Perceptual and acoustic correlates of speech in a bilateral cochlear implant user

Aline Neves Pessoa[1,2]; Beatriz CAC Novaes[2]; Sandra Madureira[1]; Zuleica A Camargo[1]

[1] *Integrated Acoustic Analysis and Cognition Laboratory-LIAAC- LAEL-PUC-SP, BraziL*
[2] *Child Hearing Center– CeAC- DERDIC- PUC-SP, Brazil*

aline.nevespessoa@gmail.com; beatriznovaes@pucsp.br; madusali@pucsp.br; zcamargo@pucsp.br

Abstract

Acoustic and perceptual analysis procedures can be thought as useful clinical tools to investigate the speech characteristics of hearing impaired children (HIC). This research aimed at investigating vocal quality and voice dynamics and their acoustic and perceptual correlates in speech samples produced by a three- year ten-month-old male child, who uses bilateral cochlear implant (BCI). The speech samples were collected during a speech therapy session. The perceptual analysis of the vocal quality was based on the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS - Camargo & Madureira, 2008). The recorded corpus was analyzed by means of the *SG Expression Evaluator script* (Barbosa, 2009) running on the free software *Praat* v5.2.10. The measures, which were automatically extracted, comprise the fundamental frequency-f0, first f0 derivative, intensity, spectral slope and long-term mean spectrum. The findings reinforce some correlations between the acoustic and perceptual data, which are relevant to be considered in rehabilitation processes.

Index terms: Auditory Perception; Acoustic Analysis; Cochlear Implant; Speech Corpora

1. Introduction/Background:

The current challenge in terms of speech perception and production studies for cochlear implant users (CI) goes beyond the analysis of spectral discrimination and detection aspects. It involves the analysis of temporal processes and the combination of frequency, intensity and duration parameters offered by sound amplification devices [1-6, 25-27].

Such combinations may affect the oral communication conditions of CI infant users acquiring language. Taking BCI users into account, the benefits to the oral ability development, such as, the identification of the sound source, the improvement of speech perception in noisy environments, and better accuracy in the central auditory processing of acoustic events should be considered.

Perceptual and acoustic analysis procedures can be thought as useful clinical tools to investigate the speech characteristics of the hearing impaired children (HIC)[5,6,8]. The vocal quality and the vocal dynamic descriptions can provide evidence about the oral language acquisition process for this population and, especially, for the speech therapy intervention.

Hence, this research takes into account perception and speech production aspects and interactions between segmental and prosodic aspects [7,8]. It derives from clinical issues related to the treatment of HIC and who use cochlear implants (CI). The speech samples were collected during a speech therapy session [6]. The research subject is a child who uses a bilateral cochlear implant device.

It is worth pointing out that the process to obtain the samples does not involve standardized speech tasks inasmuch as it comes to pass in the therapeutic environment, characterized as semi-spontaneous productions.

The application of the Vocal Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS)[9], based on VPAS 2007, (figure 1)[10] enabled the perceptive description of prosodic elements from two modules: vocal quality and voice dynamics. The vocal quality settings are taken as the result from the combined actions of the larynx and the supralaryngeal vocal tract [10-16]. Furthermore, it aims at describing the long-term tendencies that characterize vocal quality settings which can be regarded as products of the respiratory, laryngeal/phonatory, supralaryngeal/articulatory systems and muscular tension conditions. For the voice dynamics evaluation the BP-VPAS, the model provides the possibility to judge pitch and loudness parameters, the use of pauses, speech rate and respiratory support.

Speaker: Date of recording: Judge: Recording ID:

	FIRST PASS		SECOND PASS						
	Neutral	Non-neutral	SETTING	moderate			extreme		
				1	2	3	4	5	6
A. VOCAL TRACT FEATURES									
1.Labial			Lip rounding/protrusion						
			Lip spreading						
			Labiodentalization						
			Minimized range						
			Extensive range						
2. Mandibular			Close jaw						
			Open jaw						
			Protruded jaw						
			Extensive range						
			Minimized range						
3.Lingual tip/blade			Advanced tip/blade						
			Retracted tip/blade						
4. Lingual body			Fronted tongue body						
			Backed tongue body						
			Raised tongue body						
			Lowered tongue body						
			Extensive range						
			Minimized range						
5.Pharyngeal			Pharyngeal constriction						
			Pharyngeal expansion						
6.Velopharyngeal			Audible nasal escape						
			Nasal						
			Denasal						
7. Larynx height			Raised Larynx						
			Lowered Larynx						
B. OVERALL MUSCULAR TENSION									
8. Vocal tract tension			Tense vocal tract						
			Lax vocal tract						
9. Laryngeal tension			Tense larynx						
			Lax larynx						

