

# Synthesis of sung Spanish vowels in lyrical singing by sopranos using the Fant source-filter theory and the Rosenberg glottal pulse model

L. Barrientos, E. Cataldo

**Abstract**—This paper aims to synthesize the sung Spanish vowels with covering considering the soprano vocal category of lyrical singers, including variation of sustained pitches with vibrato effect. The Fant source-filter theory is used to model the production of the sustained vowels. The source is based on the Rosenberg glottal pulse model and filter (the vocal tract) is composed by an all-pole filter model with formant frequencies and bandwidths from the vowels of Spanish language. All the sounds synthesized are available to be accessed and they were submitted to a group of listeners and they received a very good evaluation with respect to their intelligibility and naturalness.

**Keywords**—Glottal signal, Rosenberg model, singing voice, vibrato effect, voice synthesis.

## I. INTRODUCTION

The synthesis of the singing voice is one of the most challenging topics studied by researchers in speech synthesis. From the first works published in the '60s [1], the synthesis of singing voice has attracted the attention of many researchers that have developed innumerable synthesizers [2], [3]. These synthesizers have been improved continuously, in terms of intelligibility and naturalness, over the years [4]. Nowadays, the singing voice synthesizers have been utilized not only for the hobby but also for professional music production, singing robot, vocal technique class, and have become so useful in the entertainment industry [5]. However, the synthesis of the female singing voice needs more effort to achieve a naturalness similar to a real professional singer [6].

According to Sundberg [7], a soprano singer usually sings musical notes whose fundamental frequency ( $F_0$ ) is higher than the first formant frequency ( $F_1$ ) of the spoken vowels. That is, in the high register of the female singing, the  $F_0$  values are higher than  $F_1$  of the speech, especially in the vowels (/i/, /e/, /o/, /u/). In this case,  $F_1$  doesn't reinforce the amplitude of  $F_0$ , and hence the sound emitted is weak. However, the sopranos increase the mouth opening, almost involuntarily, to increase the frequency of  $F_1$  and make it coincide with the values of  $F_0$ , allowing the first formant frequency to reinforce the amplitude of  $F_0$ . That increase of the first formant frequency usually makes the values of  $F_1$  closer to the first formant frequency of the vowel /a/,

and perceptually, this vowel is identified. Therefore, the vocal techniques used by lyric soprano to increase the sound level of the notes emitted in her higher range produce, in turn, an adverse effect on the intelligibility of vowels [8]. That behavior is known as formant matching and was observed by Johan Sundberg through an experimental technique to estimate formant frequencies in female singing [7], [9].

Another issue of research on the female lyric singing is the glottal behavior in the high soprano range investigated by Garnier and Henrich [10]. That investigation supports the existence of two distinct laryngeal mechanisms in the high soprano range as well as pitch jumps associated with changes in formant tuning were also observed.

Investigations that relate to the synthesis of the singing voice and the vowel intelligibility have been carried out in American English, French, German, and Japanese [2], [3], [5], [13]. However, no studies were found in the Spanish language.

This study combines signal analysis, synthesis, and psychophysics to discover details about the intelligibility of the Spanish vowels that are affected in the highest range of the lyric soprano. To develop this study, we synthesized the sung Spanish vowels by using a formant synthesizer based on the source-filter model (Fant) [14], [15].

The source is composed of glottal pulses based on the Rosenberg model and with some acoustic features of the soprano singers such as vibrato and tremolo. The glottal source developed in this work doesn't model phenomena as jitter [11], [12], and only one glottal pulse shape, considering the glottal behavior in the high soprano range investigated by Garnier and Henrich [10], was chosen for the synthesis.

Through an all-pole filter was modeled the vocal tract effects, and from the experimental voices of two professional soprano singers were obtained the formant frequencies and their respective bandwidths. A group of listeners assessed the synthesis of the sung Spanish vowels and the results of the evaluation were analyzed and later discussed in this work. All the sounds are available, and they can be easily accessed.

