Computer Fraud and Security
ISSN (online): 1873-7056

Research of Infant Voice Pitch Based on Speech Analysis Technology

Lei Guo[1,*], Yingxin Gao[2], Zhiwen Zhang[2]
[1]Key Laboratory of Linguistic and Cultural Computing Ministry of Education, Northwest Minzu University, Lanzhou, Gansu, China
[2]Key Laboratory of Language Science and Cultural Fusion Computing, Northwest Minzu University, Lanzhou, Gansu, China
**Corresponding Author.*

Abstract:
This paper takes the pitch of infants' speech as the research object, chooses 4 infants as the research target, studies the pronunciation of infants before the age of two. The infant speech database was established, and the infant speech analysis platform was developed, including automatic segment segmentation of speech samples and analysis of related acoustic parameters. After two years of tracking and recording, the characteristics and rules of phonetic pitch pronunciation of infants in pre-language stage and language stage were studied by the method of speech experiment. This study expands the vision of infant phonetics research and summarizes the phased characteristics and rules of infant phonetics development on the basis of large corpus analysis.

Keywords: speech analysis technology, matlab speech analysis platform, infant voice, pitch, fundamental frequency

INTRODUCTION

Experimental evidence suggests that language acquisition has a universal organic basis, based on the same development of language perception pathways observed across different languages, populations, and cultures [1-3]. Language is the most important communication tool for human beings, the development and change of language carries the history of human development and progress. As early as the Warring States period, the Confucian scholar Gu Liang Chi in the Spring and Autumn Gu Liang Biography said people were human because we could talk. If language is so important, why is it that only humans can make sounds, and how do humans acquire language? Infants from birth to babbling, in just a few years to master the expression and understanding of mother tongue spoken language, and what process they experienced? Babies and very young children almost miraculously develop the ability to speak without obvious effort or even teaching, while teenagers or adults struggle in a foreign language classroom and never seem to reach the same level of proficiency in their native language as a five-year-old. People take it for granted that language acquisition is something that comes naturally, just as any normally developing person can learn to walk, at some point, without obvious effort or study people can always accomplish the same things in a certain language, no matter what language they grow up in. The great linguist Noam Chomsky (1994) once said, "We are born to walk, and it is impossible for others to teach us to walk. " It's the same with language, no one teaches it, and you can't actually stop a child from learning it. "

Human beings need to go through a long process of language learning and development, and there is a fixed period that is easier to acquire language than any other period, and the infant period is the key period of language acquisition [4,5]. The study of children's language is a worldwide topic and a special case in language research, which has attracted the attention of scholars all over the world. For more than two hundred years, scholars around the world have conducted research on children's learning and understanding of language from different perspectives and different methods. The study of children's language has gone through four different stages, beginning with auditory recognition, adopting the method of biographical journals in the late 19th century, applying digital technology to the recording and playback of audio signals in the 1950s, and entering the stage of widespread use of computer-aided signal processing methods and various physiological instruments and speech analysis software after the 21st century [6-8]. The study of children's language involves all aspects of language study, such as phonetics, vocabulary, syntax, grammar, and pragmatics. The study of infant language can not only tell us what language is, but also show us the role of language in human life. Therefore, the study of infant language has become the focus of many linguists and has certain academic value.

The human language system is a dynamic complex adaptive system, while linguistics is the science of human language Linguists develop and test scientific hypotheses, and many linguists advocate the use of statistical

analysis, mathematics, and logical formalism to explain the linguistic models. In order to further study children's speech development, this study uses an advanced speech analysis platform based on MATLAB software. The platform can automatically segment speech samples and analyze relevant acoustic parameters. Through this method, we can accurately track and record the phonetic changes of infants and young children, thereby gaining a deeper understanding of the phonetic characteristics of young children. Speech analysis uses complex signal processing and speech acoustic analysis algorithms to identify and quantify the speech pronunciation parameters of infants and young children of different ages, such as the fundamental frequency parameter of measuring pitch and the formant parameter of analyzing vowel pronunciation. The application of signal processing and acoustic analysis methods to the study of phonetic pronunciation provides us with a richer and more systematic perspective, which can help us better study the uniqueness of infant phonetic pronunciation and try to summarize the regularity from the analysis of a large number of data.

