

Analysis on Application Field of Ultrasonic Imaging Technique in Linguistic Study

ZHANG Jinxi⁺, LI Yonghong and KOU Yun

Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education,
Northwest Minzu University, Lanzhou, Gansu 730030, China

Abstract. Based on various use method and technical means in linguistic study, this paper mainly discusses the application of ultrasonic imaging technique in linguistic study, particularly the study on articulation physiological part of tongue body, introduces the ultrasound imaging equipment, summarizes the application area in linguistic study and introduces the latest development of this technique in linguistic study, thus providing references for relevant researchers.

Keywords: ultrasonic imaging, linguistic study, physiological phonetics

1. Introduction

New experimental methods and experimental installations play very important role in linguistic study, interpreting many linguistic phenomena in a more scientific and reasonable way. In the aspect of language cognition, it discusses the brain mechanism in the speech production and language acquisition process from multi aspects by using myoelectricity electroencephalograph, eye-tracking and other equipment according to linguistics, cognitive science and computer science. In the aspect of phonetic physiology, it obtains multiple physiological signals in articulation, such as speech, voice, breath, myoelectricity, air current, palatal, nasal flow and lip by using airflow barometer, electroglottography, respiratory band sensor, nasal flow meter, palatal instrument, dynamic electropalatography, ultrasound recorder analyzer and 3-D motion capture equipment for the study of speech physiological multimodality. In the aspect of voice acoustics and oral culture, voice recording equipment, non-linear editing, professional video recorder and teleprompter can be used for the construction of voice resource base digitalization and acoustic analysis of voice and oral culture.

Advanced experimental installations and experimental methods play very important role in interpreting language phenomena, which is of theoretical and practical significance for us to understand the principle of human speech production mechanism particularly in the aspect of the study on phonetic physiology.

2. Ultrasound imaging experimental installation

As early as the late 1960s, ultrasound was used in phonetics research^[1] and linguistic training^[2]. Ultrasound can capture the dynamic tongue shape, causing the language phenomenon that is difficult to understand due to the interaction of tongue root, sagittal furrow, vowel and consonant. With the development of technology, ultrasound has changed from one-dimensional to two-dimensional mode(B mode, Figure 1 is ultrasound device widely used in laboratory and clinic) and now three-dimensional mode(time resolution is too low for measuring most natural languages). High-frequency (3-16MHz) sound waves of diagnostic ultrasound can easily penetrate skin, fat and muscle, but they are absorbed by bones and reflect the boundary of air. The ultrasonic detector needs to be close to the chin below the neck in order to see the tongue as full as possible. So the sound wave penetrate the tongue from below (See figure 2). The sound wave sensor

⁺ Corresponding author. Tel.: + 18754177660; fax: +0931-2977686
E-mail address: 582361484@qq.com.

passes through the jawbone and hyoid bone, and reflects the air through the tongue muscle surface. We can get the median sagittal or coronal images according to the direction of the sensor. Figure 3^[3] shows typical sagittal tongue images obtained by ultrasound, with the tongue root and hyoid shadow on the left and tongue tip and jaw shadow on the right.



Fig. 1: Ultrasound (B mode) equipment



Fig. 2: Ultrasound penetrates the tongue from below

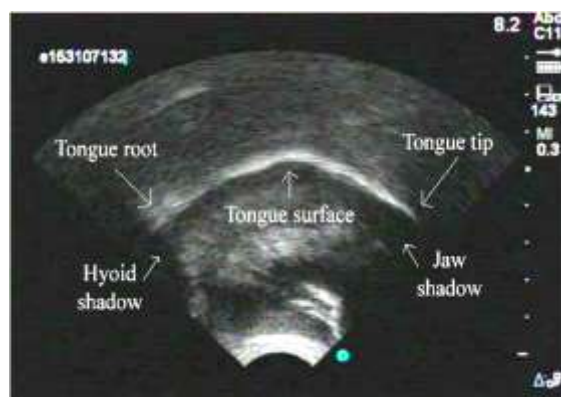


Fig. 3: Image of /z/ median sagittal tongue curve

3. Application Area of Ultrasonic Imaging Technique in Language Study

3.1. Articulation physiology

Hamlet et al^[4], who studied Laryngeal trills through the penetration by ultrasound through the throat, believed that Laryngeal trills were a kind of continuous pulsating sound, greatly changing in fundamental frequency and intensity, about 7 times per second. The changes of frequency and intensity are not always in