C. PHONATION FEATURES									
	SETTING	Present		Scalar Degree					
		Neutral	Non-neutral	Moderate			Extreme		
				1	2	3	4	5	6
10. Voicing type	Voice								
	Falsetto								
	Creak								
	Creaky								
11. Laryngeal frication	Whisper								
	Whispery								
12.Laryngeal irregularity	Harsh								
	Tremor								

		Neutral	SETTING	moderate			extreme		
				1	2	3	4	5	6
D. VOICE DYNAMICS									
13.Pitch	**Mean**		High						
			Low						
	Range		Minimised range						
			Extensive range						
	Variability		High						
			Low						
14. Loudness	**Mean**		High						
			Low						
	Range		Extensive range						
			Minimized range						
	Variability		High						
			Low						
E. TEMPORAL ORGANIZATION									
15. Continuity			Interrupted						
16. Rate			Fast						
			Slow						
F. OTHER FEATURES									
17. Respiratory support			Adequate						
			Inadequate						
18. Dyplophonia			Absent						
			Present						

Vocal Profile Analysis Scheme – VPAS (Laver, Mackenzie-Beck, 2007)

Figure 1: Vocal Profile Analysis Scheme (VPAS) [13]

From the acoustic point of view, vocal quality and voice dynamics have been analyzed according to the following parameters: fundamental frequency (f0), first f0 derivate, intensity, spectral slope, and long-term average spectrum [2-17-20].

Such correlations, described upon dynamic models and methodological procedures of Experimental Phonetics, allude to knowledge of physiological, acoustic and cognitive basis, implied in production and speech perception [8-10,16] in context composed by speakers with and without language acquisition disorders. Hence, it is claimed whether or not perceptual and acoustic correspondent features of vocal quality and vocal dynamic allow the identification of cues that indicate the oral language acquisition evolution in children that use CI. [21-23].

Furthermore, some questions are brought to light: What are the vocal quality settings presented by a HIC BIC user and how are they blended on the laryngeal, supralaryngeal and tension aspects? How are vocal dynamic aspects (pitch, loudness, pauses, speech rate and respiratory support) organized in the speech of subjects who use CI?

The purpose of this research was to describe semi-spontaneous speech samples, produced during a speech therapy session, from a subject who uses a bilateral cochlear implant so as to correlate vocal quality and voice dynamics aspects from the perceptual and acoustic descriptions.

2. Methods

For this study, speech samples from a child who uses a bilateral cochlear implant device were selected. The subject is a male child who is 3 years and 10 months old. (Table 1). The instruments used to record the samples were a unidirectional *Le son* lapel microphone and a *Sony* MD digital recorder model MZ- R70.

The edition, treatment and sample analysis processes were carried out at the Acoustic Analysis and Cognition Integrated Laboratory (LIAAC) of PUC-SP. The recordings were digitalized at the sample frequency of 22050 Hz and 16 bits with the wav extension, using the Sound Forge software (version 7.0).

Table 1- *Subject characterization*

Subject	Times	Audiological diagnose	Auditory responses Speech Terapy
M., male	On the recording day: Three-year ten-month-old male child / bilateral CI surgery: two-year one month old	Bilateral profound sensorineural hearing loss congenital. Early diagnose at the maternity ward and systematic use hearing aids since the subject was 3 months of age until the CIs implant surgery.	Minimal auditory responses for in pure tone fields with both devices: mean around 500 Hz,1KHz, 2Khz and 6KHz: 30db. Speech Therapy taking place since subject was 3 months of age, twice a week lasting 45 minutes.

The perceptual analysis was carried out with the use of the VPAS-PB [9,10] by two experienced judges.

The acoustic analysis procedures was carried out by the *SG Expression Evaluator script* [17,19] running on the software *Praat* v5.2.10. The script generates f0 (median, semi-amplitude interquartile, quantile 99.5% and skewness), first f0 derivate (mean, standard deviation (SD) and skewness); intensity (skewness), spectral slope (mean, SD and skewness) and LTAS (SD) measures [17,19].