EXPERIMENTAL METHODS

The research of this paper is based on phonetics, speech physiology and speech signal processing technology, and the research method mainly adopts the combination of modern speech experiment and traditional listening experiment. The speech experiment mainly uses speech software and language analysis platform to analyze and study the parameters of daily pronunciation of 4 2-year-old infants. In the analysis, clear samples are first segmented and classified, and then required acoustic parameters were extracted and analyzed through data statistics and charts. Finally, the characteristics and the overall change of the sound expression in infants under 2 years old were summarized. The listening experiment mainly adopts some traditional methods, such as listening and recording, to record the order of children's speech acquisition and calculate the accuracy of children's pronunciation.

Speaker Information

The corpus collected by the audio recording was tracked, and the corpus collected by the audio recording was selected on the subjects. Four infants under 2 years old with similar age were selected as the research objects, among which three were boys and one was girls. All the four infants were born full term and healthy. On the one hand, the selection of subjects of the same age was conducive to long-term and uninterrupted recording for better comparability of the results. The main recordings were made by the infant's parents and the author herself. The sampling interval was every few days, and the recording time was no less than 30 minutes. The duration of each recording varied depending on the infant's language ability and willingness to pronounce. Table 1 shows the basic information of 4 subjects.

Table 1. Relevant information of subjects

Name	Gender	Number of Recordings	First Language
Ss1	Male	350	Standard Chinese
Ss2	Male	179	Standard Chinese
Ss3	Male	123	Standard Chinese
Ss4	Female	461	Standard Chinese

Corpus Construction

The construction of corpus includes corpus design, recording, segmenting and sample classification. The construction process of the corpus is shown in Figure 1. In the process of the establishment of the corpus, the operation of any part has its corresponding special software implementation, and different units use different software in the design process of the corpus.

The original data contents in the corpus include: (1) Establish a voice expression corpus for infants under two years old. The content of the corpus is divided into emotional expression and vocal expression, in which emotional expression includes crying signal and laughter signal, while verbal vocal expression is mainly speech signal. Crying is mainly physiological crying laughter is divided into spontaneous vocal laughter and induced laughter. (2) There are a total of 1113 voice fragments in the corpus, of which 350 are recorded in Ss1, 179 in Ss2, 123 in Ss3, and 461 in Ss4. The duration of each voice fragment is different, and each corpus is not less than 5 minutes,

and the longest time is not more than 30 minutes. If each corpus is 5 minutes, the recording duration of all corpus is at least 92 hours.

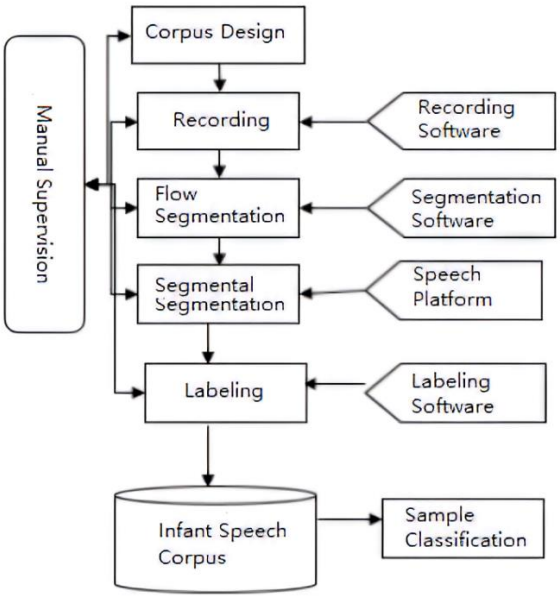

Figure 1. Diagram of corpus construction

Main Function Modules of the Voice Analysis Platform

In order to conduct more in-depth research on speech signals, we have written a speech signal analysis platform with Matlab under the Windows platform [9], which is used to carry out targeted analysis, marking and extraction and preservation of relevant parameters of the collected signals. Its main functions include: wav file reading, signal marking, mark automatic saving and marked voice file reading, calculation and saving time, formant, fundamental frequency and other related parameters, parameter file batch processing, etc. The modules of the platform include:

Signal reading and display module

The following functions are mainly completed: read the speech signal with waveread () function, calculate the energy according to the frame, then draw the energy graph. By Fourier change, the signal is converted from the time domain to the frequency domain, finally displayed by broadband three-dimensional graph. The horizontal axis represents time, the vertical axis represents frequency, and the third dimension using color depth represents energy. The basic processing functions including signal amplification, reduction, cutting, saving, etc.

Smoothing of speech signals

Infants' speech signals are unstable, thus many infant speech signals are calculated with a lot of subtle high-frequency noise, the low-pass smooth filtering is carried out. in order to reduce the impact of noise. The filter adopts zero-phase digital filter filtfilt (b, a, x). The filtfilt calls the filter function to complete zero-phase digital filtering by reverse and forward processing of the input data. The filtering process is shown in Figure 2.

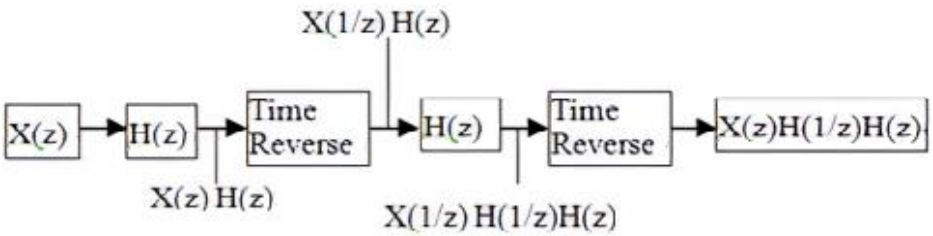

Figure 2. Filtering process of speech signal

Extraction of formant parameters

From the point of view of algorithmic, we use two ways to estimate the formant with LPC: one is to find the root method. First we calculate the prediction coefficient (that is the coefficient of the prediction error filter A (z)) with LPC analysis, next is to find the roots of A (z), these roots are the poles of the frequency characteristics of the sound channel, the bandwidth and frequency of the formant can be calculated from these poles. Second approach is the peak selection method. First of all, we use the LPC analysis method to find the speech spectrum envelope, then search the local maximum on the envelope which corresponding to the formant.

Extraction of fundamental frequency parameters

The fundamental frequency parameter represents the speech intonation in linguistic sense, but the the change of the pronunciation organ will slightly affect intonation level, the change of the resonance cavity will cause the change of the distribution of nasal energy and accent energy, and explain the relationship between tone and nasal degree from the perspective of speech production. In this paper, we adopts the traditional short-time auto-correlation function algorithm, the formula one is as follows:

$$R_n(j) = x_n(m)\omega(n-m)x_n(m+j)\omega(n-m-j), 0 \le j \le p \quad (1)$$

In the formula, Rn(j) represents the auto-correlation function corresponding to time “n”, the independent variable “j” represents the lag time of the auto-correlation function, and the subscript “n” indicates that the short-time auto-correlation function is calculated for the NTH paragraph of speech. Obviously, the value of n should change every 10~20ms, thus the auto-correlation function is calculated every other frame time (a total of p+1)

Implementation of Speech Analysis Platform

The speech signal analysis platform developed by Matlab can realize wav file reading, signal marking and signal segmentation, etc. The platform can automatic saving of segmentation files and reading of marked voice files; Calculate formant, duration and fundamental frequency parameters, and save the parameters; Solid line parameter batch processing function.

