

Margins of tolerance and reference values for the formant vowels for use in voice therapy for the deaf in commercial computer

Margens de tolerância e valores de referência para os formantes de vogais orais para uso em terapias de voz para surdos em computador comercial

André Luis Maciel Leme¹
Márcio Abud Marcelino^{2,3}
Pedro Paulo Leite do Prado³

ABSTRACT

Purpose: This study presents the margins of minimum and maximum tolerances for the frequencies of the first three formants (F1, F2, and F3) in the pronunciation of vowels of Brazilian Portuguese for use in voice therapy for the deaf. **Methods:** The frequencies were obtained from the voluntary collaboration of 53 adults who had their voices recorded and converted into digital signals during the phonation of each of the seven vowels (/a/, /e/, /ɛ/, /i/, /o/, /ɔ/, /u/) sustained for approximately one second. The samples were divided into two groups: male and female. The recording and extraction of the formants were conducted by software developed exclusively for this purpose in MATLAB platform using the eight-coefficient LPC (Linear Predictive Coding) algorithm. **Results:** The results showed that a consistent reference for the mean values of the F1, F2, and F3 frequencies can be obtained through graphical and statistical analysis of the samples collected from the voice signals. **Conclusion:** The reference values obtained were analyzed and can be used for the calibration of devices and serve as a basis for training oralization for the deaf.

RESUMO

Objetivo: Este trabalho apresenta as margens de tolerâncias mínimas e máximas para as frequências dos três primeiros formantes (F1, F2 e F3) na pronúncia das vogais orais do português brasileiro para utilização em terapias de voz para surdos. **Método:** As frequências foram obtidas a partir da colaboração voluntária de 53 indivíduos adultos que tiveram as vozes gravadas e convertidas em sinais digitais durante a emissão de cada uma das sete vogais (/a/, /e/, /ɛ/, /i/, /o/, /ɔ/, /u/), de maneira sustentada por cerca de um segundo. As amostras foram distribuídas em dois grupos, um masculino e outro feminino. A gravação e a extração dos formantes foram efetuadas através de *software* próprio desenvolvido para este fim na plataforma MATLAB, utilizando o algoritmo LPC (Linear Predictive Coding) de oito coeficientes. **Resultados:** Os resultados mostraram que uma referência consistente para os valores médios das frequências de F1, F2 e F3 pôde ser obtida através da análise gráfica e estatística das amostras de sinais de voz coletada. **Conclusão:** Os valores de referência encontrados foram analisados e podem ser usados para calibração de dispositivos e podem servir de base para o treinamento de oralização para surdos.

Keywords

Formants
Oralization
Deaf
Vowels
Tolerance

Descritores

Formantes
Oralização
Surdo
Vogais
Tolerância

Study carried out at the Programa de Pós-graduação em Engenharia Mecânica, Universidade de Taubaté – UNITAU - Taubaté (SP), Brazil.

¹ Instituto Federal de São Paulo – IFSP - Bragança Paulista (SP), Brazil.

² Universidade Estadual Paulista – UNESP - Guaratinguetá (SP), Brazil.

³ Universidade de Taubaté – UNITAL - Taubaté (SP), Brazil.

Financial support: nothing to declare.

Conflict of interests: nothing to declare.

Correspondence address:

André Luis Maciel Leme
Instituto Federal de São Paulo – IFSP
Avenida Francisco Samuel Lucchesi
Filho, 770, Penha, Bragança Paulista
(SP), Brazil, CEP: 12929-600.
E-mail: andre.m.leme@ifsp.edu.br

Received: May 20, 2015

Accepted: June 23, 2015

INTRODUCTION

The deaf present difficulties in speaking simply because they cannot hear, except for those who have problems in the vocal tract that prevent them from emitting sounds. Hearing individuals repeat the sounds they hear in order to build their oral repertoire. The ability to recall sounds found in hearing individuals is limited or absent in the deaf. This article addresses this issue considering that the establishment of reference frequencies for the first three formant vowel sounds can assist in the vocal training of deaf adults who wish to improve or acquire oralization in Brazilian Portuguese.

Historically, the deaf are surrounded by “*ouvvinista*” (of those who can hear) - a term coined by Skliar⁽¹⁾ to note that the deaf have to follow all the concepts of hearing individuals⁽²⁾ - ideology, power, and interests.

The deaf can be fully literate in Portuguese and LIBRAS (Brazilian Sign Language) and have lip-reading ability, and still not be able to properly express their ideas and thoughts to hearing individuals who are not LIBRAS users because of the language barrier, in the same way that Brazilians who speak only their native language would have difficulty being understood by Russian speakers who also know only their language. The aspect of orality for the deaf pervades the history of education since ancient times. In Greco-Roman antiquity, speech was directly associated with thought, to the point of considering that thought could not be developed without speech⁽³⁾. The first initiatives that sought a means of education for the deaf had financial motivation, and were limited to the children of wealthier families concerned with the passage of assets to their deaf offspring. Although not of inclusive nature, these initiatives have flourished, because they served as a basis for other educators who reproduced the teaching methods from the contact with deaf children of noble families who were educated by these methods⁽⁴⁾.