The perceptual an acoustic results were statistically analyzed by means of Xlstat software [16]. Hierarchical and agglomerative cluster analysis and canonic correlation analysis procedures [24] were used. The hierarchical and agglomerative cluster analysis was applied intragroup (figure 2) and the canonic correlation analysis intergroups (figure 3). There were two groups under analysis: the perceptual judgments and acoustic measures results.

This research was approved by the Ethics Committee at PUC-SP (#135/2009). The recording of the *corpus* took place in a therapeutic context, in a speech therapy room.

3. Results

The hierarchical and agglomerative cluster analysis data are presented in dendrograms for the set of perceptual and acoustic data (figure 2).

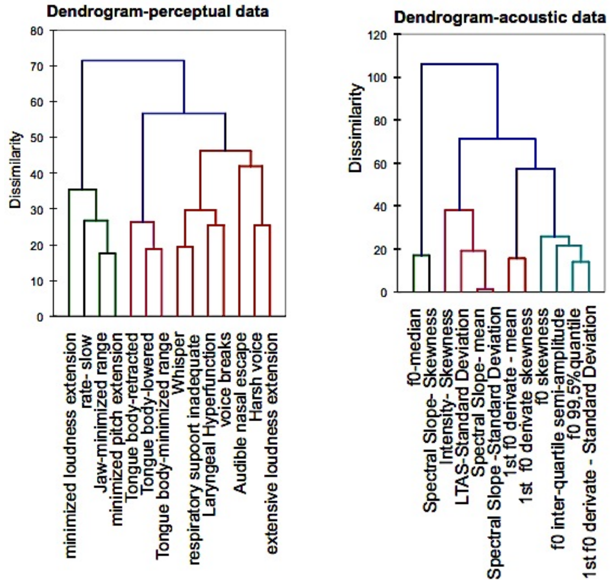

Figure 2: Dendrograms of perceptual data related to vocal quality judgements (BP-VPAS) and acoustic measures extracted by means of the Expression Evaluator script concerning the speech samples of a cochlear implant user

The cluster analysis applied to perceptual data yielded three classes: Class 1 (28,57%): minimized loudness and pitch extension, jaw mimimized range and low speech rate: Class 2 (21,42%): minimized tongue body range, lowered and retracted tongue body; Class 3 (50%): laryngeal hyperfunction, harsh voice, inadequate respiratory support, extensive loudness extension, whisper, voice breaks and audible nasal scape.

The cluster analysis applied to acoustic data yielded four classes: Class 1 (16,66%): f0 (median) and spectral slope (skewness), Class 2 (33,33%): spectral slope (mean and standard deviaton), intensity (skewness) and LTAS (standard deviation), Class 3 (16,66%): first f0 derivate (mean and skewness) and Class 4 (33,33%): f0 (interquartile semiamplitude, 99,5% quantile and skewness) and first f0 derivate measures (standard deviation) values.

The acoustic values are presented in table 2.

Table 2. *f0, First derivate, intensity, spectral slope and LTAS acoustic measures from a HIC user of bilateral IC*

Variant	Observations	Mínimum	Máximum	Mean	SD	Denormalized
f0-median	41	-0,690	1,050	0,339	0,262	271,6829268
f0 inter-quartile semi-amplitude	41	0,170	1,530	0,747	0,350	121,4941463
f0 99,5%quantile	41	0,130	1,490	1,078	0,318	
f0 skewness	41	-0,270	0,460	0,093	0,144	
1st f0 derivate - mean	41	-7,030	7,650	-0,607	2,736	-0,140177561
1st f0 derivate - SD	41	0,030	0,350	0,135	0,069	0,031269512
1st f0 derivate skewness	41	-0,860	0,950	-0,072	0,322	-0,724390244
Intensity- Skewness	41	0,140	1,190	0,580	0,272	5,8
Spectral Slope- mean	41	0,210	0,390	0,296	0,044	2,958536585
Spectral Slope -SD	41	0,240	0,410	0,321	0,040	
Spectral Slope- Skewness	41	1,110	1,460	1,238	0,081	
LTAS-SD	41	0,430	2,220	1,276	0,476	12,76341463

The circular diagram derived from the canonic correlation analysis between perceptual and acoustic data is presented in Figure 3.