RESEARCH ON INFANT VOICE PITCH

Children's language consists of simple, repetitive utterances, spoken slowly and with long pauses [10,11], in addition, it also has high pitch, wide range, short duration and other acoustic characteristics [12]. Pitch (basic frequency, F0) is a key phoneme of our voice [13,14], this paper makes a comparative analysis of the pitch characteristics of 4 infants' speech from four stages. In this stage, infants start to make some coo sounds similar to back vowels or round lip sounds, and gradually develop into strings of connected sounds and sentences, which are as follows: 8 to 20 weeks, 21 to 50 weeks, 1 year to 1 year and a half years, 1 year and a half years to 2 years old, of which 1 year old is the pre-language stage, 1 year to 2 years old is the language development stage.

The analysis method is as follows: First, the speech samples of four stages were selected from the daily speech corpus of the subjects through secondary segmentation, including 50 speech samples in the first two stages, 75 from 1 to 1 and a half years old, and 55 from 1 and a half years old to 2 years old. A total of 230 valid samples were obtained, excluding cries, laughter and other growth noises; Secondly, the Wave final speech analysis software is used to label each speech sample. Finally, the speech analysis platform was used to extract the pitch curves of each stage of the speech samples, and the pitch patterns of the children at this stage were counted.

Analysis of Infant Voice Pitch Energy

In order to further analyze the development and change of children's pitch pattern, the fundamental frequency values of 30 measuring points of each tone of the speaker were calculated using the fundamental frequency extraction platform, and the values were input into the SPSS statistical software to analyze the energy distribution of pitch of 4 infants in the four stages respectively. Since the pitch patterns of infants in the first three stages were mainly flat and out. In order to make a comparative analysis between the stages, only the tone energy distribution (TED) of level tone and falling tone is analyzed in this paper, statistical data are shown in Figure 3 and Figure 4.

