2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE) Mexico City, Mexico. September 11-13, 2019

Analysis of Audio Vocalizations in the Context of the Teaching and Learning of Singing

Luis Alberto Martinez
Electrical Engineering Department
Center for Research and Advanced Studies of the National Polytechnic Institute
Mexico City, Mexico
lamr@cinvestav.mx

Gisela Gracida
Electrical Engineering Department
Center for Research and Advanced Studies of the National Polytechnic Institute
Mexico City, Mexico
gisgracida@msn.com

Rafael Ángel Urrutia
Electrical Engineering Department
Center for Research and Advanced Studies of the National Polytechnic Institute
Mexico City, Mexico
raurza@live.com

Eladio Cardiel
Electrical Engineering Department
Center for Research and Advanced Studies of the National Polytechnic Institute
Mexico City, Mexico
ecardiel@cinvestav.mx

Manuel Mauricio Lara
Electrical Engineering Department
Center for Research and Advanced Studies of the National Polytechnic Institute
Mexico City, Mexico
mlara@cinvestav.mx

Pablo Rogelio Hernández
Electrical Engineering Department
Center for Research and Advanced Studies of the National Polytechnic Institute
Mexico City, Mexico
pablo.rogeli@cinvestav.mx

Abstract- **Since antiquity, the teaching process of singing has been by imitation. In this field , the work on the appreciation of the voice has been sustained mainly in auditory perception. Nowadays, the development of new technologies has supported the teaching-learning process of singing, through visual observation and auditory perception of the digital audio records. In this case, auditory and visual perceptual systems has been used as feedback to reinforce the learning process when vocalizations, from vowels to opera segments, are analyzed. In this sense, this project proposes the objective analysis of the audio signal of a set of vocalizations using the fundamental frequency as the essential feedback parameter. The vocalization of the vowel /a/ with five trials executed for four seconds was used. During the phonation of vowels, 3 zones were identified for the analysis: an initial(1z) , middle(2z), and the final(3z) of the vocalization recording. The middle zone had oscillations that correspond to the typical vibrato of the human voice, while unexpected oscillations occured in the other zones.**

Keywords—singing voice, visual feedback, multidisciplinarity, audio, digital signal processing

I. INTRODUCTION

Learning to sing is a complicated process [1] [2], which usually uses the teacher-apprentice approach, where the instructor sets practical examples so that the student imitates them in order to receive a feedback, indicating the quality with which the example was reproduced. Usually, this feedback is verbal [3], it uses modeling, and it is the primary way for the student of getting better. This "oral-auditive" process depends directly on the singing capacities of the teacher, and the auditive sharpness of himself and of his pupil [4].

The singing learning process is surrounded by feedback signals (auditory, visual, kinestesia, etc.), and the human brain processes these signals naturally. It is important to note that traditional teaching-learning procedure requires and uses consistently auditory feedback to the extent that the teacher and the student are heard each other frequently and mutually. The human ear functions as a differentiator of frequencies, and a musically trained ear can discern intensities, tones, rhythms, and sequences [16].

The technological advances that allow the development of monitoring and analysis strategies of vocal phenomenon, both spoken and sung, are very broad and generous [5] [6] [7] [8]. In the same way, musical and physical parameters associated with this human activity are also varied [9] [10] [11] [12]. The speech processing field, whose development is well consolidated, offers a set of knowledge that allows approximated observations to elementary parameters in the field of singing processing such as the musical intonation [8] [13]. A strategy that has been used in the process of studying singing voice is that of using consolidated commercial tools or free software to discover vocal phenomena that clarify its operation processes and thus generate technological innovations based on these findings [13] [14] [15].

There are fields of study currently very active, which are developed by research communities with whom we share some tasks to be developed, but with different objectives. For example, the development of user interface for data interaction

978-1-7281-4840-3/19/$31.00 ©2019 IEEE

with audio files [17], details of how to link information of different formats such as audio and image [18], and music information retrieval tasks such as detection of singing activity, melody estimation, musical genre classification, and intonation estimation [8].