This historical mention to the education for the deaf refers to the fact that deaf individuals have been the focus of attention because of their difficulty to communicate with a predominantly hearing society. In Brazil, there are approximately 2 million deaf citizens⁽⁵⁾, classified as having “severe hearing loss” according to the Convention on the Rights of Persons with Disabilities ratified by the Brazilian government. This is a controversial aspect. The deaf are not disabled; they are simply different and need other means to communicate. In this sense, the history of education for the deaf in Brazil is intertwined with the history of Speech-language pathology⁽⁶⁾. In general, this professional is the one who has the first contact with deafness, along with the physician. The role of Speech-language pathology extends beyond the application of bimodal, oralist, or bilingual techniques. Speech-language therapists play a more comprehensive role in the contact with deafness, because their professional practice does not involve only the treatment, it makes an enormous difference in a person’s life. Considering either approach,

the work should, indeed, promote the global development of individuals regardless of the form of communication they use. Global development refers to linguistic, intellectual, social, and academic development, and especially to identity preservation⁽³⁾.

This paper aims to offer a small contribution to support Speech-language pathology professionals in their work with the adult deaf community.

PURPOSE

The objective of this study is to identify the margins of tolerance and reference values for the frequencies of the formants F1, F2 and F3, according to gender and age range, for the seven vowels of Brazilian Portuguese (/a/, /e/, /ɛ/, /i/, /o/, /ɔ/, /u/), which constitute the core of the syllables of the words of that language⁽⁷⁾.

It is expected that these reference values may allow adult deaf individuals, supervised and guided by a Speech-language therapist, to train oralization through a computational resource that provides them with a graphic display of the reference values and the values of their sound production, helping them to calibrate their voice as close to the displayed reference as possible. The use of a regular computer, whether a desktop or a laptop, connected to a simple microphone, will provide the Speech-language therapist with an affordable tool to support voice therapy. In this context, this study seeks to find values to compose the parameters to be used in the configuration of a system for the vocal training of deaf adults.

The reference values found will be used in the construction of a computer system according to the specifications of the “Interactive System to Aid the Deaf” patent application⁽⁸⁾ which seeks to provide the Speech-language therapist with a tool to support the development of oral language therapy in individuals with severe hearing impairment. The system will include graphic features in three dimensions of the frequencies produced by vocalization performed by the deaf, collected and calculated in real time. The frequencies formed in the vocal tract will be displayed overlapped, in the same scenario, on the reference values obtained for the F1, F2, and F3 frequencies. The margin of tolerance will be represented by semi-transparent, three-dimensional geometry. The whole system operation will be performed by a Speech-language therapist. Therefore, the deaf adult will count on visual support on the computer screen to allow equalization between the production of their speech and the references obtained from individuals of the same age and gender.

METHODS

The process of voice production and all the dynamics associated with it, depicting the physical, mechanical, and acoustic aspects of the whole vocal tract involved in this process, is widely described in the literature on speech acoustics. The studies by Chiba and Kajiyama⁽⁹⁾, two Japanese researchers,

have been considered first milestone in the development of the knowledge area on the study of acoustics involved in the speech production process. The book “The Vowel”, from 1941, consolidated their research and has been widely cited to date.

Another researcher who greatly contributed to speech production studies was Gunnar Fant⁽¹⁰⁻¹¹⁾, a Swedish scholar who devoted his life to research related to vocalic production since his graduation at the Department of Telegraphy and Telephony of the Royal Institute of Technology⁽¹²⁾. In 1960, Gunnar Fant published a study⁽¹³⁾ that had significant impact on the evolution of vowel production studies. His experiments and calculations allowed us to clearly observe the relationship between the acoustic parameters of vowel production and the artificial signals produced by a model.

Authors such as Stevens⁽¹⁴⁾, O’Shaughnessy and Deng⁽¹⁵⁾, and Flanagan⁽¹⁶⁾ address not only the physiological aspects, but also the physics involved in the speech production process. These authors demonstrate, mathematically, how the frequencies observed during speech can be calculated. They address variables such as viscosity and thickness of the vocal tract and the section considered to obtain the frequency.

One of the aspects of this process is the identification of the frequencies of resonance waves that characterize a particular sound produced. These specific characteristics are obtained from the mathematical analysis of the transfer function of the passage of the acoustic wave along the vocal tract, from its production in the vocal folds to its exit through the lips. Elements such as air density and velocity, vocal tract length, cutting area where the wave is analyzed, etc., compose the modeling of this system, resulting in the calculation of the sound pressure produced^(11,14-15).

Several techniques are used to treat the voice signal. In general, these techniques are based on small anatomic variations of the vocal tract during phonation. The sound wave produced by the vibration of the vocal folds resonates throughout the vocal tract, finding greater or lesser constriction to its passage depending on the configuration of the system components (tongue, soft palate, hard palate, etc.). One of the techniques used for modeling the speech production process is called Linear Predictive Coding (LPC), which was proposed by the United States Department of Defense in 1984, initially as a model for human speech coding⁽¹⁷⁾.

The most important aspect of LPC is the linear prediction filter, which allows the value of the next sample to be determined by a linear combination of the previous samples⁽¹⁷⁾.

This and other technical approaches have been used in the computational process for extracting formants based on collected voice signals, as described ahead. The recording of voice signals was performed in a laptop computer connected to a microphone. Sampling rate was 11025 Hz. The study sample comprised 53 individuals, 40 men ($\cong 75\%$) aged 17 to 59 years, 12 women ($\cong 23\%$) aged 17 to 55 years, and one 6-year-old female child ($\cong 2\%$). None of the participants presented speech problems. The participants were volunteers who signed an

Informed Consent Form (ICF) approved by the Research Ethics Committee (REC) of “Universidade de Taubaté” under protocol no. 985459.