Finally, an analysis of naturalness and intelligibility in the synthesis of sung Spanish vowel, as well as the possible singer's formant in sopranos presented by Johan Sundberg in 1988 [7], [9] was performed.

## II. THE SOURCE-FILTER THEORY

The complete model presented here is based on the source-filter Fant theory [14], illustrated in Fig. 1.

Luis Barrientos, Electrical and Telecommunications Graduate Program, Universidade Federal Fluminense, Niterói-RJ, e-mail: eduardo\_sandoval@id.uff.br; Edson Cataldo, Telecommunications Engineering Department and Electrical and Telecommunications Graduate Program, Universidade Federal Fluminense, Niterói-RJ, e-mail: ecataldo@id.uff.br.

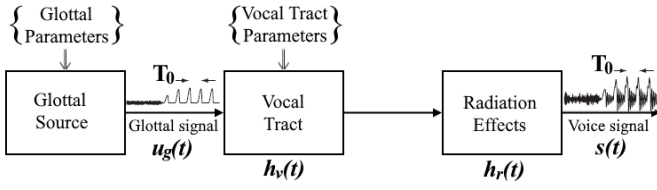


Fig. 1. Sketch of the Fant source-filter model.

The voice signal generated,  $s(t)$ , is given by the convolution of the glottal signal  $u_g(t)$ , the corresponding impulse response function  $h_v(t)$  of the filter that models the vocal tract, and the radiation by the mouth for which the impulse response function is  $h_r(t)$ . Equation (1) is then obtained:

$$s(t) = (h_r * (h_v * u_g))(t), \quad (1)$$

or, in the frequency domain,

$$\hat{s}(\omega) = \hat{h}_r(\omega) \hat{h}_v(\omega) \hat{u}_g(\omega), \quad (2)$$

where  $\hat{\cdot}$  means the Fourier Transform. With this formulation, there is no coupling between the vocal folds and the vocal tract, simplifying the model. In this paper,  $u_g(t)$  is constructed using the Rosenberg glottal pulse model with a specific modulation to obtain effects in the sound produced such as vibrato;  $h_v(t)$  is one coming from the literature and modified according to formants obtained from two professional lyrical singers, and  $h_r(t)$  is a first order high-pass FIR (finite impulse response) filter, such as suggested in [17].

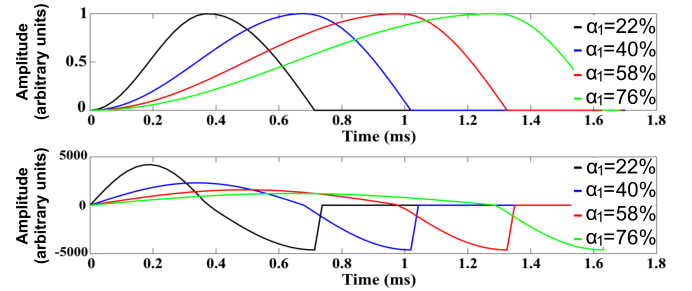
#### A. The glottal signal

The glottal signal was generated from Rosenberg glottal pulse model [18], given by Eq. 3:

$$u_g(t) = \begin{cases} 0.5A_v[1 - \cos(\pi t/(T_P))] & , 0 \leq t \leq T_P \\ A_v \cos(\pi(t - T_P)/(2T_N)) & , T_P < t \leq T_P + T_N \\ 0 & , T_P + T_N < t \leq T_0 \end{cases} \quad (3)$$