synchronization. The sounding pulsing is not completely regular. The periodic pattern in the ultrasound data corresponds to the intensity in the sound track. Moisik et al^{[5][6]} proposed a method to quantify the change of laryngeal height by using the optical flow analysis of laryngeal ultrasound imaging data, which is able to accurately quantify the change of laryngeal height and evaluate the influence of laryngeal height on vowel resonance frequency. Wodzinski et al^{[7][8]} carried out preliminary study on Velar fronting and Velar-vowel by using ultrasonic. It is pointed out in the research that the closing positions of Velar consonants are consistent in middle back vowel. For the front vowel, Velar front degree seems to be positively related to vowel. The measuring method of dual vowel closure position abides by back vowel mode of the word. For non-words, the closing position is affected by the bias quality of front dual vowel and the starting quality of next dual vowel. McAllister et al^[9] analyzed the Covert contrast of Velar fronting by using acoustics and ultrasound measurement method. The results showed that there were differences between Velar and alveolar consonant.

The key of articulation physiological part tongue is how to obtain the information of tongue movement. The tongue morphological data processed can interpret phonetic phenomena. Palo et al^[10] recorded the tongue movement before articulation by using ultrasound and carried auto analysis of ultrasonic recording of the tongue movement. Pixel differences^{[11][12]} is used for the analyzing the tongue movement.

3.2. Language teaching and second language acquisition

Gick^[13] describes the application of ultrasound imaging in second language acquisition and makes a brief introduction to related concept and then gives specific application case. Tsui^[14] carries out ultrasound training study on the second language English /l/ and /ɹ/ among Japanese adults. The results show that all participants improve the accuracy of various vocabulary position and English produced in context. Suzuki^[15] has developed a visual phonetic pronunciation application program, which connects the audio frequency, front and side video, MRI and ultrasound video, thus enabling users to watch articulation video and record their own audio frequency and video frequency, and play them synchronously with the model for comparison.

3.3. Field research

Hand-carried ultrasound(HCU) has been proved to be a very useful field language research tool. Hand-carried ultrasound and PC-based ultrasound, availability of digital video recording equipment and image processing and analysis software and their low price make it possible to conduct field research. Gick^[16] has carried out field investigation of language and phonetics, mainly focusing on tongue root. The first problem is about the harmony of the tongue root in Kinande(Bantu in Congo), and the second is about the haryngealization and vowel reduction in Nuu-chah-nulth(vaca mountains language on the west coast of Vancouver).

Gick, Bird^[17] has discussed the language imaging technique in related field applications by using ultrasound. This study determines the important factors of control head/sensor movement in the field environment, the range of head movement under different voice background and the influence of tongue compression on tongue imaging data.

3.4. Phonology

Adam Buchwald^[18] et al have studied the phonemic augmentation of an aphasic patient, who inserts a central vowel similar to a non-stressed vowel into a consonant (e.g. clone and c[ə]lone). This experiment makes a comparative analysis with a vowel word group(C1VC2)by using consonant (C1C2) group (e.g. clue). All articulations are recorded by ultrasound imaging. It provides the sgittal tongue imaging in the articulation process of tongue movement, as shown on the left of Figure 10. The result is shown on the right of Figure 10 and there are no obvious differences between C1C2 group(glue, clue, in active line) and C1VC2(galoot, collude, in dotted line). The perception of vowels inserted between consonants only reflects the dynamic articulation posture errors associated with the two consonants.