The volunteers were given an explanation on the project content and the importance of their contribution to research aimed at human development and assistive technology therapies⁽¹⁸⁾. A demonstration of the sound recording procedure was conducted so that the volunteers became familiar with the process. The procedure consists in breathing lightly, obtaining sufficient air so that the sound production does not occur forcibly, which would alter the proposal of natural phonation of a vowel. After that, the individual approaches the microphone to a distance of 2 cm of the mouth and emits the sound of the requested vowel for approximately 1 sec. The graphical interface of the software displays a green quadrangle at the top right of the screen, a graph of the first three formants, and the message “Recording Performed Successfully”, as shown in Figure 1.

The data were stored in one-dimensional vector format ready for use by MATLAB scripts. Thus, the formant extraction algorithm could be applied to each individual sample (individual and vowel) following the three main sub-routines: Signal filtering - hamming windows, pre-emphasis, and high-pass filter; Formant extraction: 8-coefficient LPC; and Track selection - selects tracks > 90 Hz (formant) and passband < 400 Hz.

Of the 53 participants, the most significant sample was composed of male adults aged between 19 and 59 years; therefore, values from the substrate of 38 male individuals were considered for this study. Among the 38 samples, all from individuals without speech problems, 22 are from Sao Paulo state, 12 from Minas Gerais state, and the others from other Brazilian states. Age distribution in years was as follows: two with 19, four with 20, four with 21, three with 22, four with 23, two with 25, two with 26, one with 27, three with 29, one with 34, two with 38, two with 39, one with 50, one with 51, two with 52, one with 57, one with 58, and two with 59.

Regarding data analysis, some utility programs were created to optimize and accelerate the extraction of formants of the whole sample to guarantee the integrity of the data, that is, ensure that the extraction of formants of each sample followed the same process. Individual data were stored in a repository in text format to be imported into an Excel spreadsheet to aid in the assembly of statistical graphs and calculations.

Data analysis was performed using statistical concepts (standard deviation, median, and coefficient of variation)⁽¹⁹⁾. Each formant and phoneme received individualized treatment for data collection. To illustrate the method used to obtain the reference values, we took as an example the data of the first formant (F1) of the phoneme /a/ as from the indicated sample, as shown in Figure 2.

The number of classes (block groups) was defined according to the Sturges’ rule⁽²⁰⁾ and the interval between classes was based on the number of samples used⁽²⁰⁾. The value for the

margin of tolerance was calculated using the coefficient of variation found. The value of the median (reference center) was used for graphical representation of the frequencies of each formant. The margin of tolerance was designed by subtracting and adding the amount resulting from the application of the

coefficient of variation on the median value. These values enabled the drawing of a three-dimensional figure around the central point, as shown ahead.

Finally, in this study, the algorithm was structured in a single program to ensure the integrity of the extracted data.