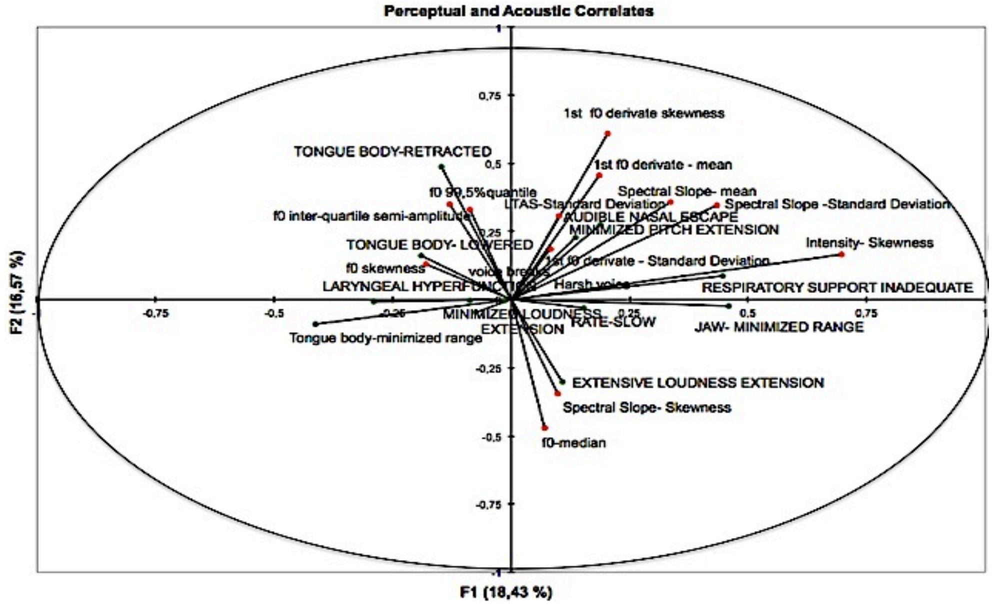

Figure 3: Circular diagrams from canonic correlation analysis: correlations between acoustic and perceptual data from a CI user

The correlations shown in Figure 3 concern the most frequent vocal quality settings and voice dynamics parameters (minimized range of tongue body, lowered tongue body, minimized range of jaw, laryngeal hyperfunction, harsh voice and minimized pitch extension) and acoustic measures.

4. Discussion and Conclusions

The perceptual data analysis allowed the observation of several levels of interaction between classes in canonic analysis (figure 3). Classes 1 and 2 in cluster analysis grouped the vocal tract settings (minimized range of tongue body and jaw associated with lowered and retracted tongue body and intermittent audible nasal escape) voice dynamics elements (pitch, loudness and speech rate) and intermittent occurrences (voice breaks). Class 3 combined muscular tension settings (laryngeal hyperfunction) and laryngeal settings, such as harsh voice, whisper and inadequate respiratory support from voice dynamics elements. These tendencies reveal the interaction between articulatory mechanisms and their interaction with laryngeal (phonatory) events in language acquisition. In articulatory arena, lingual, jaw and velopharyngeal settings were related to loudness, pitch and speech rate elements. In phonatory domain, the tension (laryngeal hyperfunction) and laryngeal (harsh voice and whisper) settings groups were found to be very productive.

These findings can be interpreted as being derived from mobilizations and adaptations of the articulators to achieve specific articulatory targets. Both language development issues and speech therapy strategies are factors which influence these speech maneuvers.

The findings concerning the voice dynamics analysis indicate minimized pitch extension, extensive loudness extension, low speech rate and inappropriate respiratory support. There were intermittent occurrences, such as vocal breaks. Moreover, there were intermittent vocal quality settings of nasal air escape.

The association of laryngeal hyperfunction to high habitual pitch may be compatible to minimized range settings in jaw and tongue [22]. Such combinations are commonly described as mechanisms yielding vocal tract and laryngeal hyperfunction, especially if conditions related to the developmental stages of the vocal apparatus are considered.

From the acoustic point of view, the canonic analysis showed tendencies of grouping classes 2 and 3 from cluster analysis, comprising spectral slope, intensity, LTAs and some of the first f0 derivate measures. In general, f0 measures grouped separately.