Next, a methodology for recording and the basis for the audio analysis of vocalizations are described. This enable to do visual and auditory observations showing the development of a strategy of visual and auditory feedback so as to offer measurements that contribute to the process of teaching and learning of singing.

II. Methodology

A. *Participants, vocalizations, and indications about the vocal exercises*

There were three participants who generated the set of vocalizations. A woman (singer) and two men with differences in vocal training. These participants will be referred as woman, man 1, and man 2 in this document. They were asked to generate an /a/ vowel in their natural voice frequency, with a duration of 4 seconds to promote the accommodation and learning of the phonation apparatus. The vocalization was repeated five times consecutively. Each of them choose the fundamental frequency they feel comfortable with. The distance between the microphone and the person were 20 cm. The participants voiced in a confortable standing position.

B. *Recording system and environment*

A Sennheiser microphone (mic) e620 model was used, which is a permanent polarization condenser microphone with a transmission range of 40 Hz to 20 kHz with flat frequency response under specific conditions of distance between sound source and microphone. The mic signal is digitized with an audio interface Behringer UMC202HD which has a response in frequency from 10 Hz to 50 kHz; the interface was set to apply a 44100 Hz sampling rate. Recordings were made in a low-noise (less than 20dB) room. Finally, the recording process was made through the Audacity software installed on a laptop with OS Windows 10, Core i7 @2. 50 GHz and 8 GB RAM.

C. *Process of measuring and plotting the fundamental frequency of vocalizations.*

The three audio signals of the vowel /a/ were generated with 5 trials, they conditioned with Audacity software version 2.1.1 resulting in 15 audio recordings. The procedure to identify the recorded sound of the /a/ vowel consisted on a visual review through graphical observation and an auditory analysis with the audio playback. In this way, the times in which the vowel is present are defined, and then we proceed to arbitrarily trim the segment 100ms before and after each /a/.

For the measurement of the fundamental frequency (F0) the Praat software version 6.0.46 was used. This software was set to use the autocorrelation analysis method, considering 100 Hz to 280 Hz measurement limits and a maximum audio duration of 25s. The time-F0 values were then saved in a text file, which offers the list contained in the file with the extension PitchTier.

Finally, Praat text files of the vocalizations were exported to Matlab environment to be plotted and analyzed, using a set of Matlab functions reported by [19].

III. Results and Discussion

The figure 1 exhibits the graphics of the 15 vocalizations performed by the three participants. Those graphics of the five /a/ vowels vocalizations are ordered to the time of onset according to the colors black, red, green, blue and purple, the black segment being the first /a/ vowel up to the purple segment for the fifth /a/.

It is observed in Fig. 1 differences in F0 between of the vocalization of /a/ vowel of the three participants. Also, differences in time duration between participants are noticed. In general, it is also possible to corroborate graphically that every participant was able to repeat their /a/vowel executions.

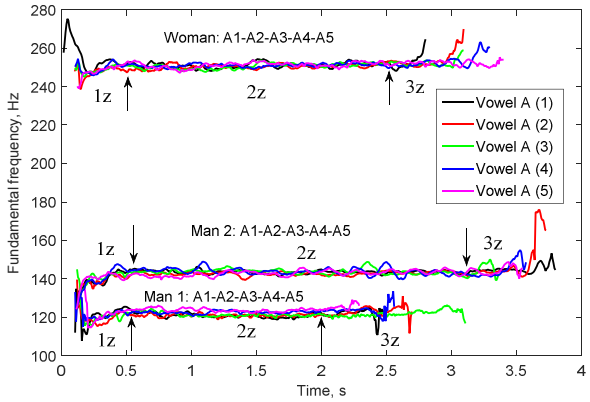

Fig. 1. All five /a/ vowel vocalizations (A1-A2-A3-A4-A5), produced by three participants, are shown. 1z, 2z and 3z are the identified zones in each vocalization from all participants. Each arrow indicate the boundary time point between zones. F0 and duration differences between participants are evident.