Where  $T_0$  is the pitch period;  $A_v$  is the constant related to the amplitude of the glottal pulse;  $T_P$  is the opening time, portion of the pulse with positive slope and  $T_N$  is closing time, portion of the pulse with negative slope. The relative opening and closing times are given by  $\alpha_1 = T_P/T_0$  and  $\alpha_2 = T_N/T_0$  respectively. The glottal open quotient  $OQ$  is defined as the ratio between the duration of the glottal open phase and the fundamental period. And from the relation  $OQ = (T_P + T_N)/T_0$ , the values for  $\alpha_1$  and  $\alpha_2$  could be calculated [13], [19]. Different shapes of the Rosenberg glottal pulse were developed considering the glottal behavior in the high soprano range investigated in [10]. In that investigation, it was observed that for most sustained productions using the full head mechanism,  $OQ$  increased smoothly with  $F_0$ . The maximum values of  $OQ$ , higher than 0.8 and mostly around 0.9, were reached at the upper limit of the laryngeal transition and it is possible to deduce the minimum value 0.5 for lower pitches of  $F_0$ . Subsequently, four shapes of glottal pulse were generated using the range:  $0.42 \leq OQ \leq 0.96$  and maintaining the opening phase larger than the closing phase, as usual. Figure 2 shows in time domain, the effects of

these variations when  $\alpha_1$  increases from 22 % to 76 % with steps of 18 %. The parameters  $A_v = 1$ ,  $F_0 = 589 \text{ Hz}$  and  $\alpha_2 = 20\%$  are fixed. In this paper, the values considered were  $\alpha_1 = 58\%$  and  $\alpha_2 = 20\%$ . The chosen fundamental frequency  $F_0$  amounts to a note of the vocal range of the sopranos singers, likewise the volume of phonation  $A_v = 1$ . With the information of a complete glottal pulse, a glottal signal was generated, calculating the convolution of the glottal pulse with a train of periodic impulses without perturbations.


 Fig. 2. Temporal variations of the glottal pulse (above) and its derivative (below) for four values of  $\alpha_1$  with  $A_v$ ,  $F_0$  and  $\alpha_2$  fixed.

#### B. Vibrato effect

The vibrato corresponds to a periodic fluctuation of the fundamental frequency, and according to [20], vibrato is characterized by four parameters: rate, extent, regularity, and waveform. The vibrato rate ( $V_{Rate}$ ) determines the number of fluctuations per second, typically 4 Hz - 14 Hz. The vibrato extent ( $V_{Ext}$ ) is associated with the amount of pitch variation with average values between 5 ms to 10 ms. Regularity tends to vary most during the negative phase, and the waveform is approximately sinusoidal. In this work, it was added a tremolo effect to the signal to make the synthesis of the sung voice more expressive. The tremolo refers to the periodic variation of intensity to produce changes in the volume of a pitch [21]. This effect is carried out by passing an input signal through a modulating gain amplifier, and hence tremolo can be described by two parameters: amplification level ( $A_{Level}$ ) and tremolo rate ( $T_{Rate}$ ). The tremolo rate follows the same range of vibrato rate, as well as amplification level follows the range of vibrato extent. For generating the vibrato effect in the sung voice, the values of the parameters are  $V_{Ext} = 0.4 \text{ ms}$  and  $V_{Rate} = 4.5 \text{ Hz}$ . On the other hand, the tremolo effect was created by using amplitude modulation, and the parameters were picked out following the parameters of vibrato, hence the  $A_{Level} = 0.002$  and  $T_{Rate} = 4.5 \text{ Hz}$ . Finally, Fig. 3 shows the mouth acoustic pressure for the sung voice with vibrato and tremolo effect.

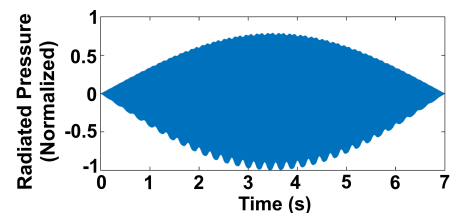


Fig. 3. Mouth acoustic pressure of the sung voice with vibrato effect.