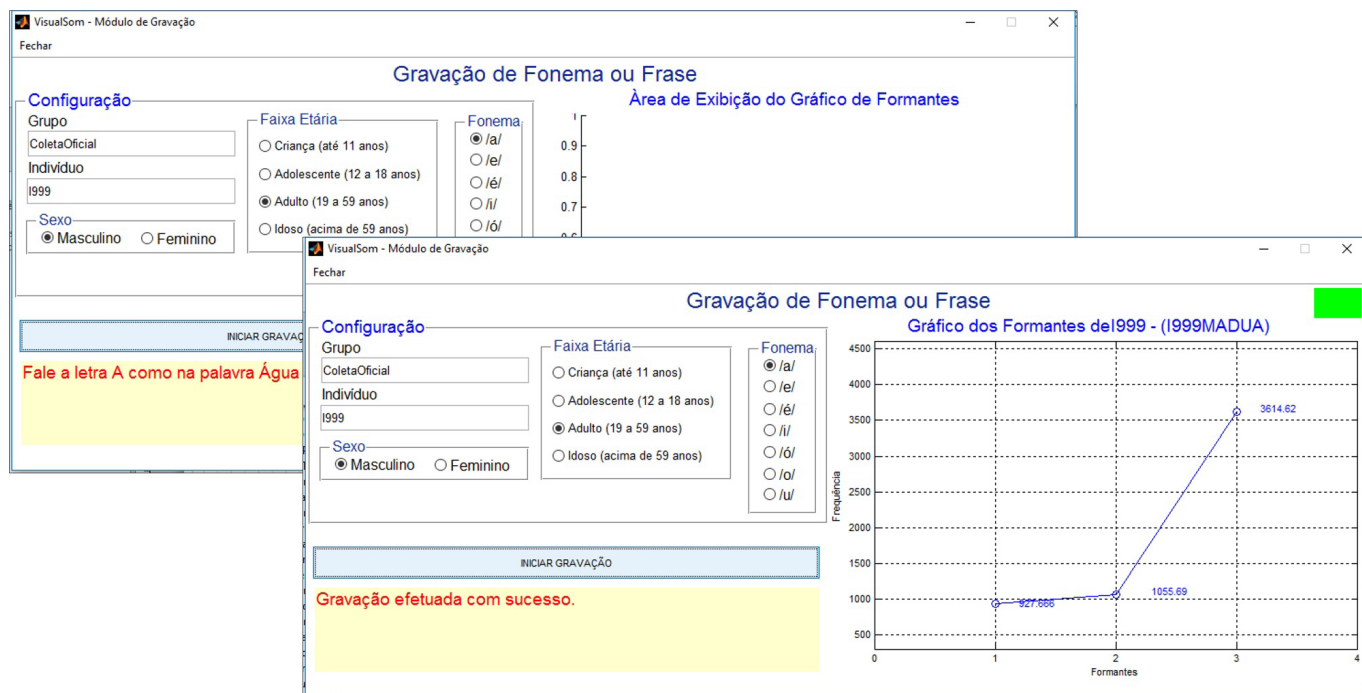


Figure 1. Graphical interface software program for sample collection

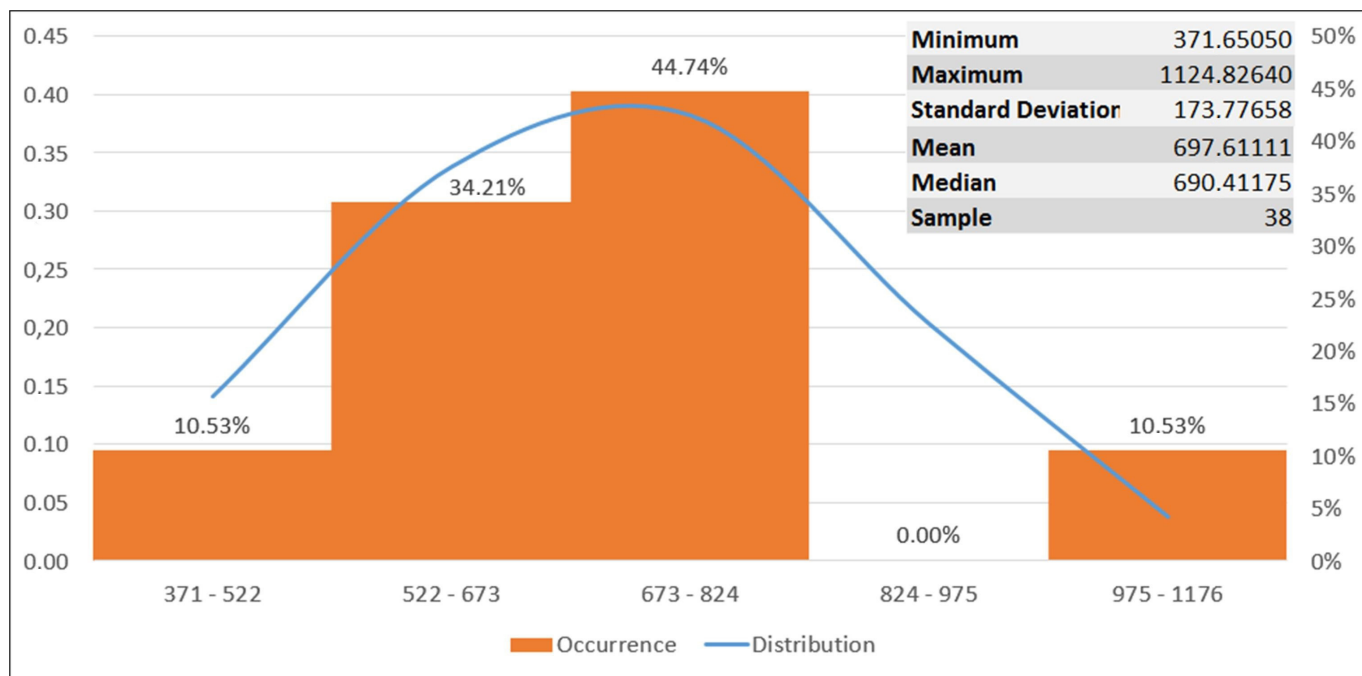


Figure 2. Data of the frequencies of the first formant (F1) of the phoneme /a/ extracted from a voice recording sample of 38 male adults without vocal problems

RESULTS

The data obtained with the method described, for the selected samples, are consolidated in Table 1, according to each phoneme. The option for using the sample median rather than its mean is due to the fact that the sample outlying frequencies (in Hz) could affect the result. The differences are small, but the purpose of this study is to find the most accurate possible approximation between the reference values of the formants of vowels.

To corroborate the results of this research, we adopted frequencies of formants derived from the production of vowels of Brazilian Portuguese from previous studies.

In her study, Oliveira presented a table (Table 4 on page 112; see references)⁽²¹⁾ with the mean values of the formants F1, F2, and F3 for the phonation of phoneme /a/ by two male individuals without vocal problems aged 18 to 21 years, respectively. The frequency values extracted from the referred table were obtained when the pronounced word, according to the research method cited, contained the phoneme /a/ in its stressed syllable, as follows: F1=769 Hz and F2=1325 Hz, for the first individual; F1=845 Hz and F2=1371 Hz, for the second individual.

A classic study by Behlau⁽²²⁾ presents a table of values for F1 and F2 of the stressed vowels for men, women, and children. The frequency values for men for the phoneme /a/ are F1=807 Hz and F2=1440 Hz.

A more recent study, from 2013, shows a comparison of the production of vowels among children, men, and women. In this article, the frequency values for men for the phoneme /a/ are F1=620 Hz and F2=1478 Hz⁽²³⁾.

Table 2 shows a comparison between the frequency values obtained in this research and those of the other three studies.