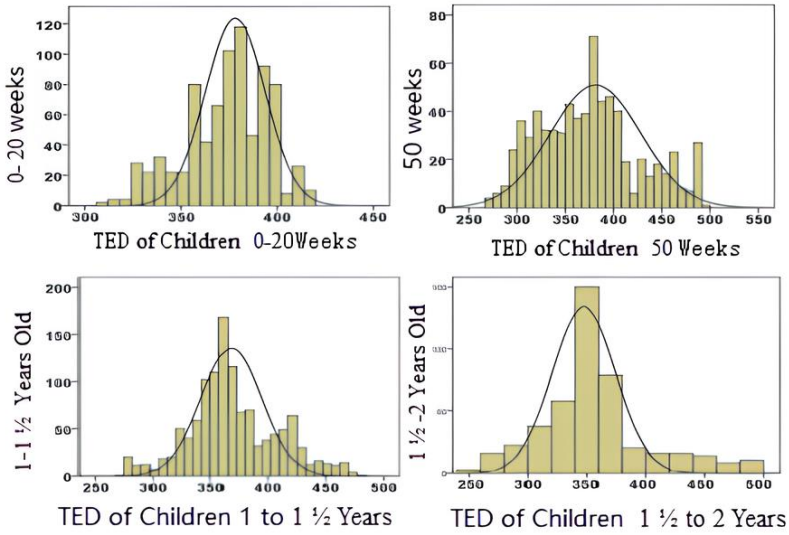

Figure 3. Level tone energy distribution of children 0-2 years

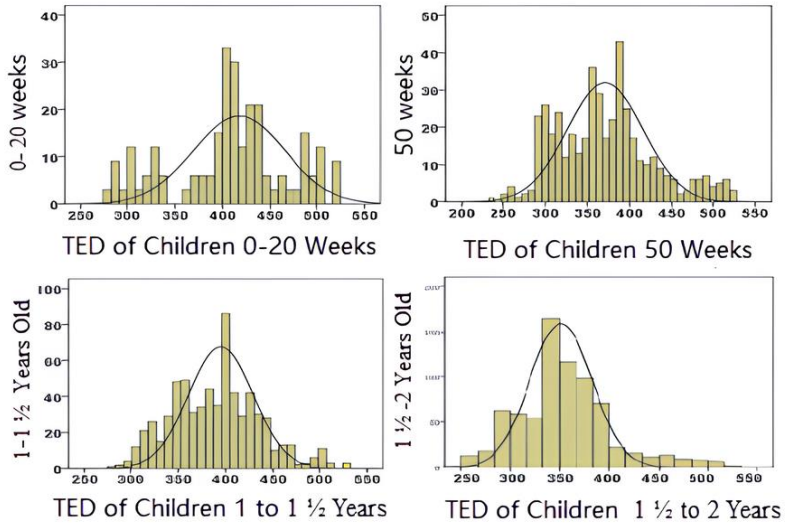

Figure 4. Falling tone energy distribution of children aged 0-2 years

8-20 weeks

At this stage, the pitch pattern of children is mainly flat tone. A total of 31 effective level tones and 9 falling tones were counted from 50 speech samples, and a total of 31 30+9 30=1200 effective dates were obtained. From the tone energy distribution of 8-20 weeks, the highest energy of level tone is concentrated around 380Hz, accounting for about 13% of the total number of samples, followed by about 370Hz, accounting for about 9% of the total number of samples. In this stage, the number of samples to falling tone is relatively small, and the highest energy is concentrated around 410Hz, accounting for 20% of the total number of samples

21 weeks to 50 weeks

This stage including 25 level tones and 16 falling tones, a total of 25 30+16 30=1230 valid dates were obtained. In this stage, the overall amplitude of the level tone energy decreases, and the gap between each frequency domain segment decreases. The highest energy is concentrated around 360-370Hz, accounting for about 8% of the total sample number, followed by about 390 Hz, accounting for about 6% of the total sample number, and the high-frequency energy of the sound is concentrated around 390 Hz, accounting for about 10% of the total sample number.

1 Year -1 and a half years old

In this stage, there were 41 level tones, 24 falling tones, a total of 41 30+24 30=1950 effective dates. This stage, with the growth of the larynx and vocal cords, children's language ability significantly improved, level tone energy overall decline, high frequency energy concentrated in 360 Around Hz, accounting for about 10% of the total number of samples, followed by around 370 Hz, accounting for about 6% of the total number of samples, the falling tone energy is concentrated around 390 Hz, accounting for about 11% of the total number of samples.

1 ½ -2 years old

In this stage, a total of 15 level tones and 25 falling tones are counted, and a total of 15×30+25×30=1200 effective dates are obtained. In this stage, the fundamental frequency range of children's level tone is mainly concentrated around 350 Hz, while frequency energy of falling tone is mainly concentrated between 350 Hz and 360 Hz, and the overall fundamental frequency energy is significantly reduced compared with the previous stage.

According to the data energy statistics of each measuring point of the two types of tones in the above four stages, the development and change of children's pitch are closely related to physiological characteristics [15-17], and the energy of level tone in the four stages shows a downward trend as a whole, and the energy curve in the figure also gradually flattens out.

Statistical Analysis of Infant Voice Pitch Contour

The mean value of 30 measuring points of each tone in the four stages was calculated by SPSS analysis software, and the final mean value of the fundamental frequency was used to draw the pitch contour distribution diagram of each stage.