### C. Spanish vowels formants

The resonant frequencies of the vocal tract are called formant frequencies, and they are fundamental in the construction of the voiced sounds. The first five formants are the most important: the two lowest formants determine most of the vowel color, while the third, fourth, and fifth formants contribute to the personal voice timbre [7]. To the vocal extension, the singer's voice can be classified into different vocal categories. For female voices, they are contralto, mezzo-soprano, and soprano. And, for male voices, they are bass, baritone, and tenor. [16], [22]. The most considerable difference between the spectral characteristics of the sung vowel phonemes by male singers and the ones pronounced by nonsingers is on the singer's formant. It is a prominent peak of the spectrum envelope that appears in the proximity of 3 kHz in all vowel specters sung by male singers and belongs to the typical characteristics of a sung vowel. The singer's formant is generated from the clustering of the upper formants (third, fourth, and fifth) with frequencies close to each other. The amplitude of the spectral power of the singer's formant (in dB) depends on the vocal classification. It is lower for the bass category and higher for the tenor category. Regarding sopranos, the amplitude of the peak is lower than in the other vocal categories. It can even be considered as a third or fourth normal formant. Likewise, the level of the peak also varies depending on the intensity of phonation. Considering the center frequency, the peak varies depending on the voice category. In bass singers, the center frequency is around 2.2 kHz; in baritones, around 2.7 kHz; in tenors, around 3.2 kHz; and in altos, around 2.8 kHz. These frequency differences seem to contribute significantly to the differences of the timbres between these voice categories [7], [23]. In this research, the formant frequencies of the sung Spanish vowels were obtained from experimental voice signals produced by 2 lyrical singers, in a total of 23 utterances. The voice signals were recorded in similar conditions: quiet environment (recording studio) and voices recorded directly on a desktop. The subjects were 2 women trained professionally in lyrical singing in the vocal category of soprano, aged 22 and 24. They both have at least 6 years of regular lyrical singing training. Each singer sustained for 9 s each one of the Spanish vowels in the different notes of their vocal range. They started singing the vowel /a/ on the lower pitch of their vocal extension, and continuously, they reached the highest pitch they could produce, following a diatonic scale. The first singer, named soprano 1, produced utterances from the musical note F4 up to E5 (349 Hz - 659 Hz), and the second singer, named soprano 2, produced utterances since the musical note E3 until F5 (167 Hz - 698 Hz). The value of the first five resonance frequencies and their respective bandwidths were extracted using the speech analysis and synthesis Software (Praat), version 5.3. Likewise, the fundamental frequency for each musical note intoned was saved in a spreadsheet in order to be used subsequently in the synthesis of sung voice. This information can be observed in <https://cutt.ly/ayFS4qr>.

## III. RESULTS

### A. Synthesis of sung voice by sopranos

The program developed to the synthesis of sung voice takes, at the beginning of the simulation, data from the soprano singer previously contained in a spreadsheet. These data are as follows: fundamental frequency of the tone (musical note), the first five formant frequencies, as well as their respective bandwidths. When the glottal signal was produced and submitted to modulation, the sound was similar to a *beep sound* with a pitch period corresponding to the soprano singer (soprano 1 or soprano 2). This *beep* presented variations at the start and the end of the utterances, as expected, generating naturalness in the sound. On the other hand, when the algorithm of the vocal tract model was connected, the synthesis of the Spanish vowels was created with acoustic characteristics of the singing voice, such as vibrato and tremolo. Finally, the five Spanish vowels (/a/, /e/, /i/, /o/, /u/) were synthesized with acoustic characteristics of the singing, mimicking the vocal extension of each of the singers (soprano 1 and soprano 2). The audio of all the sung vowels can be accessed at <https://cutt.ly/UyHHQDf>.

### B. Naturalness degree

The listening test was performed by ten Brazilians and ten Colombians, aged from 20 to 53 years. The group of listeners chosen for this test is made up of eight women and twelve men in total. They were asked to classify 14 synthesis of the singing voice by sopranos as natural and expressive or synthetic, without receiving any additional information about the stimuli. The classification could be between a range of 0 to 5, being 0 incomprehensible synthesis, and 5 an excellent natural sound. This test was performed by the listeners individually. The test performed here follows studies by [24], [25] in tests applied to estimate the naturalness of synthetic sounds generated by models of the vocal tract.