As it can be observed, the differences between values are within the margin of tolerance found for formants F1 (171.983 Hz) and F2 (410.408 Hz).

Graphic presentation of results

In a comparative graphic approach between F1, F2, and F3 of the seven vowels, establishing the median of the sample frequencies as the central point, surrounded by a parallelepiped that represents the margin of tolerance found, visual distinction between the vowels can be observed - one of the secondary objectives of this study - providing visual feedback of the vowel pronunciation, as shown in several forms in Figure 3.

The graphical display in three different views reinforces research in which the pattern of formants characterizes the vowels according to the vocal tract configuration⁽²⁴⁾. The predominance of F1 and F2 in vowel characterization is well known, but the literature reports the first three forms as the reference for vowel identification⁽²⁵⁾.

DISCUSSION

By approaching and plotting all formants and their margins of tolerance in mutual, two-dimensional, and three-dimensional view, it was possible to observe that the frequency values allow the use of these frequency references as a basis for vocal training, considering that they show the distinction between the vowels of Brazilian Portuguese through the first three formants.

To validate the reference values obtained, an experiment was conducted using a vocal training interface designed for this study, in which reference formants are displayed for a hypothetical oralization training of the phoneme /a/.

Table 1. Consolidation of margin of tolerance and reference values of the frequency of formants extracted from the voice recording of 38 male adults without vocal problems

Phoneme	Median			Margin of tolerance		
	F1	F2	F3	F1	F2	F3
/a/	690.412	1320.648	3053.270	171.983	410.408	513.845
/e/	348.628	2108.066	3080.864	52.681	255.704	317.906
/ɛ/	508.730	1920.590	2872.906	77.775	222.408	273.125
/i/	275.810	2339.031	3250.589	39.006	448.630	271.076
/o/	388.536	1437.203	3208.146	81.623	793.545	432.587
/ɔ/	589.081	1667.094	3338.410	96.529	865.978	420.796
/u/	325.173	1129.021	3234.571	57.121	653.504	403.147

Table 2. Comparison of the frequencies of formants F1 and F2 between the values found in this study and the values obtained in three other studies

Values obtained in this study	F1 (Hz)		F2 (Hz)		Differences	
	(a) 690.41	(A)1320.65	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)
Values obtained by Oliveira ⁽²¹⁾						
Individual 1	(c) 769	(C) 1325	(c)-(a) 78.588	(C)-(A) 4.352		
Individual 2	(d) 845	(D) 1371	(d)-(a) 154.588	(D)-(A) 50.352		
Values obtained by Behlau ⁽²²⁾	(e) 807	(E) 1440	(e)-(a) 116.588	(E)-(A) 119.352		
Values obtained by Brod ⁽²³⁾	(f) 620	(F) 1478	(f)-(a) -70.412	(F)-(A) 157.352		