In the field of music, it is known that measurements of duration of musical notes, gaps or silences, and musical intervals are the basis for understanding the rhythm concept.

On the other hand, in Fig. 1 it can be identified that the three participants show three regions on each produced vowel, which may be named as the beginning of vowel, stable vowel and end of vowel. Noting this same figure and supported with the images of the Fig. 2 to Fig. 6 it can be seen a tendency to relapse in the mechanism that creates this notable change between the region of stable vowel and the regions of the beginning and end of the vowel. The woman has a beginning of vowel which tends to a F0-peak followed by a minimum F0-peak to consequently being directed towards the area of stable vowel. The end of the vowel in woman trial tends to increase its F0. Meanwhile, man 1 has an onset of vowel with a decline of F0 and an end of vowel with non-uniform oscillations. Finally, man 2 has an onset of vowel with a rise of the F0 and a trend towards less noticeable rise compared with the woman in the end region of the A.

In Fig. 2 to Fig 6 it can be shown each of the A vowels, separetaly and it can be visually compared differences among productions made by each participant according to the vowel number.

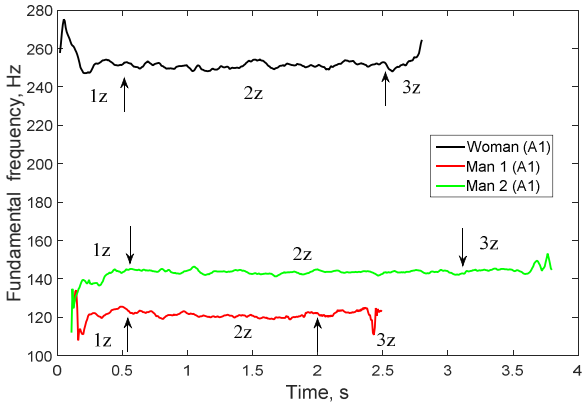

Fig. 2. F0 of the three participants when first A is produced. 1z, 2z and 3z are the identified zones in each vocalization from all participants. Each arrow indicate the boundary time point between zones.

On the one hand, if the stable region of all the A repetitions and the three participants are observed, it is possible to identify that man 2 presents more pronounced swings as it increases the number of repetitions of the vowel A. These oscillations, which could be identified as a certain expression of a *vibrato*, are more visible in man 2 than in the other 2 participants and less visible in man 1 than in the other 2 participants. Related to the oscillations of woman vocalization in the middle region of the vowel, it is distinguished that these F0 oscillations were of ± 3.1921 Hz.

Finally, man 1 shows less extensive and less repetitive fluctuations of ± 2.0634 Hz in F0 along the time of the three participants considering the stable region of the repetitions of the vowel A. In addition, also in the stable region, graphs show a tendency of ascending and descending in various areas of the stable region (Fig. 2, Fig. 4, Fig. 5 and Fig. 6) which suggests that man 1 is the participant with less vocal training, which coincides with the facts.

It is important to notice that although all three participant were asked to produce the vowel /a/ lasting 4 second, Man 2 was the closest with 3.69 s in A1 and Man 1 was the smallest with 2.115 s in A5. This suggests duration as a parameter to characterize participant singing skills (TABLE I).

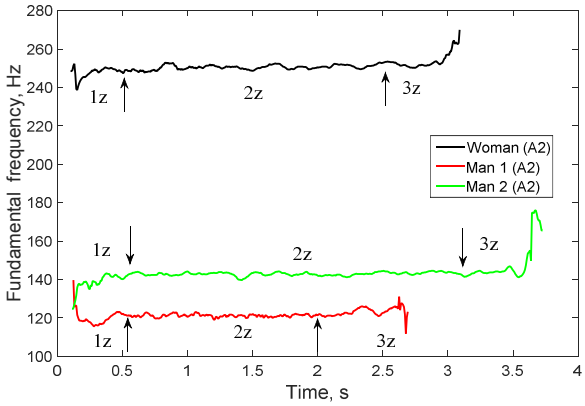

Fig. 3. F0 from vocalizations of the three participants when second /A/ vowel is produced. 1z, 2z and 3z are the identified zones in each vocalization from all participants. Each arrow indicate the boundary time point between zones.

