Intelligent acoustic data fusion technique for information security analysis

Ying Jiang, Yize Tang, Wenda Lu, Zhongfeng Wang, Zepeng Wang and Luming Zhang

1State Grid Zhejiang Information & Telecommunication Company
2Zhejiang Huayun Information Technology CO., LTD
3Department of CSIE, Hefei University of Technology

Abstract. Tone is an essential component of word formation in all tonal languages, and it plays an important role in the transmission of information in speech communication. Therefore, tones characteristics study can be applied into security analysis of acoustic signal by the means of language identification, etc. In speech processing, fundamental frequency (F0) is often viewed as representing tones by researchers of speech synthesis. However, regular F0 values may lead to low naturalness in synthesized speech. Moreover, F0 and tone are not equivalent linguistically; F0 is just a representation of a tone. Therefore, the Electroglottography (EGG) signal is collected for deeper tones characteristics study. In this paper, focusing on the Northern Kam language, which has nine tonal contours and five level tone types, we first collected EGG and speech signals from six natural male speakers of the Northern Kam language, and then achieved the clustering distributions of the tone curves. After summarizing the main characteristics of tones of Northern Kam, we analyzed the relationship between EGG and speech signal parameters, and laid the foundation for further security analysis of acoustic signal.

1. Introduction
Among all human language, tone is an essential component of tone languages, and is used to build words much as consonants and vowels do [1]. For instance, in Chinese mandarin, the syllable /hú/, when pronounced with a level pitch pattern, means “embankment”; with a rising pattern, the meaning is “whistle”; with a falling-rising pattern, the meaning is “bottom”; and with a falling pattern, the meaning is “land”. Thus, semantic distinction in tonal languages depends not only on articulatory composition, but also on tone patterns. Thus, more and more researchers pay attention to research the distinguishing characteristics of different tones. Here, in continuation of this work, we use the physiological and acoustic signals information to describe the tones in a tone language.

From a linguistic or physics point of view, the voice generation of larynx towards the lip involves three components: the voice source (phonation), the acoustic resonance (articulation), and the lip radiation (loudspeaker) (Fig. 1) [2]. The vibration of vocal cords generates speech signal, and the vowels and consonants are highly connected to the vibration difference of vocal cords, namely the phonation types. The vibration of vocal cords is a synergistic effect of different parts of a throat. During the process of generating voice, F0 is decided by the vibration frequency of vocal cords [3]. After tuning by soundtrack, mouth cavity and nasal cavity, the source signals from vibration can be made into different sounds. This is called resonance, which can be seen as a modulating process to the source signals. Finally, the sounds are transferred into the air through the radiation effect of mouth to form the speech signals. Thus, the generation of speech is a process of stimulation and modulation. In the research of traditional linguistics and speech processing, what we collect is the speech signals, and many in-depth studies have been done as well as numerous acoustic models are raised. Meanwhile, the generation of tones is also
fully analyzed. However, tone is much more than just a physical quantity. As a linguistic term, tone is a kind of rational cognition of pitch by linguist, belonging to the categorical perception.

![Figure 1. Voice generation of an Sine curve.[4]](image)

Though F0 is the frequently used physical quantity in the present speech research, F0 cannot indeed represent the original vibrational state of vocal cord with the influence of phonation. An electroglottograph (EGG) signal, for example in Fig. 2, can be used to measure the vocal information directly and index the degree of contact between the vocal folds, so it has been a useful tool in measuring and describing phonations [5-11]. Childers and Lee used EGG measures to extract source related features for modal, vocal fry, falsetto and breathy voice, which are produced by subjects with and without vocal disorders [12]. Four factors were found to be the most important in distinguishing the above four phonation types: glottal pulse width, pulse skewing, abruptness of glottal closure, and turbulent noise. Results for glottal pulse width showed that modal phonation was produced with medium glottal width, vocal fry with a short glottal width, falsetto and breathy voice with a long glottal width. For pulse skewing, modal phonation was characterized by medium skewing, vocal fry with high skewing, falsetto and breathy voice with low skewing. In terms of abruptness of closure, modal voice and vocal fry were characterized by abrupt closure, while falsetto and breathy phonations were characterized by progressive closure. Finally, breathy phonation was the only phonation produced with high turbulent noise.

![Figure 2. The voice signal and Electroglottograph in Northern Kam Language /?a5 $/ and /la22/](image)

2. Methods

**Corpus:** In this paper, a corpus of synchronous EGG and speech signals is established. It is acquired
under clean environment from a 53-year-old man who is familiar with Kam language and mandarin. Speech signals are recorded by FieldPhon software with sampling frequency 44100Hz, and the strict processing of listening, identifying and tuning is implemented after recording. The corpus includes 770 monosyllable words, covering all tones on 9 unchecked syllables and 4 checked syllables in Northern Kam Language. Moreover, the universality and proportionality of the samples of tones, consonants and vowels are ensured as far as possible. EGG signals are synchronously acquired with the recording of speech signals.