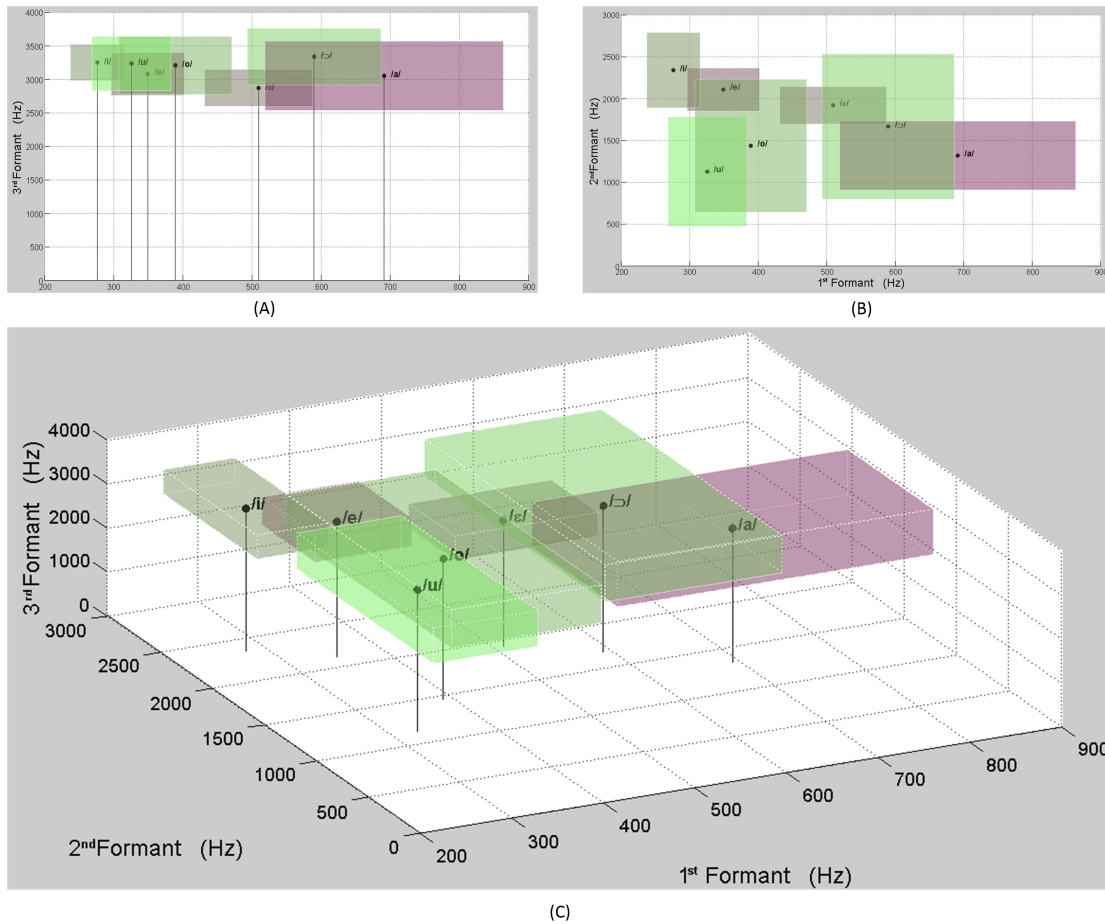


Figure 3. Comparative overview between the frequencies of reference values and margin of tolerance of the first three formants of the seven vowels extracted from the voice recording of 38 male adults without vocal problems. (A) shows the comparison between F1 (horizontal) and F2 (vertical), (B) shows the frequencies of F3, and (C) shows the three-dimensional view of the three formants - F1 (horizontal axis to the right), F2 (horizontal axis to the left), and F3 (vertical axis)

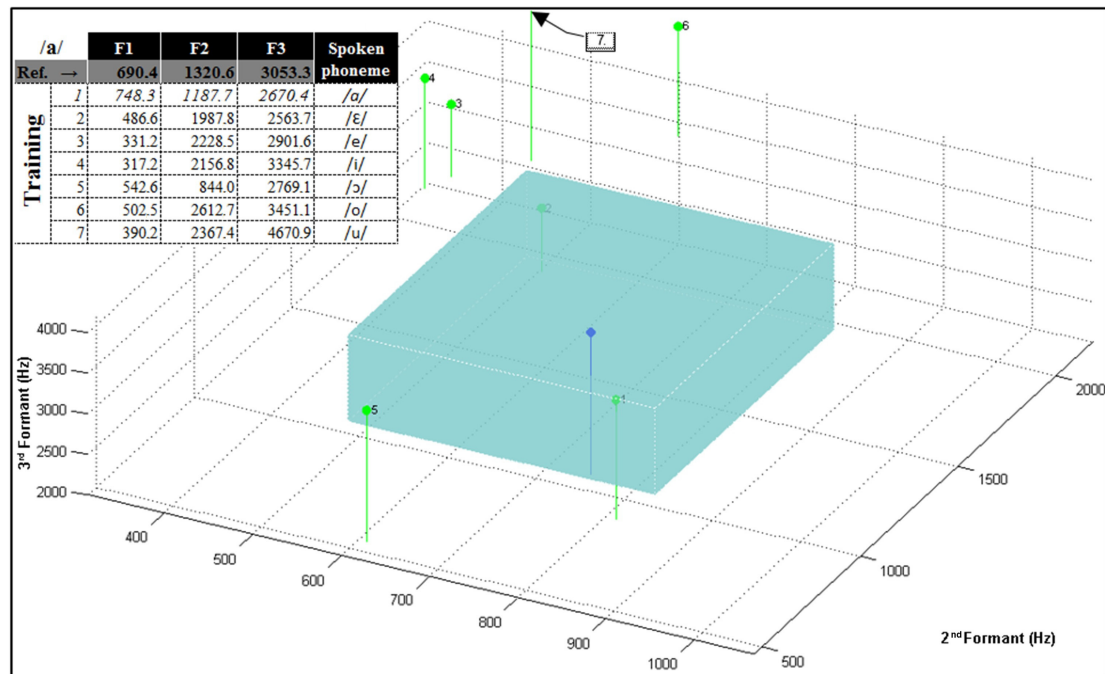
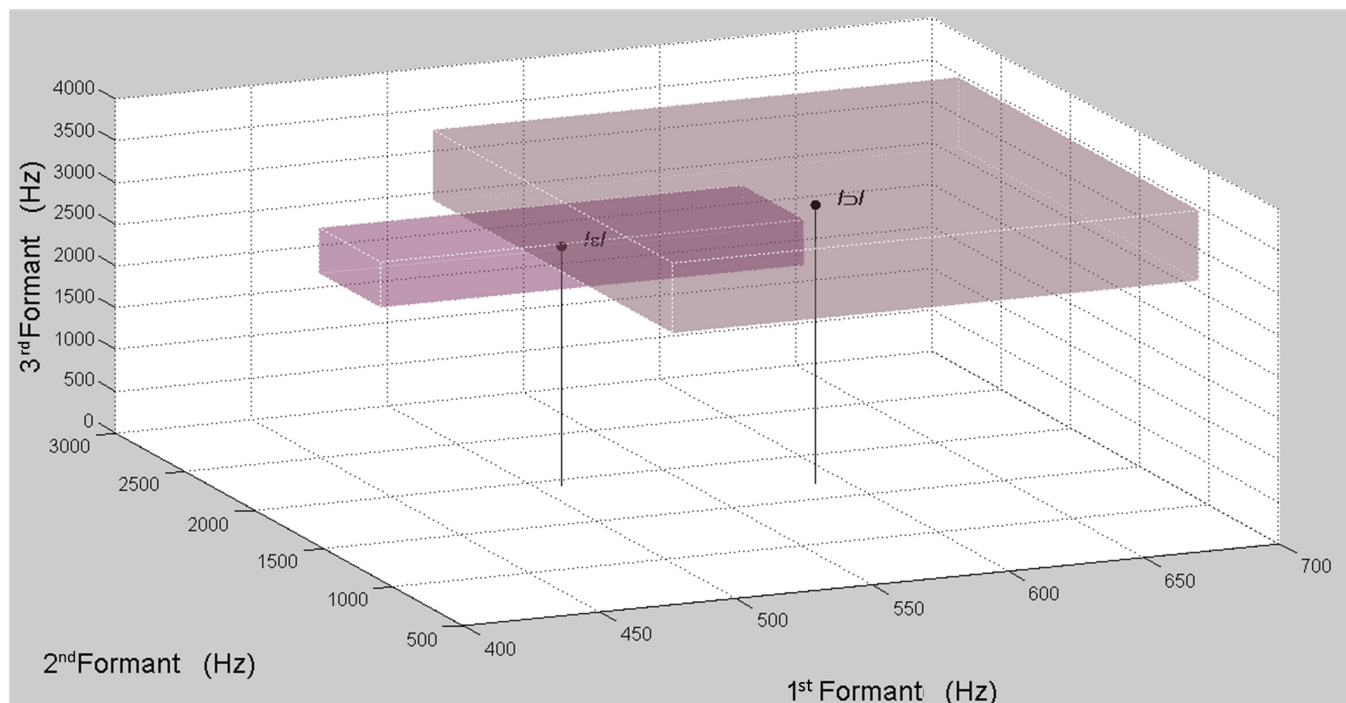