3.5. Dialect

Lu et al^[19] have studied the tongue movement model of Xi'an dialect by using ultrasonic tongue image. This paper classifies static Viseme of Xi'an dialect, records the temporal and spatial attributes of tongue activity in VCV and CVC speech by ultrasound imaging, and develops the relevant programs that can

automatically track the tongue movement outline, classify the extracted visual information and define the visual position for creating the dynamic visual system of the tongue in Xi'an dialect.

3.6. Coarticulation

Zharkova^[20] has carried out relevant studies on the co-articulation development of child language, describing the language co-articulation mode among Scottish English-speaking children aged 3 to 10. This study carries out tongue prediction co-articulation analysis of the four initial consonants of different restraint degree by using ultrasound tongue imaging data. Krebs^[21] has studied co-articulation problem of some vowels and consonants and the result shows that [k]it has low resistance to co-articulation and tongue root participates in the co-articulation of front vowel of the back tongue.

3.7. Speech therapy and pathological phonetics

Bernhardt, Gick et al^[22] expound the role of ultrasound in speech therapy and study the speech therapy of teenagers and adults with severe hearing impairment, continuous language barrier and English accent, and get feedback from dynamic two-dimensional ultrasound. The use of ultrasonic imaging technique can improve the accuracy of phonetic pronunciation. Michelle^[23] corrects the North American English/r/ pronunciation by using ultrasound biofeedback technology. This non-invasive biofeedback technology allows a person to see the movement of their tongue shape characteristics, thus enabling them to correct their pronunciation according to the correct articulation mode. Preston, Brick et al^[24] carry out treatment experiment among six children with apraxia of speech aged 9 to 15 years old. The visual feedback of real-time ultrasound imaging is used and children are required to adjust their tongue movement.

3.8. Speech synthesis

Jaumard-Hakoun^[25] has proposed singing sound synthesis based on ultrasound tongue imaging and lip for extraction of characteristics in ultrasound tongue image and lip image sequence and nonlinear prediction by using layered construction, thus getting the most satisfactory synthetic quality effect.

3.9. Speech recognition

Acoustic features are used in traditional language recognition. The extraction of ultrasound image and visual characteristics of optical image can also be used in language recognition. Such synthesizers can only be driven by articulation physiological data, which is called Silent Speech Interface(SSI)^[26].

Hueber^{[27][28]} acquires continuous speech recognition from ultrasound tongue imaging and optical image of lip, uses automatic speech correction program and strong visual feature extraction technology for phonetic marking in the treatment of speech corpus, estimates in visual corpus and acoustic corpus based on HMM (Hidden Markov Model) random model and makes a comparison between the performance of visual speech recognition system and traditional vision based recognition system. Wang^[29] proposes a new method to integrate dynamic information into ultrasound based silent speech interface, which weights the reliability of static and dynamic visual feature information by using Multi-stream Hidden Markov Model(MSHMM, Multi-stream Hidden Markov Model).

4. Latest development of ultrasonic imaging technique in language study

In recent three years, there are more and more studies for improving and enhancing ultrasonic imaging technique in language study and studies on language phenomena by using ultrasonic imaging technique.

Relying on the application of deep neural network (DNN) technology, new methods and new algorithms, researchers have more in-depth studies on ultrasound-based imaging silent speech interface. Csapó T G et al^[30] use the two-dimensional ultrasound of tongue as the input information of deep neural network by using DNN, thus giving the preliminary results of ultrasound tongue movement based speech conversion. Lu Wenhuan et al^[31] have proposed three hybrid characteristic extraction method based on the realization of ultrasonic silent speech interface, namely, extraction feature of wavelet coefficient by principal component analysis(WaveletPCA), divided discrete cosine transform principal component analysis(block DCTG PCA) and block WalshHadamard transform principal component analysis (block WHTG PCA). The experiment

result shows that hybrid feature extraction method can better extract the important characteristics in ultrasound image of tongue.