Taking into account the perceptual and acoustic data distribution in canonic analysis (figure 3), results from perceptual analysis in Classes 1 and 2 (figure 2) grouped with classes 2 and 3 (figure 2) from acoustic analysis.The spectral slope, intensity and LTAS measures grouped with tongue body, jaw, and laryngeal hyperfunction settings and loudness and pitch range and speech rate in voice dynamics domain. These findings reinforce the interaction between some supralaryngeal mobilizations, specially tongue body and spectral slope and LTAS measures [10,14]. In the samples analysed, these data reinforce the possibility of some vocal loading in other to achieve some articulatory targets and, again, reinforce the complex interactions between supralaryngeal settings, voice dynamics elements and spectral measures in language acquisition in the HIC children [2,5-6, 15]

Continuing to explore the interactions between perceptual and acoustic information (figure 3), data from perceptual analysis in Class 3 (figure 2) grouped with class 4 (figure 2) from acoustic analysis. The f0 measures (99,5% quantile, interquatile semi-amplitude and skewness) grouped with tension (laryngeal hyperfunction), laryngeal (harsh voice, whisper) settings and voice dynamics elements (inadequate respiratory support, extensive loudness extension and voice breaks). All these perceptual findings are physiologic compatible [12-13], meaning that the irregular pattern of vocal fold closure and vibration (harsh and whisper settings) is related to inadequate respiratory support and laryngeal hyperfunction. These findings reinforce the complex interactions between pitch control and laryngeal (harsh voice and whisper) and muscular tension settings (laryngeal hyperfunction) [4, 6, 22-25], reflected in f0 acoustic measures.

The mean negative f0 first derivate indicate a small amount of pitch variability. These data differ from findings in the phonetic literature which refer to extreme and abrupt pitch variations not only for AASI users, but also for CI users [8,22,25]. These findings indicate the influence of laryngeal hyperfunction and aperiodicity on pitch extension and variability.

Differently from other studies carried out with hearing aids and unilateral CI users [2,4,6,8,15,22], the f0 values matched the values found for male hearing children within the same age group (mean of 270 Hz) [28].

It is worth pointing out that besides the diagnosis and early intervention being important for the prognostics, specific rehabilitation procedures concerning the oral sensorimotor system, voice and speech seem to be crucial for a good oral-verbal language development and for the acoustic feedback. [3,6].

The sensitive auditory experiences provided by hearing technologies such as bilateral cochlear implant devices can foster speech perception and production links. [8-10,16].

The findings reinforce some correlations between the acoustic and perceptual data, which are relevant to be considered in rehabilitation processes.

5. Acknowledgements

We acknowledge Dr. Plínio Barbosa from UNICAMP, for the revised version of the SG Expression Evaluator Script and Dr. Luiz Carlos Rusilo for statistical advice.

This study was sponsored by a FAPESP grant (2009/10644-7).

6. References:

[1] Stuchi RF; Nascimento LT; Bevilacqua MC and Brito Neto RV. “Linguagem oral de crianças com cinco anos de uso do implante coclear”. Pró-Fono 2007 Abr-Jun;19(2):167-76.

[2] Pessoa, A.N; Novaes, B. C. A. C. and Camargo, Z. “A fala em usuários de implante coclear: evidências de análise perceptivo-auditiva e acústica”. In: 19º. Congresso Brasileiro e 8º Internacional de Fonoaudiologia, 2011, São Paulo. Rev da Sociedade Brasileira de Fonoaudiologia- Suplemento Especial. São Paulo: Sociedade Brasileira de Fonoaudiologia, 2011. v. 16. p. 1247-1247.

[3] Novaes BCAC and Mendes BC. “Habilitação Auditiva: Intervenção em Bebês e Crianças”. In: Caldas Neto, S; Mello Jr, JF; Martins, RHG; Costa, SS. [eds]. Tratado de Otorrinolaringologia - São Paulo: Edit Roca; p. 371-80, 2011.

[4] Ubrig MT, Goffi-Gomez MVS, Weber R, Menezes MHM, Nemr NK, Tsuji DH and Tsuji RK. “Voice Analysis of Postlingually Deaf Adults Pre - and Postcochlear Implantation”. Journal of Voice 2011 Nov; 25(6): 692-99.