Figure 4. Plotting of reference formants of the vowel /a/ and the phonation of the remaining vowels to demonstrate that the reference characterizes the vowel /a/. Point 1 plotted is the pronunciation of the vowel /a/, which appears next to the reference and within the margin of tolerance

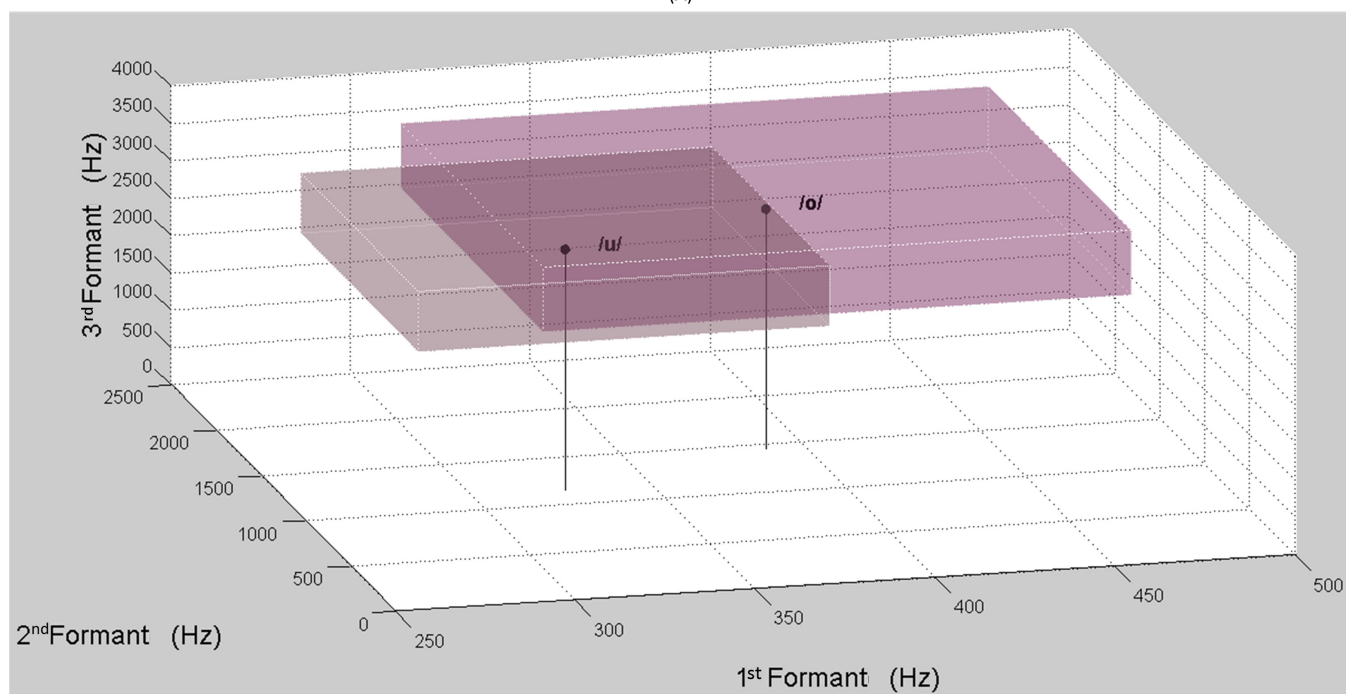
The lighter points (in green) represent the pronunciation of the other phonemes as well as that of the /a/ phoneme performed by a hypothetical individual under therapy, as shown in Figure 4. The result of this experiment shows that it is possible to use the obtained references as a basis for characterizing each of the seven vowels of Brazilian Portuguese, providing Speech-language pathology professionals and researchers with an extra support tool.

It is worth noting that two pairs of phonemes - /ɛ/ and /ɔ/; /o/ and /u/ - showed an overlap of values, as presented in Figure 5.