**Data processing:** Seven kinds of F0 parameters are acquired based on the basic acoustic parameters extracted by STRAIGHT [22]: 1) the amplitude difference between starting point and end point of fundamental frequency (pitch variation), 2) the length of fundamental frequency curve (duration), 3) the maximum of fundamental frequency, 4) the minimum of fundamental frequency, 5) the extreme points of fundamental frequency curve, 6) the slope of fundamental frequency curve, 7) the slope variation of fundamental frequency curve. Meanwhile, two parameters of EGG, CQ and DECPA, are obtained based on the parameters including CQ, Peak Vel and Min Vel extracted by EggWorks and processed with Matlab.

**Analysis Methods:** The rule of minimum opposition is adopted for analyzing the experimental data so as to accurately attain the influence of EGG parameters and acoustic parameters on the acoustic characteristics. There is only one kind of characteristic difference (tone, consonant or vowel difference) for data comparison. For example, when the opposition characteristic is consonant, the rule of minimum opposition ensures the consistency of vowel and tone. The opposite characteristic is tone, and the consonant and vowel are consistent. During analysis, the pathological voice is ignored while the outliers and interferences of erroneous recording are eliminated.

The method of minimum opposition used in this paper for one person and one language system can also be applied to multi-person and multi-language system. However, the person and the language type should be respectively taken as the condition of opposition for analysis and discussion.

**3. Results**

**3.1. Analysis of Tone Identification by EGG Signal**

At present, the tone definition is mainly confined to the value and modes of pitch. The value of pitch in physiology field corresponds to the vibration speed of vocal fold, and in acoustics or information field corresponds to the value of the fundamental frequency. On the other hand, tone is a linguistic term. We can think that the tone is the combination of phonation type and vibration frequency. Phonation type is a suprasegmental feature related to timbre, while vibration frequency determines the value of fundamental frequency. As shown in Fig. 3, with the influence of resonance, labial radiation and consonant, it is often difficult to accurately acquire the value of F0 through the speech signal. Meanwhile, EGG can obtain the glottal information of larynx directly without the influence of the above disturbances. So the value of F0 extracted by EGG signal is more precise than by the speech signal. As shown in Table 1, F0 extracted by EGG signal has smaller variation range and is smoother compared to F0 extracted by the speech signal.
Figure 3. F0 extracted by EGG and speech signals. The red line indicates F0 extracted by speech signal, and the blue line indicates F0 extracted by EGG. F0 extracted by EGG has smaller variation range and is smoother.

Table 1. Analysis of F0 of /wən/ extracted by EGG and speech signals. F0 slope is calculated after fitting for F0 values. F0 range is the variation range of F0 values. DRF is the variation range after five points smoothing for F0 values.

|                        | F0 slope | F0 range | DRF   |
|------------------------|----------|----------|-------|
| F0 extracted by EGG    | 0.027    | 11.27Hz  | 6.17Hz|
| F0 extracted by speech | 0.167    | 31.92Hz  | 21.64Hz|

3.2. Analysis of Acoustic and EGG Parameters of Checked and Unchecked Syllables
In the Northern Kam Language, like most Eastern Asian languages, the same tone value can correspond to the checked and unchecked syllables. For example, as shown in Fig. 3, tone7 is on unchecked syllable and tone10 checked syllable. They have the same tone value 55, but with the different durations, the ends of the vowel and the variation ranges of F0. In Tai-Kadai languages, the tones on checked and unchecked syllables differ in duration, and the checked syllables end in the consonants, leading to big difference in the phonation type.

Figure 4 and Table 2 show the parameters of the checked and unchecked syllables with the same tone value. We can find that DECPA of the checked syllable is higher than that of the unchecked syllable with the same tone value, but the trends of DECPA of the checked and unchecked syllables are similar. Furthermore, from Table 2, the difference between the values of DECPA of the checked and unchecked syllables is beyond 100. The F0 slope of the checked syllable is also higher, which is related to the higher DECPA of the checked syllable. Meanwhile, the values of CQ of the checked and unchecked syllables with the same tone value are approximately equal. From Fig. 4, the vowel is /a/ in syllables with different tone values, and the values of CQ are around 0.55.
Figure 4. EGG parameters of the checked and unchecked syllables with the same tone. The red line indicates the checked syllables, and the blue line indicates the unchecked syllables.

Table 2. Acoustic and EGG parameters of unchecked and checked syllables.

| tone value | 55   | 31   | 13   | 25   |
|------------|------|------|------|------|
|            | unchecked | checked | unchecked | checked | unchecked | checked | unchecked | checked | 5.588 | 0.585 | 0.564 | 0.587 | 0.552 | 0.589 | 0.571 | 0.524 |
| Sa         |       |      |      |      |      |      |      |      | 636.77 | 682.84 | 431.03 | 639.78 | 275.91 | 391.65 | 365.78 | 432.30 |
| F0 range   | 0.1925 | 0.0171 | -1.2271 | -1.2523 | 0.4467 | 0.5925 | 1.2591 | 3.8649 |

3.3. Analysis of Similar Tones

The similar tones, for example, the tone4 and tone9 shown in Fig 4 have much resemblance in tone contour. The curves of them are nearly parallel, and the durations of them are approximately equal, with only the values of F0 are different. For these similar tones with little difference in tone characteristics, it is difficult for the current tone recognition algorithms to distinguish them.

