

Clinical Feedback and Technology Selection of Game Based Dysphonic Rehabilitation Tool

Zhihan Lv

FIVAN, Valencia, Spain
Email: lvzhihan@gmail.com

Chantal Esteve

FIVAN, Valencia, Spain
Email: chantal@fivan.org

Javier Chirivella

FIVAN, Valencia, Spain

Pablo Gagliardo

FIVAN, Valencia, Spain
Email: pablog@fivan.org

Abstract—An assistive training tool software for rehabilitation of dysphonic patients is evaluated according to the practical clinical feedback from the treatments. One stroke sufferer and one parkinson sufferer have provided earnest suggestions for the improvement of our tool software. Seven pitch estimation algorithms have been evaluated and compared with selected patients voice database. A series of benchmarks have been generated during the evaluation process for technology selection.

Keywords—*dysphonic, gamification, voice game, pitch detection*

I. INTRODUCTION

Using video game as the vision feedback has been already demonstrated as a useful HCI approach for rehabilitation assistive method [10]. Some previous research have proved the usability of the game in speech therapy [30], for children with autism [5], and parkinsons disease patients [6]. However, none of these researches has considered the success rate of recognition of patients' voice by different pitch estimation algorithms from the view of computer technology.

In this research, we evaluated seven pitch estimation algorithms using our patient voice database, and then gave a preliminary benchmark of the performance of pitch recognition of patient voice. Video game is employed as vision feedback tool to evaluate the efficacy of the patient voice exercise, which can be considered as the long-time continued evaluation of the severity of dysphonia. Pitch is detected as the evaluation factor which is also the controllable input of the games. We define our research as a biometric information management system design, in which the biometric refer to voice pitch specifically. In our previous research [13], we mentioned that our design science based software development just go forward a step, and thus 'the next step is to iteratively improve it and to develop experiments to test it before eventually moving on to conducting clinical tests as the assistive tool for rehabilitation of dysphonic patients'. Therefore, we recorded and assessed the feedback from the treatments for improving the assistive tool. In other side, we supposed the therapist could supervise and train the patient side by side since the microphone arrays of kinect is enabled to filter ambient noise. In this paper, we discussed this idea according to the practical clinical application as well as the direct feedback from the treatment.

II. CLINICAL FEEDBACK FROM TREATMENT

Previously, we have discussed with a rich-experience therapist regarding general exercise process in order to identify and define the problem pragmatically. The first version of the



Fig. 1. New user interface of the proposed tool which is suitable for tablet with normal microphone.

assistive tool is designed according to the therapist's suggestions and considering 'Intuitive', 'Entertainment', 'Incentive' as main design factors. The latest version of the software proposed in this paper supports common mic and gets rid of Kinect2. Because we have realized that the treatments are like the driving training, the independent driving isn't expected at the early training usually. Especially when the game is out of control but the patient cannot increase the pitch at all, the therapist should rapidly make voice and help the patient to complete the game, which gives patient a paradigm and can reduce anxiety and boost confidence of the patients. The new user interface is shown as in figure 1. The left side of the UI is the game logic, the right side is the estimated pitch values monitor which presents frequency, pitch, amplitude, midi note, midi note number, sample, duration.

The tool is evaluated in the clinical treatment for three patients. The first patient is a stroke sufferer who always sits inside the wheelchair. The patient cannot speak, but he can still make voice. After exercise four times, the patient control the avatar in the game successfully. The therapist suggested that the distance between the obstacles in the game should be shorter so that the patient could control the avatar to easily evade the obstacles as they want. During the game, we can see the patient struggle to talk and ask the questions about the games, but he cannot make voice clearly. The stroke patient was evaluated again at the next day. After we modified the density of the planets and the critical value of pitch, the patient can play with the game well, since the game it's much easier.