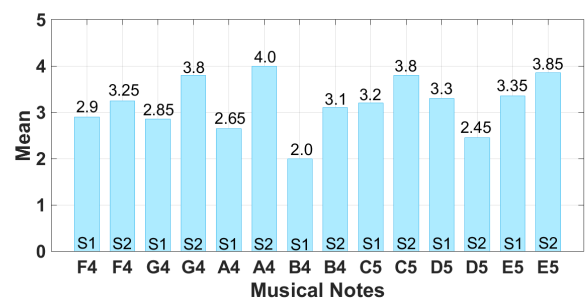


Fig. 4. Bar graph of the perceptual test of the synthesis of sung vowels by sopranos considering the mean of naturalness.

The bar-graph of the Fig. 4 contrasts all the synthesis of the sung voice with resonance parameters of the sopranos 1 and 2 in the range F4 – E5. As a result, 6 of 7 synthesis of the sung voice by soprano 2 obtained a higher mean of naturalness compared to the synthesis of the sung voice by soprano 1. The note A4 ~ 440 Hz presents a mean of naturalness equivalent to 80%, likewise, in the E5 ~ 659 Hz it presents a mean of naturalness corresponding to 77%, both were produced with the resonance frequencies of the soprano 2. The pitches also

produced with the resonance frequencies of the soprano 2 that presented a naturalness mean not less important were  $C5 \sim 523 \text{ Hz}$ , and  $G4 \sim 523 \text{ Hz}$  both with the 76 %. The only pitch synthesized with resonance parameters of soprano 1 that presented the highest degree of naturalness compared to the same synthesized pitches for soprano 2 was  $D5 \sim 587 \text{ Hz}$ .

From the results of the listening test, it was observed that most of the tones that have the best mean correspond to the synthesis generated through the formant frequencies of the soprano 2. This result can be due to the vibrato effect is more compatible with those resonance frequency values [5]. Likewise, The shape of the Rosenberg glottal pulse offers a higher quality to the tone, and therefore, better naturalness.

### C. Intelligibility degree

The same assessments of the previous listening test were carried out in the second test and permitted us to know what of the synthesis of the sung Spanish vowels by the sopranos understudy is more intelligible. These syntheses were separated by musical notes and contain all the Spanish vowels in the following order: (/a/, /e/, /i/, /o/, /u/). The listeners were asked to classify the synthesized Spanish vowels as intelligible without receiving any additional information about the stimuli. The classification could be between a range of 0 to 5, being 0 unknowns, and 5 intelligible. This test was also performed by the listeners individually, and the sounds of the sung vowels correspond to those used in the naturalness test. The bar-graph of Fig. 5 represents the five Spanish vowels with the highest intelligibility degree in the perceptual test, and the audio corresponding to these syntheses can listen at <https://cutt.ly/4yVGUcr>.

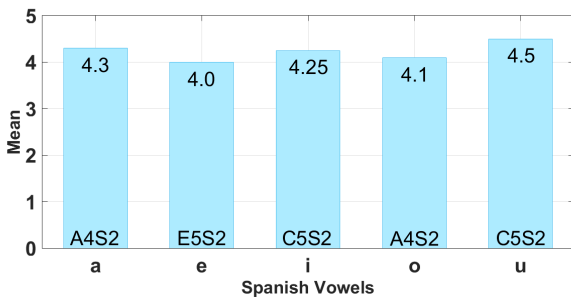


Fig. 5. Bar graph of perceptual test of the synthesis of vowels sung by sopranos considering the mean of intelligibility of vowels.