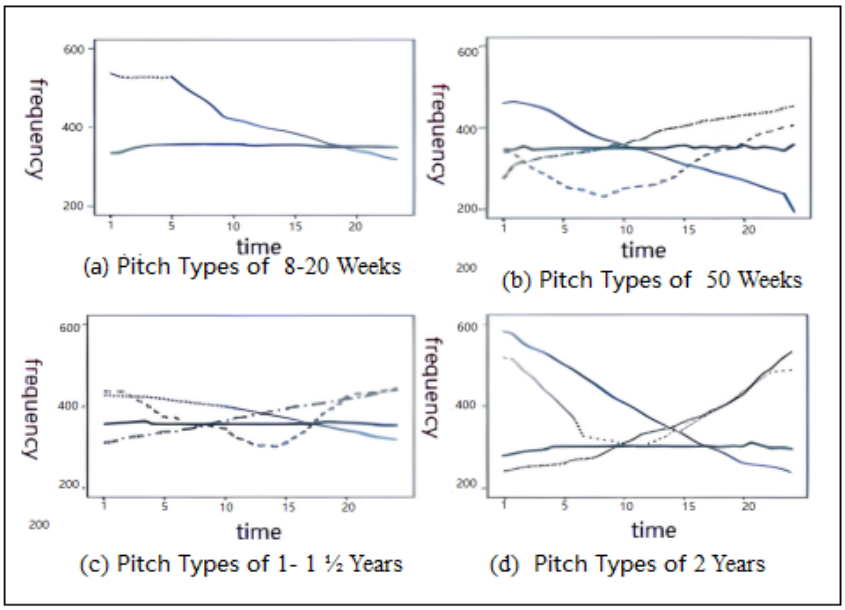

Figure 5. Pitch types of infants at four stages

The figure above are the tone distribution diagram of children at the four growth stages of 8-20 weeks, 21-50 weeks, 1-1/2 years old, and 2 years old respectively. Before 20 weeks old, children only have level tone and falling tone, while between 21 weeks and 50 weeks old, rising tone and falling-rising tone appeared. On the whole, the fundamental frequency energy of the child before the age of 2 is relatively high, and the tuning range is mainly between 300 and 500Hz. Through the analysis in the figure above, the conclusion can be drawn as follows.

First, the pitch pattern of infants at 8-20 weeks was mainly level tone and falling tone. From 21 weeks to 50 weeks, the pitch pattern was mainly level tone and falling tone, with a few rising tone appeared. At the age of 1 to 1 and a half, the main pitch patterns were still level tone and falling tone, and the number of level tone increased significantly. At the age of one and a half to two years old, children appear four tones of Mandarin Chinese, and the change of tone level has a function of meaning discrimination.

Second, child has been able to gradually issue the four tones of Mandarin Chinese during the period of 6 months to 1 and a half years old, of which the number of level tone is the largest, followed by the falling tone and the level tone and the falling-rising tone. This also shows that it is the most difficult for children to acquire the rising tone and falling-rising tone. Due to the size of the vital capacity and height, weight has a great relationship for sound production, children's vital capacity and the number of alveoli are less than adults [18], so there is not enough air flow for the production of rising tone. The production of falling-rising tone has two stages with first falling and then rising, thus the pronunciation is more difficult until the child at 1 ½ -2 years old who has been able to issue the tone of the four basic tones in the mother tongue.

Comparative Analysis of the Pitch Patterns of Infants' Crying and Speech Signal

Babies come into the world with crying [18]. It can be said that crying is the first vocal expression of babies [19,20]. Before babies have no language expression ability, they have very limited ways to convey information, so crying becomes an important way for them to express various requirements and wishes, full of rich emotional colors Babies' cries can express different physiological needs and reactions. Crying is a special language of babies. Different babies' cries have certain common characteristics and individual characteristics.

The production of infant crying requires the cooperation of three important systems: respiratory system, vocal fords and vocal tract, the main source of crying is the vibration of the vocal folds of the pharynx cavity. In this paper, the cry signal of 4 infants will be collected, all kinds of cries of infants in different stages will be recorded respectively, and the acoustic parameters of cry units of 4 infants at 4 stages will be classified respectively, and the speech analysis software Speech lab will be used to mark each crying unit with a period mark, and the voice platform developed by Matlab will be used to extract the fundamental frequency curve, as showing in Figure 6.

