *INTERSPEECH*, 2013, pp. 2410–2414.

[4] Q. Li and M. J. Russell, "An analysis of the causes of increased error rates in children's speech recognition," in *Seventh International Conference on Spoken Language Processing*, 2002, pp. 2337–2340.

[5] M. Karina, C. Felix, H. Horst-Udo, and P. Rico, "Herausforderungen an sprachinterfaces für kinder," in *Electronic Speech Signal Processing*, 2010, pp. 180–187.

[6] S. D'Arcy and M. Russell, "A comparison of human and computer recognition accuracy for children's speech," in *Ninth European Conference on Speech Communication and Technology*, 2005, pp. 2197–2200.

[7] A. Kazemzadeh, H. You, M. Iseli, B. Jones, X. Cui, M. Heritage, P. Price, E. Anderson, S. Narayanan, and A. Alwan, "Tball data collection: the making of a young children's speech corpus," in *Ninth European Conference on Speech Communication and Technology*, 2005, pp. 1581–1584.

[8] S. Lee, A. Potamianos, and S. Narayanan, "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," *The Journal of the Acoustical Society of America*, vol. 105, no. 3, pp. 1455–1468, 1999.

[9] M. S. Eskenazi, "Kids: A database of children's speech," *The Journal of the Acoustical Society of America*, vol. 100, no. 4, pp. 2759–2759, 1996.

[10] R. Cole, P. Hosom, and B. Pellom, "University of Colorado prompted and read children's speech corpus," Technical Report TR-CSLR-2006-02, University of Colorado, Tech. Rep., 2006.

[11] R. Cole and B. Pellom, "University of colorado read and summarized stories corpus," Technical Report TR-CSLR-2006-03, Center for Spoken Language Research, Tech. Rep., 2006.

[12] K. Shobaki, J. P. Hosom, and R. A. Cole, "The OGI kids' speech corpus and recognizers," in *Sixth International Conference on Spoken Language Processing*, vol. 4, 2000, pp. 258–261.

[13] A. Potamianos and S. Narayanan, "Spoken dialog systems for children," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 1. IEEE, 1998, pp. 197–200.

[14] R. Leonard, "A database for speaker-independent digit recognition," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'84.*, vol. 9. IEEE, 1984, pp. 328–331.

[15] A. Batliner, M. Blomberg, S. D'Arcy, D. Elenius, D. Giuliani, M. Gerosa, C. Hacker, M. Russell, S. Steidl, and M. Wong, "The PF_STAR children's speech corpus," in *Ninth European Conference on Speech Communication and Technology*, 2005, pp. 2761–2764.

[16] M. Gerosa, "Acoustic modeling for automatic recognition of children's speech," Ph.D. dissertation, Ph. D. thesis, University of Trento, 2006.

[17] L. Bell, J. Boye, J. Gustafson, M. Heldner, A. Lindström, and M. Wirén, "The Swedish NICE Corpus–spoken dialogues between children and embodied characters in a computer game scenario," in *9th European Conference on Speech Communication and Technology*. ISCA, 2005, pp. 2765–2768.

[18] L. Bell and J. Gustafson, "Child and adult speaker adaptation during error resolution in a publicly available spoken dialogue system," in *Eighth European Conference on Speech Communication and Technology*, 2003, pp. 613–616.

[19] L. Cleuren, J. Duchateau, P. Ghesquiere *et al.*, "Children's oral reading corpus (CHOREC): description and assessment of annotator agreement," in *Language Resources and Evaluation*. European Language Resources Association (ELRA), 2008, pp. 998–1005.

[20] F. Csatári, Z. Bakcsi, and K. Vicsi, "A Hungarian child database for speech processing applications," in *Sixth European Conference on Speech Communication and Technology*, 1999, pp. 2231–2234.

[21] C. Tobias, S. Izumi, T. Tomoki, S. Hiroshi, and S. Kiyohiro, "Development of preschool children subsystem for ASR and QA in a real-environment speech-oriented guidance task," in *Eighth Annual Conference of the International Speech Communication Association*, 2007, pp. 1469–1472.

[22] F. Burkhardt, M. Eckert, W. Johannsen, and J. Stegmann, "A database of age and gender annotated telephone speech." in *LREC*, 2010.

[23] H. Grissemann and M. Linder, "Zürcher Lesetest," *Bern: Huber Verlag*, 2000.

[24] D. Iskra, B. Grosskopf, K. Marasek, H. Heuvel, F. Diehl, and A. Kiessling, "Speecon-speech databases for consumer devices: Database specification and validation," 2002.

[25] D. Ingram, "Phonological rules in young children," *Journal of child language*, vol. 1, no. 1, pp. 49–64, 1974.

[26] R. D. Kent and H. K. Vorperian, "Speech impairment in Down syndrome: A review," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 1, pp. 178–210, 2013.

[27] B. Aarti and S. K. Kopparapu, "Spoken Indian language identification: a review of features and databases," *Sādhanā*, vol. 43, no. 4, p. 53, 2018.

[28] L. D. Shriberg and J. Kwiatkowski, "Phonological disorders I: A diagnostic classification system," *Journal of Speech and Hearing Disorders*, vol. 47, no. 3, pp. 226–241, 1982.

[29] M. A. Faircloth and S. R. Faircloth, "An analysis of the articulatory behavior of a speech-defective child in connected speech and in isolated-word responses," *Journal of Speech and Hearing Disorders*, vol. 35, no. 1, pp. 51–61, 1970.

[30] N. Andrews and M. E. Fey, "Analysis of the speech of phonologically impaired children in two sampling conditions," *Language, Speech, and Hearing Services in Schools*, vol. 17, no. 3, pp. 187–198, 1986.

[31] L. C. Yang, "The expression and recognition of emotions through prosody," in *Sixth International Conference on Spoken Language Processing*, 2000.

[32] S. Safavi, M. Russell, and P. Jančovič, "Automatic speaker, age-group and gender identification from children's speech," *Computer Speech & Language*, vol. 50, pp. 141–156, 2018.

[33] S. M. Prasanna and B. Yegnanarayana, "Extraction of pitch in adverse conditions," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1. IEEE, 2004, pp. 109–112.

[34] A. Potamianos and S. Narayanan, "Robust recognition of children's speech," *IEEE Transactions on speech and audio processing*, vol. 11, no. 6, pp. 603–616, 2003.

[35] R. C. Snell and F. Milinazzo, "Formant location from lpc analysis data," *IEEE transactions on Speech and Audio Processing*, vol. 1, no. 2, pp. 129–134, 1993.

[36] M. Gerosa, D. Giuliani, and F. Brugnara, "Acoustic variability and automatic recognition of children's speech," *Speech Communication*, vol. 49, no. 10-11, pp. 847–860, 2007.