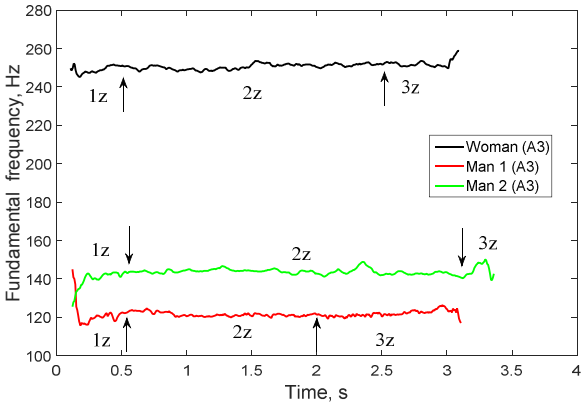

Fig. 4. F0 of three participants when third A is produced. produced. 1z, 2z and 3z are the identified zones in each vocalization from all participants. Each arrow indicate the boundary time point between zones.

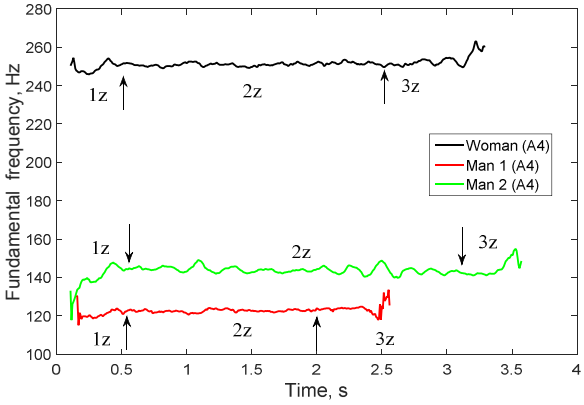

Fig. 5. F0 of three participants when fourth A is produced. produced. 1z, 2z and 3z are the identified zones in each vocalization from all participants. Each arrow indicate the boundary time point between zones.

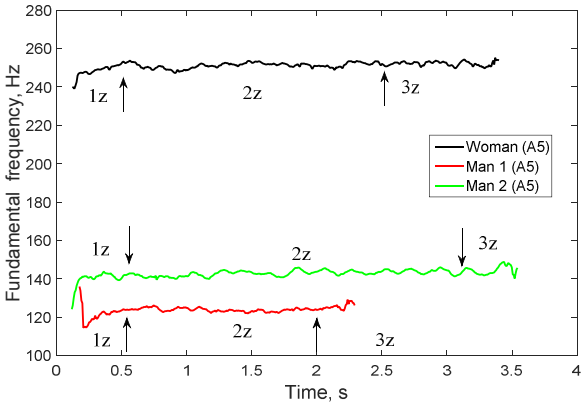

Fig. 6. F0 of three participants when fifth A is produced. 1z, 2z and 3z are the identified zones in each vocalization from all participants. Each arrow indicate the boundary time point between zones.

TABLE I. DURATIONS OF /A/-VOWEL EXECUTION

Participants	Duration (s)				
	A1	*A2*	*A3*	*A4*	*A5*
Woman	2.7828	2.9922	2.9853	3.1872	3.2777
Man 1	2.3628	2.5725	2.9853	2.408	2.115
Man 2	3.6903	3.6152	3.24	3.465	3.427

According to the results, and as it was mentioned before, both the auditory and visual perceptual systems could be used as feedback to reinforce the teaching-learning process of singing, respect to those that use one perceptual system only.

IV. CONCLUSIONS

The addition of a mechanism of visual feedback to reaffirm or correct the way of producing a vocalization provides clear information on aspects of the vocal phenomenon that the auditory perception of the individual does not retain easily. This visual feedback in this paper can be analyzed together with a process of identifying and interpreting the audio files recorded that represent the vocalizations produced. As a result, this suggests that the contribution of the auditory information plus visual information results in an observation that it is greater than the sum of its parts.