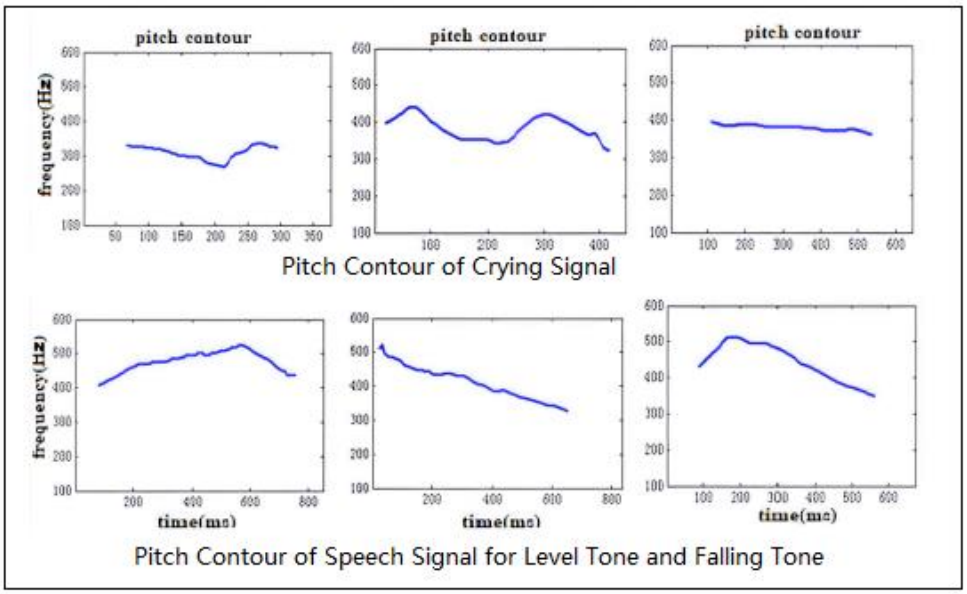

Figure 6. Pitch patterns of infants' crying and speech signal

Although crying and voice are both vocal expressions, but actually they are quite different [21,22]. Compared with crying, the expression of voice is ever-changing, voice pronunciation is not only the resonance of the air flow in the voice cavity, but also the tuning of the mouth, the types of pronunciation and pronunciation methods are diverse. However, the types of crying are different and much simpler than the voice, many cries are the constant repetition of crying units with regularity, and the mouth is tuned to the crying sound [23,24]. Because babies' vocalizations are influenced by their physiological structure, crying sounds and speech signals are very different. First, from a physiological point of view, infants' vocal cords are short and thin, so the F0 of infants is generally higher than that of voice. Compared with voice, the fundamental frequency range of infant cries is between 400-500Hz, which is significantly higher than the fundamental frequency of voice. Second, the voice segment of crying is less correlated before and after crying, and there are multiple sound units.

SUMMARY

In this paper, we used the speech signal analysis platform to analyzing the pitch pattern and pitch fundamental frequency characteristics of infants and young children from birth to 2 years old. Through a long time of speech recording, speech segmentation, speech classification and speech analysis, we analyzed the vocal expression in the pre-speech stage and speech signals in the speech stage of infants and young children. It is concluded that the pitch pattern of infants and young children is gradually complicated with the increase of age, from a single falling tone to 4 basic tones, and the fundamental frequency energy value shows a downward trend with the increase of age, the change of fundamental frequency ability is related to the maturity of infants and young children's physiological vocal organs, and the fundamental frequency value decreases with the growing and changing of vocal fords. The corpus of this paper is based on four speakers, and it is hoped that there will be more subjects in the future in order to get more general conclusion about pronunciation in young children and the origins of human speech. units.

ACKNOWLEDGMENT

This paper was subsidized by Central Universities Fundamental Research Funds for the project: The Ideological and Political Construction and Practice of College English Courses in Ethnic Universities under the guidance of the Consciousness of the Chinese Nation Community (31920230109).

REFERENCES

[1] Hoff, E. Language development at an early age: Learning mechanisms and outcomes from birth to five years. Encycl. Early Child. Dev. 7-10, 2009.

[2] Arenillas-Alcón, S., Costa-Faidella, J., Ribas-Prats, T., et al. Neural encoding of voice pitch and formant structure at birth as revealed by frequency-following responses. Sci Rep 11, 6660, 2021.