III. ALGORITHM COMPARISON

The second patient is a parkinson sufferer who can sit and stand independently. The therapists played the game to instruct the game rule at the beginning and then the patient tried to play with the game. After the patient repeated the game four times, the patient complete the game successfully. The patient's voice is not clear, so the pitch estimation algorithm cannot detect it. A more sensitive algorithm which is suitable for patient's voice is needed. The critical values of the player pitch for patients are usually vastly different from the normal people. We set the critical value as half as that for the normal people at the beginning, but it's still impossible for patients to reach. We modified the setting iteratively. It doesn't work until the value of the setting is eight times easier than normal people. The patient is unable to change the pitch, so we try to use extreme low pitch ($50Mel$) as the critical pitch.

The therapist wish a bigger game logic user interface (UI), because the patient's vision is not enough good to see the whole scene, another reason is that the movable space for the avatar is too limited. The current avatar in the game is also too small to manipulate. Similarly, the patient thinks the obstacles are also too small, he cannot see them very clearly. We have observed that the patient struggled to try to play with the game UI by touch during the game, which is because of that the current interaction approach is not idiomatic.

The most intolerable problem is that the algorithm is still not sensitive enough for patients voice. Sometimes, the detection of some pitches are missing. During the game, we noticed that different voice or different timbre got different success rate of detection. For example, 'Wo' was recognized with higher success rate than 'Aa'. The detection of 'Yi' had a very good performance. By making voice of 'Yi', the patient passed the easiest level of the game. The detection of 'Ai' didn't work in our game.

After twice treatment for each patient, the therapist start to supervise the patients to play with the game without our attending. After three times treatments in one week, the patients can control the game by their pitch very well, even they can make the voice of 'Ai' effortlessly. Because they have understood the game rule now and distinguished the pitch and loudness. Previously, the patients was struggling to make voice but actually they only increased the loudness which was not the controller factor of the designed game. Now the patients don't put forth their strength to make voice any more, instead, they let their voice be slightly and increase the pitch skillfully, as the therapist teach them, as well as the experience that they realized from the game. The easiest levels of the games are already suitable for both patients. We are going to improve both UI and pitch estimation algorithm of the software as the therapist suggested so that it can be suitable for most of the dysphonic patients.

The situation of the third patient is better than the previous two patients. The therapist tried to treat this patient using senior level of the designed game. However, the result indicated the senior level was still too difficult for the patient even if he has better situation. The main reason is the lack of the sensitivity of the pitch estimation algorithm. The minor reason is that the great disparity between normal people and dysphonic patients.

As we have mentioned, in our previous work [13] the frequency of the patient voice is identified by FFT (Fast Fourier Transform) based frequency-domain approach which turns the voice wave into a frequency distribution. Further more, the pitch (tone) and midi number are calculated depend on frequency. The therapist evaluated the pitch estimation approach with piano and proved the reliability and accuracy of it. We have considered the voice characteristics like non-continuity and low-loudness of dysphonic patients, so we concluded that the sensitivity and suitability of the pitch extracting method are the decisive factors beside the accuracy rate. In the clinical feedback section, we have introduced that the current FFT based pitch estimation algorithm can only detect some certain pitches. Therefore, there is a practical need that creating or finding a stable and sensitive enough pitch estimation algorithm for the special voice of patients in different conditions.

In this paper, we will evaluate seven pitch estimation algorithms, which are also known as pitch detection or pitch recognize algorithms. The seven pitch estimation algorithms are as followed: AdvancedAutoCorrelator [1], Dynamic Wavelet Algorithm [7] which is implemented at [27], FastYin [24] which is a faster implementation of Yin [4], MPM [25] by Philip McLeod and Geoff Wyvill, Yin [4] which is implemented at [2], FFT [26] which is implemented at [35], classical AutoCorrelator. The implementation is based on C# with WPF, the algorithms refer to the source code of TarsosDSP [28].

We filtered the pitch above $400Mel$ and only keep the pitch below $400Mel$ which is the limit of patients voice, because sometimes super high pitches ($> 2000Mel$) were detected.

In the first step, we compared the performance and time-consuming of seven algorithms with different buffer size (1024, 2048, 4096, 8192, 16384) using sine wave as input. The sine wave is a mathematical curve that describes a smooth repetitive oscillation, also on behalf of the analog signal. Figure 2 left indicates that with the amplification of the buffer size, the curve tends to smooth.