As the proposal is to offer supporting tools for professionals working in this area, this overlapping of frequencies should be considered in an individual basis at the time of their use and within the context of the approach adopted for its therapeutic utilization.



(A)



(B)

Figure 5. Different views of the reference values of F1, F2, and F3 frequencies extracted from the voice recording of 38 male adults without vocal problems. (A) shows the overlapping of vowels /ɛ/ and /ɔ/. (B) shows the overlapping of vowels /o/ and /u/

CONCLUSION

This study presented a valuable range of information for use as reference frequencies to characterize the seven vowels of Brazilian Portuguese. These reference values can be used to support device calibration and also for the production of new assistive technologies directed to voice therapy.

It can be clearly observed that the three-dimensional view proposal presents visual distinction between the vowels. When they are displayed in the same scenario, it is possible to observe the gap between them. The reference values place them in different spatial regions. The margins of tolerance found provide the individual under voice therapy with a reference for phonation within frequencies that make the sound intelligible and distinguishable by a hearing individual. The repetitive vocal training will assist the deaf in positioning the vocal tract properly when they have to pronounce one of the vowels.

REFERENCES

- Schubert SEM, Coelho LAB. A matemática e a surdez: existem barreiras na aprendizagem da disciplina? In: X Congresso Nacional de Educação; 2011 Nov 7-10; Curitiba, PR. Anais. Curitiba: PUC-PR; 2011. p. 2087-99.
- Skliar C. Uma perspectiva sócio-histórica sobre a psicologia e a educação dos surdos. Porto Alegre: Mediação; 1997.
- Moura MC, Lodi ACB, Harrison KMR. História e educação: o surdo, a oralidade e o uso de sinais. São Paulo: Roca; 1997.
- Lane H. When the mind hears: a history of the deaf. New York: Vintage Books; 1989.
- IBGE: Instituto Brasileiro de Geografia e Estatística. Estatísticas de gênero [Internet]; 2010 [citado em 2013, Ago 13]. Disponível em: <http://www.ibge.gov.br/apps/snig/v1/?loc=0&cat=-1,-2,-3,128&ind=4648>
- Nascimento LCR. Fonoaudiologia e surdez: uma análise dos percursos discursivos da prática fonoaudiológica no Brasil [dissertação]. Campinas: Universidade Estadual de Campinas, Faculdade de Educação; 2002.
- Miranda ARM, Matzenauer CLB. Aquisição da fala e da escrita: relações com a fonologia. Cad Educ. 2010 Jan-Abr;35:359-405.
- Prado PPL, Marcelino MA, inventores. Sistema interativo para auxílio a deficiente auditivo. Brasil patent 0215612-1A2. 2011 Ago 04.
- Chiba T, Kajiyama M. The vowel: its nature and structure. Tokyo: Tokyo-Kaiseikan; 1941.
- Fant G. Acoustic theory of speech production. 2nd ed. Mouton: The Hague; 1970.
- Fant G. Speech acoustics and phonetics. 24th ed. Dordrecht: Kluwer Academic Publishers; 2004.
- Mikhailov VGG. Fant's the acoustic theory os speech production. In: XVI Session of Russian Acoustical Society; 2005 Nov 14; Leninsk Gory, Moscou. Proceedings. Moscou: RAS; 2005. p. 14-18.
- Fant G. Structural classification of Swedish phonemes. STL-QPSR. 1960;1(2):010-5.
- Stevens KN. Acoustic phonetics. 30th ed. Cambridge: MIT Press; 2000.
- Deng L, O'Shaughnessy D. Speech processing: a dynamic and optimization-oriented approach. New York: Marcel Dekker; 2003.
- Flanagan JL, Allen JB, Hansegawa-Johnson MA. Speech analysis synthesis and perception. 3rd ed. New York: Springer-Verlag; 2008.
- Bradbury J. Linear Predictive Coding [Internet]. 2000 [citado em 2014 July 30]. Disponível em: http://my.fit.edu/~vkepuska/ece5525/lpc_paper.pdf
- Filho TAG. Tecnologia assistiva para uma escola inclusiva: apropriação, demandas e perspectivas [dissertação]. Salvador: Universidade Federal da Bahia, Faculdade de Educação; 2009.
- Spinelli W, Souza MHS. Introdução à estatística. São Paulo: Ática; 1990.
- Doane DP, Seward LE. Estatística aplicada à administração e economia. São Paulo: McGraw Hill Brasil; 2014.
- Oliveira M, Pacheco V. Produção vocálica: análise acústica e Síndrome de Down. Diadorim. 2013 Dez;14:99-126.
- Behlau M. Uma análise das vogais do português falado em São Paulo: perceptual, espectrográfica, de formantes e computadorizada de frequência fundamental [dissertação]. São Paulo: Escola Paulista de Medicina; 1984.
- Brod LEM, Seara IC. As vogais orais do português brasileiro na fala infantil e adulta: uma análise comparativa. Ling Ensino. 2013 Jun;16(1):111-30.
- Hayward K. Experimental phonetics. New York: Routledge; 2000.
- Gregio FN. Configuração do trato vocal supraglótico na produção das vogais do português brasileiro: dados de imagens de ressonância magnética [dissertação]. São Paulo: Pontifícia Universidade Católica de São Paulo; 2006. (Report No. 1).

Author contributions

ALML was responsible for all phases of the research, which integrates his master's thesis; MAM was the study advisor and PPLP was the study co-advisor.