[3] Richard, Neel ML, Jeanvoine A, et al. Characteristics of the frequency-following response to speech in neonates and potential applicability in clinical practice: A systematic review. J. Speech, Lang. Hear. Res. 63, 1618–1635, 2020.

[4] E. S. Gopi. Digital Speech Processing Using Matlab (Signals and Communication Technology) 2014th Edition. Springer, 2013.

[5] Leipold S, Abrams DA, Menon V. Mothers adapt their voice during children's adolescent development. Sci Rep. 2022 Jan 19; 12 (1): 951.

[6] Ferguson CA. Baby talk in six languages. Am. Anthropol. 1964; 66: 103–114.

[7] Fernald A, Simon T. Expanded intonation contours in mothers' speech to newborns. Dev. Psychol. 1984; 20: 104–113.

[8] Piazza EA, Iordan MC, Lew-Williams C. Mothers consistently alter their unique vocal fingerprints when communicating with infants. Curr. Biol. 2017; 27: 3162–3167.

[9] Levrero F., Mathevon N., Pisanski K. et al. The pitch of babies' cries predicts their voice pitch at age. Biology Letters, Volume 14, Issue 7, 2018.

[10] E. S. Gopi. Digital Speech Processing Using Matlab (Signals and Communication Technology) 2014th Edition. Springer, 2013.

[11] Benedict, H. Early lexical development: comprehension and production. Journal of Child Language, Vol. 6, No. 02, pp. 183-200, 1979.

[12] Cabrera, L. & Gervain, J. Speech perception at birth: The brain encodes fast and slow temporal information. Sci. Adv. 6, eaba7830, 2020.

[13] Liquan Liu, Antonia Götz, Pernelle Lorette. How Tone, Intonation and Emotion Shape the Development of Infants' Fundamental Frequency Perception. Psychology of Language, Volume 13, 2022.

[14] Arvaniti, A., and Fletcher, J. "The autosegmental-metrical theory of intonational phonology," in The Oxford Handbook of Language Prosody. eds. C. Gussenhoven and A. Chen, Oxford: Oxford University Press, 78–95, 2020.

[15] Bryant, G. A. The evolution of human vocal emotion. Emot. Rev. 13, 25–33, 2021.

[16] Polver, S., Háden, G. P., Bulf, H., et al. Early maturation of sound duration processing in the infant's brain. Sci Rep 13, 10287, 2023.

[17] Trinh N., Susanne R., Anja L., et al. Sing to me, baby: Infants show neural tracking and rhythmic movements to live and dynamic maternal singing. Developmental Cognitive Neuroscience, Volume 64, 2023.

[18] A. Attaheri, Á. N. Choisdealbha, G. M. Di Liberto, et al. Goswami. Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. NeuroImage, 247. 2022.

[19] Courtney B. Hilton, Cody J. Moser, Mila Bertolo, et al. Acoustic regularities in infant-directed speech and song across cultures. Nature Human Behaviour, 2020.

[20] Soltis, J. The signal functions of early infant crying. Behavioral and Brain Sciences 27, 443–458, 2004.

[21] Farran, L. K., Lee, C. -C., et al. Cross-Cultural Register Differences in Infant-Directed Speech: An Initial Study. PLOS ONE 11, e0151518, 2016.

[22] Byers-Heinlein K, Tsui A S M, Bergmann C, et al. A Multilab Study of Bilingual Infants: Exploring the Preference for Infant- Directed Speech. Advances in Methods and Practices in Psychological Science 30, 2021.

[23] Abdellah K, Francis G, Juan RO, et al. Principal component analysis of the spectrogram of the speech signal: Interpretation and application to dysarthric speech. Computer Speech & Languag, 59:114-122, 2020.

[24] Chen, A., Stevens, C. J., and Kager, R. Pitch perception in the first year of life, a comparison of lexical tones and musical pitch. Front. Psychol. 8:297, 2017.