For example, when buffer size is 1024, Dynamic Wavelet, Yin and McLeod have error from midi note number of 35 to 40. FastYin has error from 35 to 45. FFT and AdvanceAutocorrelator have regular errors during the whole range. Autocorrelator has irregular error during the whole rage.

When buffer size is 4096, AdvanceAutocorrelator, Dynamic Wavelet, Yin have not any error in the whole range. FastYin and McLeod have errors at 40, 55 and 35 to 40 respectively. Fft and Autocorrelator have errors at 35 to 60 and 36 to 60 respectively.

When buffer size is 16384, only Fft and Autocorrelator have serious errors after 60. McLeod has tiny error before 40. Besides, all the other algorithms have normal curve.

Figure 2 right indicates that the time consuming of McLeod increases immensely with the increase of buffer size, while the time consuming of other algorithms aren't changed significantly.

In the second step, the algorithms are compared using patient voice database. Figure 3 shows the performance com-

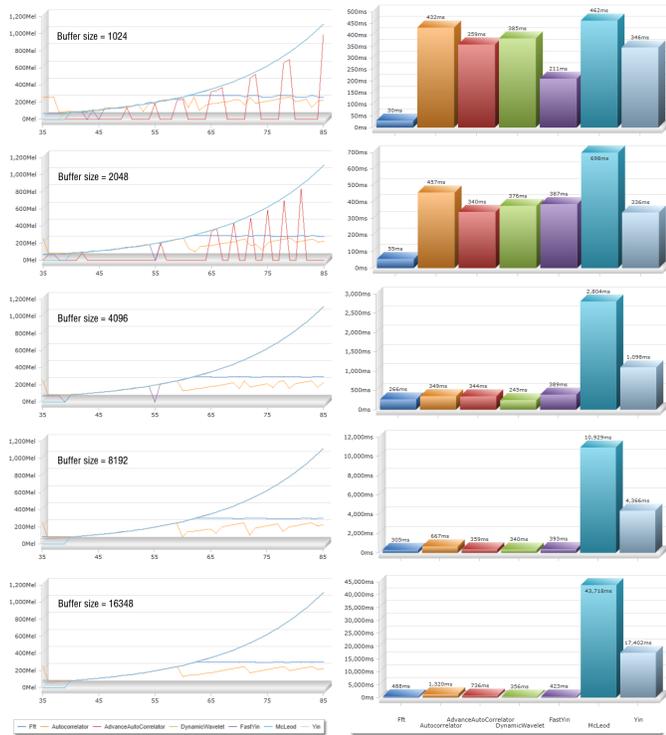


Fig. 2. The performance comparison of the seven algorithms with different buffer size (1024, 2048, 4096, 8192, 16348). Left: Pitch estimation from sine wave, the x-axis represents midi note number; Right: Time consuming of algorithms.

parison of the algorithms with 4096 buffer size using parkinson patient voice record. This voice section recorded a few of voice of 'Wo'. Mcleod cannot correctly estimate the pitch of patient voice when the buffer size is more than 1024 so we won't add it in this figure. As shown in Figure 3, all the estimation results of Autocorrelator and Fft are non-zero values, which means successful estimated pitch. Yin and FastYin have only one estimated non-zero pitch value, which is too limited for pitch controllable game. AdvanceAutocorrelator and Dynamic Wavelet get median intensive pitches between the two kinds. The comparison results indicates that Autocorrelator algorithm gets the most intensive pitches which means it is the most sensitive pitch estimation algorithm for patient voice. No zero value pitches estimated also reveals that Autocorrelator and Fft algorithms are suitable for patients' voice with non-continuity characteristic.

The flaws of Autocorrelator and Fft are that the pitch estimation errors happen when the midi note numbers are more than 60, which means the two algorithms are not suitable for high-pitched voice. Fortunately, the limit of patient voice pitch is 60 as we mentioned in our previous work [13], so the limit of the two algorithms are in the scope of the patients' voice. Therefore, Autocorrelator is the most suitable candidate algorithm for patient voice pitch estimation.