The parameters DECPA and CQ of similar tones (tone4, tone9) for /tu/ and /ta/ respectively. DECPA can distinguish the similar tones with the same vowel and consonant. For /tu/ DECPA of tone4 is larger than that of tone9, and the values of DECPA of the two tones are clearly distributed in two intervals. For /ta/ DECPA of tone9 is larger than that of tone4. At the same time, the distributions of DECPA for /tu/ and /ta/ are also different as the result that the articulation positions of /a/ and /u/ are different. So DECPA is related to not only the energy change of pronunciation but also the articulation positions.

Meanwhile, the similar tones cannot be well discriminated by CQ in /tu/ and /ta/. In addition, the curves of CQ of different tones in the same vowel are approximately similar. But the values of CQ of different vowels in the same tone are inconsistent, which may imply that CQ to some extent has distinction of the vowels.

4. Conclusion

Although EGG is a widely used instrument in physiology and medicine, it is still a relatively new means of measurement in the field of linguistics and speech. EGG signals were initially used in linguistics to examine phonation types; the EGG parameters CQ and DECPA have been mainly used to measure the opening and closing glottis states, as well as velocity variations. Because Northern Kam is a relatively rich tonal language, it is very meaningful to study its tones. In this paper, we collected EGG signals of...
all the monosyllables in the language, and then we analyzed the relationship between the EGG signals and tones.

5. References
[1] Gang P., William S.-Y. Wang. Tone recognition of continuous Cantonese speech based on support vector machines. Speech Communication 45.1, 49-62 (2005).
[2] Kong J.P. On Language Phonation. Central Nationalities University Press, Beijing, 274-291 (2001).
[3] Yip M. Tone. Cambridge University Press (2002)
[4] Hardcastle, William J., John Laver, and Fiona E. Gibbon, Eds. The handbook of phonetic sciences. Vol. 79. John Wiley & Sons, (2012).
[5] Guion, Susan G., Mark W. Post, and Doris L. Payne. Phonetic correlates of tongue root vowel contrasts in Maa. Journal of Phonetics 32.4, 517-542 (2004).
[6] Esposito, Christina M. An acoustic and electroglottographic study of White Hmong tone and phonation. Journal of Phonetics 40.3, 466-476 (2012).
[7] Childers, Donald G., and Ashok K. Krishnamurthy. A critical review of electroglottography. Critical reviews in biomedical engineering 12.2, 131-161 (1984).
[8] Michaud, Alexis. Final consonants and glottalization: new perspectives from Hanoi Vietnamese. Phonetica 61.2-3, 119-146 (2005).
[9] Esposito, Christina M. An acoustic and electroglottographic study of phonation in Santa Ana del Valle Zapotec. Poster presented at the 79th meeting of the Linguistic Society of America, San Francisco, CA. (2005).
[10] Michaud, Alexis, and Martine Mazaudon. Pitch and voice quality characteristics of the lexical word-tones of Tamang, as compared with level tones (Naxi data) and pitch-plus-voice-quality tones (Vietnamese data). Proceedings of Speech Prosody 2006 (2006).
[11] DiCanio, Christian T. The phonetics of register in Takhian Thong Chong. Journal of the International Phonetic Association 39.02, 162-188 (2009).
[12] Childers, Donald G., and C. K. Lee. Vocal quality factors: Analysis, synthesis, and perception. The Journal of the Acoustical Society of America 90.5, 2394-2410 (1991).
[13] Vu-Ngoc, Tuân, Christophe d’Alessandro, and Alexis Michaud. Using open quotient for the characterisation of Vietnamese glottalised tones. Proceedings of Eurospeech-Interspeech 2005: the 9th Annual Conference on Speech Communication and Technology. 2885-2889 (2005).
[14] Michaud, Alexis. A Measurement from Electroglottography: DECPA, and its Application in Prosody. Speech Prosody 2004, International Conference. 633-636 (2004).
[15] Long Y. Research of Kam Language Dialect. Shanghai Normal University, P R China. (2012).
[16] Long Y. and Zheng G. The Dong language in Guizhou Province, China. Arlington: SIL and University of Texas at Arlington (1998).
[17] Wang CC, Ding QL, Tao H, Li H. Comment on “Phonemic diversity supports a serial founder effect model of language expansion from Africa”. Science. 335.6069,657 (2012.).
[18] Chao, Y. R. A system of tone letters. Le maître phonétique 45, 81-83 (1980).
[19] SY, William. Phonological features of tone. International Journal of American Linguistics, 93-105 (1967).
[20] Shi, Lin. An experimental analysis of the five level tones of Gaoba Dong language. Journal of Chinese Linguistics 15.2, 335-361 (1987).
[21] Xiaonong Z. Phonetics. The Commercial Press. Beijing (2010).
[22] Han X., Li L.,Pan WY . Computer based field investigation and processing system for language. Journal of Tsinghua University (Science and Technology). 53.6, 888-892 (2013).