The synthesis of the vowels /a/ and /o/ sustained in the pitch A4 by the soprano 2 (A4\_S2) reached intelligibility means of 4.3 and 4.1 respectively, which represents the vowels (/a/ and /o/) with the highest intelligibility degree. This pitch (A4) also presented the best naturalness degree of all the synthesis of the sung voice in this research. The synthesis of the vowel /e/ sustained in the pitch E5 by the soprano 2 (E5\_S2) reached intelligibility mean of 4.0, and hence, represented the vowel that obtained the highest intelligibility degree of 14 synthesis for this vowel. The pitch E5 sustained by the soprano 2 also reached a good naturalness degree. The pitch of C5 sustained by the soprano 2 (C5\_S2) obtained

intelligibility mean of 4.25 and 4.5 for the vowels /i/ and /u/ respectively. Particularly, the synthesis of the Spanish vowels sustained in the pitches (A4\_S2), (C5\_S2), and (E5\_S2) are of the great interest for a spectral analysis since this pitches reached the highest mean of naturalness and intelligibility.

### D. Singer's Formant in Sopranos

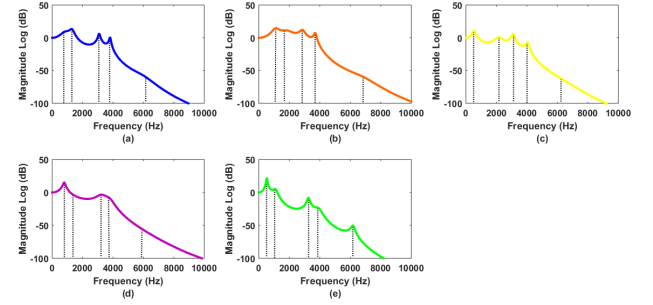


Fig. 6. Frequency response for the five Spanish vowels sung that presented the greatest intelligibility degree (a) vowel /a/, (b) vowel /e/, (c) vowel /i/, (d) vowel /o/ and (e) vowel /u/.

The Fig. 6 presents the frequency responses of the five Spanish vowels with the greatest intelligibility degree in the perceptual test. It can be highlighted a spectrum envelope peak in the vicinity of the 3 kHz, provided by the higher formant frequencies clustering ( $F_3$ ,  $F_4$ ,  $F_5$ ). Regarding this clustering,  $F_5$  is quite far from  $F_4$  in frequency, and hence doesn't adequately participate in the construction of the singer's formant. This is a common feature of the five frequency responses developed. The first five formant frequencies (in Hz) corresponding to the sung Spanish vowels with the greatest intelligibility degree in this research appear in Tab. I. Thus, it can see that the super tones sustained by the soprano 2 are very high compared to normal frequency values of the first formant in most vowels, except the vowel /a/ which is  $808 \text{ Hz}$  and indicates that for this vowel, the singer doesn't need to adjust the first formant frequency ( $F_1$ )[26], [27].

TABLE I  
FORMANT FREQUENCIES (IN Hz) FOR THE SUNG SPANISH VOWELS.

Pitch	$F_0$	Vowel	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$
A4	435	a	808	1304	3088	3808	6025
E5	646	e	1091	1891	2873	3706	6869
C5	515	i	536	2159	3104	4051	6276
A4	435	o	790	1551	3226	3795	5928
C5	513	u	526	1108	3251	3947	6179

The bandwidths of the first five resonance frequencies for the sung Spanish vowel in a super pitch of the soprano 2 are shown in the Tab. II. These bandwidths were used for the syntheses and represent the effects of loss of the vocal tract [17].

In particular, the bandwidths of the formant frequencies for the vowel /u/ are relatively low (less than  $200 \text{ Hz}$ ), reflecting low losses in the vocal tract of the soprano 2, hence, the propagation of sound through the vocal tract is more efficient than the other vowels (/a/, /e/, /i/, /o/). This characteristic

positively influences in the naturalness ( $mean = 3.8$ ) and intelligibility ( $mean = 4.5$ ) of the vowel /u/.

TABLE II  
BANDWIDTH (IN HZ) OF THE FORMANT FREQUENCIES.