It is not clear whether this pharyngeal contraction posture is also a kind of pronunciation function for languages with rhotic accent or rolling tongues through the analysis of language phenomena by using ultrasonic imaging technique. Boyce et al^[32] have the comparative study of rhotic accent or rolling tongues phenomena and describe the pharyngeal contraction. Bouavichith et al^[33] have a comparative study on the nasal sound produced by velar and palatal in Malayalam language.

5. Conclusion

We should carry out in-depth studies in language research and the traditional research technique and methods need to be updated constantly. The development of ultrasound imaging technology has promoted the in-depth study of language. Previously, some language phenomena can only be explained by theoretical speculation. Now, we can use ultrasound imaging technology for the explanation by using ultrasonic imaging technique and scientific and reasonable experimental data. The development of ultrasound imaging, image processing, three-dimensional modeling and other techniques are of great importance for the study of language practice.

6. Acknowledgements

This work was financially supported by National Natural Science Foundation of China (Grant No. 11564035).

7. References

- [1] C. A. Kelsey, F. D. Minifie and T. J. Hixon, "Applications of ultrasound in speech research," *J. Speech Hear. Res.*, 12, 564–575 (1969).
- [2] Shawker T H, Sonies B C. Ultrasound biofeedback for speech training. Instrumentation and preliminary results[J]. *Investigative Radiology*, 1985, 20(1): 90-93.
- [3] Davidson L. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance[J]. *The Journal of the Acoustical Society of America*, 2006, 120(1): 407-415.
- [4] Hamlet S L, Palmer J M. Investigation of laryngeal trills using the transmission of ultrasound through the larynx[J]. *Folia Phoniatrica et Logopaedica*, 1974, 26(5): 362-377.
- [5] Moisik S R, Esling J H. Evaluating the vowel space effects of larynx height using laryngeal ultrasound[J]. *Canadian Acoustics*, 2011, 39(3): 180-181.
- [6] Moisik S R, Esling J H, Bird S, et al. Evaluating laryngeal ultrasound to study larynx state and height[C]. *Proceedings of the 17th International Congress of Phonetic Sciences*. 2011: 136-139.
- [7] Wodzinski S M, Frisch S A. A preliminary ultrasound study of velar fronting[J]. *The Journal of the Acoustical Society of America*, 2003, 114(4): 2395-2395.
- [8] Wodzinski S M, Frisch S A. Ultrasound study of velar - vowel coarticulation[J]. *The Journal of the Acoustical Society of America*, 2006, 120(5): 3373-3374.
- [9] McAllister Byun T, Buchwald A, Mizoguchi A. Covert contrast in velar fronting: An acoustic and ultrasound study[J]. *Clinical linguistics & phonetics*, 2016, 30(3-5): 249-276.
- [10] Palo P, Schaeffler S, Scobbie J M. Pre-speech tongue movements recorded with ultrasound[C]. *Proceedings of the 10th International Seminar of Speech Production (ISSP 10)*. 2014: 300-303.
- [11] Mcmillan C T, Corley M. Cascading Influences on the Production of Speech: Evidence from Articulation[J]. *Cognition*, 2010, 117(3):243-260.
- [12] Drake, E., S. Schaeffler, and M. Corley. "ARTICULATORY EVIDENCE FOR THE INVOLVEMENT OF THE SPEECH PRODUCTION SYSTEM IN THE GENERATION OF PREDICTIONS DURING COMPREHENSION". In: *Architectures and Mechanisms for Language Processing (AMLaP)*. Marseille. (2013a).
- [13] Gick B , Bernhardt B , Bacsfalvi P , et al. Ultrasound imaging applications in second language acquisition[M]. *Phonology and Second Language Acquisition*. 2008.
- [14] Tsui M L . Ultrasound speech training for Japanese adults learning English as a second language[D]. 2012.
- [15] Suzuki, Kyori & Wilson, Ian & Watanabe, Hayato. Visual Learning 2: Pronunciation app using ultrasound, video, and MRI. *INTERSPEECH 2017: Show & Tell Contribution*. August 20–24, 2017.