Figure 4 presents a pitch estimation of a piece of voice section which records a voice of 'Aa' by the same parkinson patient. This voice section recorded a voice of 'Aa'. The pitch estimated by seven algorithms with 4096 buffer size. Yin and FastYin cannot estimate any pitch from this piece of voice. The

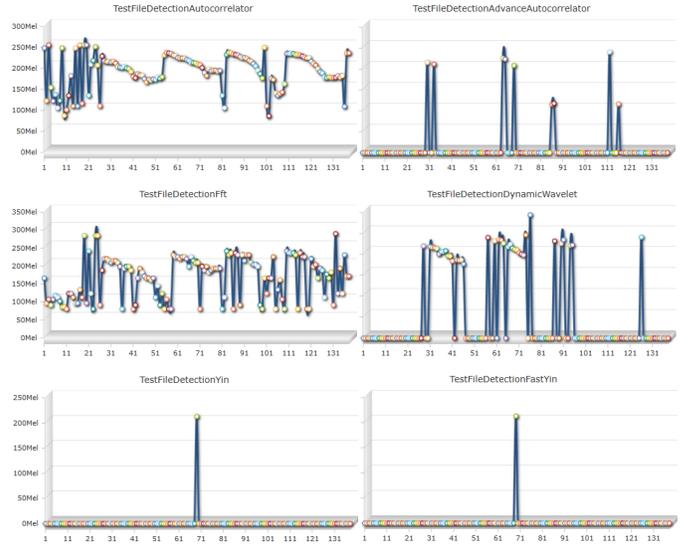


Fig. 3. The performance comparison of the seven algorithms with 4096 buffer size using parkinson patient voice record, the x-axis represents midi note number. This voice section recorded a few of voice of 'Wo'.

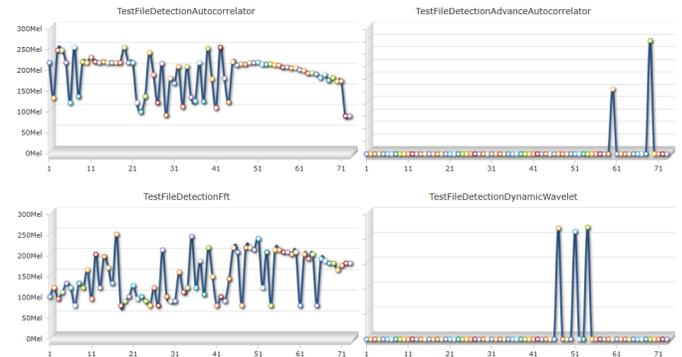


Fig. 4. The performance comparison of the seven algorithms with 4096 buffer size using parkinson patient voice record, the x-axis represents midi note number. This voice section recorded a voice of 'Aa'.

observation result of the graph also proves that Autocorrelator is the most sensitive algorithm and suitable for patients' voice with non-continuity characteristic.

IV. CONCLUSION

The contributions of this paper include two points. The first one is the evaluation of the designed assistive tool for dysphonic rehabilitation. One stroke sufferer and one parkinson sufferer have brought us their practical suggestions according to their experiences to improve the software in the treatments. The second point is the most suitable pitch estimation algorithm selection. In order to select the best algorithm, seven algorithms have been evaluated and compared with selected patient voice database. Even through the comparison results indicate that Autocorrelator is the most sensitive algorithm and suitable for patients voice with non-continuity characteristic, a custom pitch estimation algorithm which can refine patient voice is still expected.

Virtual reality has been widely applied in many fields range

from everyday entertainment and communication [12] [36] to traditional research fields such as geography [23] [31], ocean [29] [21], city [20] [9] [8] and biology [32] [22], clinical assist [13] [14] [19]. Some novel interaction approaches [11] [16] [15] [17] [18], the new network data management algorithm [33] [34], smart grid system [3] [37], neural network algorithm [39] [38] are considered to be integrated in our future work.