ACKNOWLEDGMENT

Authors deeply appreciate the collaboration of the "Conservatorio Nacional de Música de México" for the valuable advicing and didactic materials. Likewise to the Universidad Autónoma de Nayarit and Conacyt for the support given to Luis Alberto Martínez Rodríguez for the postgraduate studies.

REFERENCES

[1] A. Hong-Young, *Singing professionally: studying singing for actors and singers*. Heinemann Drama, 1995.

[2] C. Caballero, *Cómo educar a voz hablada y cantada*, 12a. Ed. D.F.: Alfa Futuro S.A. de C.V., 2012.

[3] J. Callaghan, *Singing and voice science*, 1st ed. Singular, 1999.

[4] P. Wilson, "Does real - time visual feedback improve pitch accuracy in singing?," The University of Sydney, 2006.

[5] L. Nijs and M. Leman, "Interactive technologies in the instrumental music classroom: A longitudinal study with the Music Paint Machine," *Comput. Educ.*, vol. 73, pp. 40–59, 2014.

[6] M. Taenzer and M. Stefan, "Analysis and Visualisation of Music," in *ICEIC, International Conference on Electronics, Information and Communication*, 2019.

[7] J. Gerhard and D. E. Rosow, "A Survey of Equipment in the Singing Voice Studio and Its Perceived Effectiveness by Vocologists and Student Singers," *J. Voice*, vol. 30, no. 3, pp. 334–339, 2016.

[8] E. J. Humphrey *et al.*, "An Introduction to Signal Processing for Singing-Voice Analysis," *IEEE Signal Process. Mag.*, vol. 36, no. 1, 2019.

[9] G. Gracida, "Caracterización acústica y perceptual de la expresividad vocal en el canto operístico," Universidad Nacional Autónoma de México, 2016.

[10] G. Gracida, "Programa Interactivo Para Analizar La Voz Cantada Mediante Técnicas De Procesamiento Digital De Señales," 2010.

[11] E. Rapoport, "Emotional expression code in opera and lied singing," *J. New Music Res.*, vol. 25, no. 2, pp. 109–149, 1996.

[12] R. Timmers, "Vocal expression in recorded performances of Schubert songs," *Music. Sci.*, vol. XI, no. 2, pp. 73–101, 2007.

[13] O. Babacan, T. Drugman, N. Alessandro, N. Henrich, and T. Dutoit, "A comparative study of pitch extraction algorithms on a large variety of singing sounds (Circuit Theory and Signal Processing Laboratory , University of Mons , Belgium Speech and Cognition Department , GIPSA-lab , Grenoble , France," 2013, pp. 7815–7819.

[14] B. McFee, J. W. Kim, M. Cartwright, J. Salamon, R. M. Bittner, and J. P. Bello, "Open-Source Practices for Music Signal Processing Research: Recommendations for Transparent, Sustainable, and Reproducible Audio Research," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 128–137, 2019.

[15] V. M. Ramesh and S. H. V, "Exploring Data Analysis in Music Using Tool Praat," *2008 First International Conference on Emerging Trends in Engineering and Technology*. pp. 508–509, 2008.

[16] M. Müller, *Fundamentals of music processing*, Springer Science +Business Media, 2015

[17] M. Goto and R. B. Dannenberg, "Music Interfaces Based on Automatic Music Signal Analysis: New Ways to Create and Listen to Music," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 74–81, 2019.

[18] Z. Duan, S. Essid, C. C. S. Liem, G. Richard, and G. Sharma, "Audiovisual Analysis of Music Performances: Overview of an Emerging Field," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 63–73, 2019.

[19] T. Boril and R. Skarnitzl, "Tools rPraat and mPraat, intefacing phonetic analysis with signal processing," vol. TDS 2016, no. LNAI 9924, pp. 367–374, 2016