- [16] Gick B. The use of ultrasound for linguistic phonetic fieldwork[J]. *Journal of the International Phonetic Association*, 2002, 32(2): 113-121.
- [17] Gick B, Bird S, Wilson I. Techniques for field application of lingual ultrasound imaging[J]. *Clinical Linguistics & Phonetics*, 2005, 19(6-7): 503-514.
- [18] Buchwald A, Rapp B, Stone M. Evidence for discrete phonological representations in production: Ultrasound imaging of aphasic speech[J]. *Brain and Language*, 2006, 99(1-2): 140-141.
- [19] Lu Z, Czap L. Modelling the tongue movement of Chinese Shaanxi Xi'an dialect speech[C]. 2018 19th International Carpathian Control Conference (ICCC). IEEE, 2018: 98-103.
- [20] Zharkova N. An ultrasound study of the development of lingual coarticulation during childhood[J]. *Phonetica*, 2018.
- [21] Krebs V L , Sedarous Y , Miller A L . Consonant-Vowel Coarticulation in velar plosives[C]. *Meetings on Acoustics. Acoustical Society of America*, 2013.
- [22] Bernhardt B, Gick B, Bacsfalvi P, et al. Ultrasound in speech therapy with adolescents and adults[J]. *Clinical Linguistics & Phonetics*, 2005, 19(6-7): 605-617.
- [23] Cavin M. The use of ultrasound biofeedback for improving English/r[J]. *Working Papers of the Linguistics Circle*, 2015, 25(1): 32-41.
- [24] Preston J L, Brick N, Landi N. Ultrasound biofeedback treatment for persisting childhood apraxia of speech[J]. *American Journal of Speech-Language Pathology*, 2013, 22(4): 627-643.
- [25] Jaumard-Hakoun A , Xu K , Cl  nence Leboulenger, et al. An Articulatory-Based Singing Voice Synthesis Using Tongue and Lips Imaging[C]. *Interspeech 2016*. 2016.
- [26] B. Denby, T. Schultz, K. Honda, Thomas Hueber, J.M. Gilbert, et al. *Silent Speech Interfaces*. Speech Communication, Elsevier: North-Holland, 2010, 52 (4), pp.270.
- [27] Hueber T, Chollet G, Denby B, et al. Continuous-speech phone recognition from ultrasound and optical images of the tongue and lips[C]. *INTERSPEECH 2007, Conference of the International Speech Communication Association*, Antwerp, Belgium, August. DBLP, 2007:658-661.
- [28] Thomas Hueber, Elie-Laurent Benaroya, G  rard Chollet, Bruce Denby, G  rard Dreyfus, et al. Development of a Silent Speech Interface Driven by Ultrasound and Optical Images of the Tongue and Lips. *Speech Communication*, Elsevier: North-Holland, 2010, 52 (4), pp.288.
- [29] Wang J. Integrating Dynamic Information with Multi-stream HMM in Ultrasound-based Silent Speech Interface[J]. *Journal of Information & Computational Science*, 2015, 12(13):4875-4883.
- [30] Csap  t G, Gr  sz T, Gosztolya G, et al. DNN-based Ultrasound-to-Speech Conversion for a Silent Speech Interface[C]. *INTERSPEECH*. 2017:3672-3676.
- [31] LU Wenhuan, QU Yuexin, YANG Yalong, et al. Hybrid feature extraction from ultrasound images for a silent speech interface[J]. *J Tsinghua Univ(Sci&Technol)*, 2017, Vol.57, No.11:1159-1162.
- [32] Boyce S , Hamilton S M , Campos A R , et al. Articulatory similarity in rhotic sounds: A cross-linguistic comparison[J]. *Journal of the Acoustical Society of America*, 2015, 137(4):2382-2382.
- [33] Bouavichith D A , Namboodiripad S , Garellek M . A contrastive place of articulation between palatals and velars: An ultrasound study of Malayalam palatal-velar nasals[J]. *Journal of the Acoustical Society of America*, 2018, 143(3):1755-1755.