ACKNOWLEDGMENT

The authors would like to thank Sonia Blasco, Vicente Penades and Oktawia Karkoszka. The work is supported by LanPercept, a Marie Curie ITN funded through the 7th EU Framework Programme(316748).

REFERENCES

- [1] BitFlipper. Realtime C# Pitch Tracker. <https://pitchtracker.codeplex.com>, 2011.
- [2] P. Brossier. aubio. <http://aubio.org/>, 2015.
- [3] L. Che and M. Shahidehpour. Dc microgrids: Economic operation and enhancement of resilience by hierarchical control. *IEEE Transactions Smart Grid*, 2014.
- [4] A. De Cheveigné and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.
- [5] M. E. Hoque, J. K. Lane, R. E. Kaliouby, M. S. Goodwin, and R. W. Picard. Exploring speech therapy games with children on the autism spectrum. In *INTERSPEECH*, pages 1455–1458. ISCA, 2009.
- [6] M. Krause, J. Smeddick, and R. Meyer. A digital game to support voice treatment for parkinson’s disease. In *CHI ’13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA ’13, pages 445–450, New York, NY, USA, 2013. ACM.
- [7] E. Larson and R. Maddox. Real-time time-domain pitch tracking using wavelets. *Proceedings of the University of Illinois at Urbana Champaign Research Experience for Undergraduates Program*, 2005.
- [8] X. Li, Z. Lv, J. Hu, B. Zhang, L. Yin, C. Zhong, W. Wang, and S. Feng. Traffic management and forecasting system based on 3d gis. In *Cluster, Cloud and Grid Computing (CCGrid), 2015 15th IEEE/ACM International Conference on*. IEEE, 2015.
- [9] X. Li, Z. Lv, W. Wang, C. Wu, and J. Hu. Virtual reality gis and cloud service based traffic analysis platform. In *The 23rd International Conference on Geoinformatics (Geoinformatics2015)*. IEEE, 2015.
- [10] M. A. Liebert and G. R. P. D. Virtual reality in psychotherapy: Review, June 2005.
- [11] Z. Lv. Wearable smartphone: Wearable hybrid framework for hand and foot gesture interaction on smartphone. In *2013 IEEE International Conference on Computer Vision Workshops*. IEEE, 2013.
- [12] Z. Lv, G. Chen, C. Zhong, Y. Han, and Y. Y. Qi. A framework for multi-dimensional webgis based interactive online virtual community. *Advanced Science Letters*, 7(1):215–219, 2012.
- [13] Z. Lv, C. Esteve, J. Chirivella, and P. Gagliardo. A game based assistive tool for rehabilitation of dysphonic patients. In *Virtual and Augmented Assistive Technology (VAAT), 2015 3rd Workshop on*. IEEE, 2015.
- [14] Z. Lv, C. Esteve, J. Chirivella, and P. Gagliardo. Serious game based dysphonic rehabilitation tool. In *2015 International Conference on Virtual Rehabilitation (ICVR)*. IEEE, 2015.
- [15] Z. Lv, L. Feng, S. Feng, and H. Li. Extending touch-less interaction on vision based wearable device. In *Virtual Reality (VR), 2015 IEEE*. IEEE, 2015.
- [16] Z. Lv, L. Feng, H. Li, and S. Feng. Hand-free motion interaction on google glass. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications*. ACM, 2014.
- [17] Z. Lv, A. Halawani, S. Feng, H. Li, and S. Ur Réhman. Multimodal hand and foot gesture interaction for handheld devices. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 11(1s):1–19, 2014.
- [18] Z. Lv, A. Halawani, S. Feng, S. ur Rhman, and H. Li. Touch-less interactive augmented reality game on vision-based wearable device. *Personal and Ubiquitous Computing*, pages 1–17, 2015.
- [19] Z. Lv and H. Li. Imagining in-air interaction for hemiplegia suffererl. In *2015 International Conference on Virtual Rehabilitation (ICVR)*. IEEE, 2015.