Pitch	Vowel	$Bw_1$	$Bw_2$	$Bw_3$	$Bw_4$	$Bw_5$
A4	a	282	126	53	40	808
E5	e	195	357	146	59	998
C5	i	90	202	68	76	1411
A4	o	77	1676	291	286	2153
C5	u	23	116	62	198	87

#### IV. CONCLUSIONS

High quality and naturalness were perceived in most of the syntheses of the sung Spanish vowels by the soprano 2. This result is due to the shape of the Rosenberg glottal pulse. The glottal opening quotient used for these syntheses ( $O_q = 0.78$ ), as well as the acoustic effects of vibrato and tremolo added to the synthesis, permitted achieving a maximum mean of naturalness of 4.0 to the note A4 sustained by soprano 2 (A4\_S2). The syntheses of the five Spanish vowels that obtained the highest degree of intelligibility in the perceptual test were also sustained in the pitches that obtained the highest degree of naturalness among all the syntheses of this investigation. In fact, the vowel /u/ sustained in E5 of the soprano 2 reached the maximum mean of intelligibility with a value of 4.5 and corresponds to 90%. Regarding frequency responses for the Spanish vowels synthesized with the highest intelligibility degree, it was evidenced a spectrum envelope peak in the vicinity of the 3 kHz characterizing the singer's formant, however, in the grouping of formant frequencies that create this envelope.  $F_5$  is quite far from  $F_4$  and does not adequately participate in the construction of the spectrum envelope peak. This is a common feature of the frequency responses developed for each vowel. The bandwidths for the formant frequencies of the vowel /u/ are relatively low (less than 200 Hz), causing low losses in the soprano vocal tract. This characteristic positively influenced the naturalness ( $mean = 3.8$ ) and intelligibility ( $mean = 4.5$ ) of the vowel, reaching the maximum degree among all synthesized vowels. Finally, new perspectives emerged from this work, and hence more complex models of the glottal pulse could be used to develop syntheses of the sung Spanish vowels with the resonance characteristics of the soprano 2. Consequently, these syntheses can be submitted to perceptual tests with the same listeners, thus, developing a comparison of glottal models regarding the degree of naturalness and intelligibility produced. On the other hand, different acoustic effects may also be exploring to improve the naturalness and expressiveness degree of the synthesis, especially of the soprano 1 that did not get good mean in the perceptual tests of this research.

#### ACKNOWLEDGMENTS

This work was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico - Grant 303234/2017-2) - Brazil.