- [20] Z. Lv, X. Li, B. Zhang, W. Wang, S. Feng, and J. Hu. Big city 3d visual analysis. In *Eurographics 2015*, 2015.
- [21] Z. Lv and T. Su. 3d seabed modeling and visualization on ubiquitous context. In *SIGGRAPH Asia 2014 Posters*, page 33. ACM, 2014.
- [22] Z. Lv, A. Tek, F. Da Silva, C. Empereur-mot, M. Chavent, and M. Baaden. Game on, science-how video game technology may help biologists tackle visualization challenges. *PLoS one*, 8(3):e57990, 2013.
- [23] Z. Lv, T. Yin, Y. Han, Y. Chen, and G. Chen. Webvr–web virtual reality engine based on p2p network. *Journal of Networks*, 6(7), 2011.
- [24] M. Mauch. Matthias Software:FastYin. <http://schall-und-mauch.de/artificialmusicality/software/>, 2014.
- [25] P. McLeod and G. Wyvill. A smarter way to find pitch. In *Proceedings of International Computer Music Conference, ICMC*, 2005.
- [26] H. H. Monson. Statistical digital signal processing and modeling, 1996.
- [27] A. Schmitt. Dynamic Wavelet Algorithm Pitch Tracking library. <http://www.schmittmachine.com/dywapitchtrack.html>, 2010.
- [28] J. Six, O. Cornelis, and M. Leman. TarsosDSP, a Real-Time Audio Processing Framework in Java. In *Proceedings of the 53rd AES Conference (AES 53rd)*, 2014.
- [29] T. Su, Z. Lv, S. Gao, X. Li, and H. Lv. 3d seabed: 3d modeling and visualization platform for the seabed. In *Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014.
- [30] C. T. Tan, A. Johnston, A. Bluff, S. Ferguson, and K. J. Ballard. Retrogaming as visual feedback for speech therapy. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications*, SA ’14, pages 4:1–4:5, New York, NY, USA, 2014. ACM.
- [31] J. Tan, X. Fan, and Y. Ren. Methodology for geographical data evolution: three-dimensional particle-based real-time snow simulation with remote-sensing data. *Journal of Applied Remote Sensing*, 8(1):084598–084598, 2014.
- [32] A. Tek, B. Laurent, M. Piuze, Z. Lu, M. Chavent, M. Baaden, O. Delalande, C. Martin, L. Piccinali, B. Katz, et al. Advances in human-protein interaction-interactive and immersive molecular simulations. *Biochemistry, Genetics and Molecular Biology”Protein-Protein Interactions-Computational and Experimental Tools”*, pages 27–65, 2012.
- [33] Y. Wang, W. Jiang, and G. Agrawal. Scimate: A novel mapreduce-like framework for multiple scientific data formats. In *CCGrid2012*. IEEE, 2012.
- [34] Y. Wang, A. Nandi, and G. Agrawal. Saga: Array storage as a db with support for structural aggregations. In *SSDBM ’14*, New York, NY, USA, 2014. ACM.
- [35] P. Wendykier. JTransforms. <https://sites.google.com/site/piotrwendykier/software/jtransforms>, 2014.
- [36] M. Zhang, Z. Lv, X. Zhang, G. Chen, and K. Zhang. Research and application of the 3d virtual community based on webvr and ria. *Computer and Information Science*, 2(1):P84, 2009.
- [37] M. Zhang, Y. Sun, S. Dang, and K. Petrou. Smart grid-oriented algorithm of data retrieval and processing based on crio. In *ISEEE2014*. IEEE, 2014.
- [38] X. Zhang, G. Foderaro, C. Henriquez, A. VanDongen, and S. Ferrari. A radial basis function spike model for indirect learning via integrate-and-fire sampling and reconstruction techniques. *Advances in Artificial Neural Systems*, 2012:10, 2012.
- [39] X. Zhang, Z. Xu, C. Henriquez, and S. Ferrari. Spike-based indirect training of a spiking neural network-controlled virtual insect. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 6798–6805. IEEE, 2013.