#### REFERENCES

- [1] C. C. Lochbaum and J. L. Jr. Kelly, "Speech synthesis," *Proceedings of the Speech Communication Seminar*, pp. 583-596, 1962.
- [2] M. Mellody, F. Herseth, G. H. Wakefield, "Modal distribution analysis, synthesis, and perception of a soprano's sung vowels," *Journal of Voice*, v. 15, n. 4, pp. 469-482, 2001.
- [3] M. Nishimura, K. Hashimoto, K. Oura, Y. Nankaku, k. Tokuda, "Singing Voice Synthesis Based on Deep Neural Networks," *Interspeech*, pp. 2478-2482, September 2016.
- [4] N. D'Alessandro, P. Woodruff, Y. Fabre, T. Dutoit, S. Le Beux, B. Doval, C. d'Alessandro, "Realtime and accurate musical control of expression in singing synthesis," *Journal on Multimodal User Interfaces*, v. 1, n. 1, pp. 31-39, 2007.
- [5] T. Nose, M. Kanemoto, T. Koriyama, T. Kobayashi, "Hm-based expressive singing voice synthesis with singing style control and robust pitch modeling," *Computer Speech & Language*, v. 34, n. 1, pp. 308-322, 2015.
- [6] L. Ardaillon, *Synthesis and expressive transformation of singing voice*. Ph.D. thesis, Paris 6, 2017.
- [7] J. Sundberg, "Vocal tract resonance in singing," *The NATS Journal*, v. 44, n. 4, pp. 11-20, 1988.
- [8] P. P. de Julián, "Modificación o ajustamiento de las vocales españolas en el canto lírico," *Estudios de fonética experimental*, pp. 263-293, 2016.
- [9] J. Sundberg, "Research on the singing voice in retrospect," *TMH-QPSR*, v. 45, n. 1, pp. 11-22, 2003.
- [10] M. Garnier, N. Henrich, L. Crevier-Buchman, C. Vincent, J. Smith, J. Wolfe, "Glottal behavior in the high soprano range and the transition to the whistle register," *The Journal of the Acoustical Society of America*, v. 131, n. 1, pp. 951-962, 2012.
- [11] I. R. Titze, *Principles of voice production*. Prentice Hall, 1994.
- [12] E. Cataldo and C. Soize, "Stochastic mechanical model of vocal folds for producing jitter and for identifying pathologies through real voices," *Journal of biomechanics*, v. 74, pp. 126-133, June 2018.
- [13] N. H. Bernardoni, *Etude de la source glottique en voix parlée et chantée: modélisation et estimation, mesures acoustiques et électroglottographiques, perception*. Ph.D. thesis, Université Pierre et Marie Curie-Paris VI (UPMC), 2001.
- [14] G. Fant, "Acoustic theory of speech production," sgravenhage: Mouton, *The Netherlands*, 1960.
- [15] G. Fant, "The source filter concept in voice production," *STL-QPSR*, v. 22, n. 1, pp. 21-37, 1981.
- [16] M. L. Facal, *La voz del cantante: estudio comparativo del análisis objetivo y subjetivo de la voz hablada y cantada*. Librería Akadia Editorial, 2005.
- [17] L. R. Rabiner, and R. W. Schafer, *Theory and applications of digital speech processing*. v. 64, Pearson Upper Saddle River, NJ, 2011.
- [18] A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," *The Journal of the Acoustical Society of America*, v. 49, n. 2B, pp. 583-590, 1971.
- [19] B. Doval, C. d'Alessandro, N. Henrich, "The spectrum of glottal flow models," *Acta acustica united with acustica*, v. 92, n. 6, pp. 1026-1046, 2006.
- [20] M. Hirano, S. Hibi, S. Hagino, "Physiological aspects of vibrato," *Vibrato*, pp. 9-33, 1995.
- [21] L. Regnier, G. Peeters, "Singing voice detection in music tracks using direct voice vibrato detection," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1685-1688, 2009.
- [22] S. Hertegard, J. Gauffin, J. Sundberg, "Open and covered fiber optics, inverse singing as studied by means of filtering, and spectral analysis," *J Voice*, v. 4, pp. 220-230, 1990.
- [23] T. J. Millhouse, F. Clermont, "Perceptual characterisation of the singer's formant region: a preliminary study," *Proceedings of the Eleventh Australian International Conference on Speech Science and Technology*, pp. 253-258, 2006.
- [24] D. R. Allen, W. J. Strong, "A model for the synthesis of natural sounding vowels," *The Journal of the Acoustical Society of America*, v. 78, n. 1, pp. 58-69, 1985.
- [25] E. Cataldo, F. R. Leta, J. Lucero, L. Nicolato, "Synthesis of voiced sounds using low-dimensional models of the vocal cords and time-varying subglottal pressure," *Mechanics Research Communications*, v. 33, n. 2, pp. 250-260, 2006.
- [26] G. Berndtsson, J. Sundberg, "Perceptual significance of the center frequency of singer's formant," *Scandinavian Journal of Logopedics and Phoniatrics*, v. 20, n. 1, pp. 35-41, 1995.
- [27] R. Weiss, W. Brown Jr, J. Moris, "Singer's formant in sopranos: fact or fiction?," *Journal of Voice*, v. 15, n. 4, pp. 457-468, 2001.