[5] Madureira S; Barzaghi L and Mendes B. “Voicing contrasts and the deaf: production and perception issues”. In: Windsor F; Kelly ML; Hewlett N. (Org.). Investigation in Clinical Phonetics and Linguistcs. 1:417-28, 2002

[6] Pessoa NA, Pereira LK, Novaes BCAC; Camargo Z, and Mendes B. “Using Acoustic Analysis To Follow-Up The language development of a brazilian deaf child with cochlear implant”. In: Proceedings of 11th International Conference on Cochlear Implants and Other Implantable Auditory Technologies. Stockholm; Jun 30-Jul 02. p.371, 2010.

[7] Albano E, Barbosa P, Gama-Rossi A, Madureira S, and Silva A. “A interface fonética-fonologia e a interação prosódia-segmentos”. In: Estudos Linguísticos XXVII - Anais do XLV Seminário do Grupo de Estudos Linguísticos do Estado de São Paulo-GEL’97. Campinas, p.135-43, 1997.

[8] Cukier S and Camargo Z. “Abordagem da qualidade vocal em um falante com deficiência auditiva: aspectos acústicos relevantes do sinal de fala”. Revista CEFAC, Jan-Mar. 7(1): 93-101, 2005.

[9] Camargo ZA and Madureira S. “Avaliação vocal sob a perspectiva fonética: investigação preliminar”. São Paulo: Distúrbios da Comunicação, Abr. 20(1): 77-96, 2008.

[10] Camargo Z and Madureira S. “Dimensões perceptivas das alterações de qualidade vocal e suas correlações aos planos da acústica e da fisiologia”. DELTA - PUCSP, 25(2): 285-317, 2009.

[11] Laver J, Wirz SL, Mackenzie-Beck J and Hiller SM. “A perceptual protocol for the analysis of vocal profiles”. Edinburgh University Department of Linguistics Work in Progress,14: 139-155, 1981.

[12] Laver J. “The phonetic description of voice quality”. Cambridge: Cambridge University Press, 1980.

[13] Laver, J. and Mackenzie-Beck, J. “Vocal Profile Analysis scheme-VPAS”. Edinburgh: QMUC, Speech Science Research Centre; 2007.

[14] Hammaberg B. and Gauffin J. “Perceptual and acoustics characteristics of quality differences in pathological voices as related to physiological aspects”. In: Fujimura O, Hirano M. *Vocal foldphysiology*. San Diego: Singular. 283-303, 1995.

[15] Abberton E. “Voice Quality of deaf speakers”. In: Kent RD, Ball MJ. Voice Quality Measurement. San Diego: Singular. 22: 449-59, 2000.

[16] Rusilo LC, Madureira S. and Camargo Z. “Evaluating Speech samples for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS)”. In: Proceedings of the 4rd ISCA Workshop ExLing May 25-27; Paris, p.51, 2011.

[17] Barbosa PA. “Incursões em torno do ritmo da fala”. Campinas: Pontes/FAPESP, 2006.

[18] Barbosa PA. “From Syntax to acoustic duration: a dynamical model of speech rhythm production”. Oxford: Speech Communication, 2007 Sept. 49(9): 725-42.

[19] Barbosa PA. “Detecting changes in speech expressiveness in participants of a radio program In: Proceedings of Interspeech”. Brighton. p. 2155-58, 2009.

[20] Hirst D. “The analysis by synthesis of speech melody: from data to models”. Journal of Speech Sciences,1(1): 55-83, 2011.

[21] Yoshinaga-Itano C. “From Screening to Early Identification and Intervention: Discovering Predictors to Successful Outcomes for Children With Significant Hearing Loss”. J Deaf Stud Deaf Educ, Winter; 8(1): 11-30, 2003.

[22] Wirz S. “The voice of the Deaf”. In: Fawcus M (Edit). Voice Disorders and their Management. Croom Helm 1986.

[23] Tobey EA, Geers AE, Brenner CB, Altuna D. and Gabbert G. “Factors associated with development of speech production skills in children implanted by age five”. Ear & Hearing, Feb; 24(1): 36–45, 2003.

[24] Lattin, J; Carrol, D J D, and Green, P E. “Análise de dados multivariados”. São Paulo: Cegrage Learning, 2011

[25] Xu L, Zhou N, Chen X, Li, and Schultz, Z. Vocal singing by prelingually-deafened children with cochlear implant. Hearing Research, Jun. 255: 129–34, 2009.

[26] Baudonck, ED; Dhooge, I. and Lierde, KV. “Objective vocal quality in children using cochlear implants: a multiparameter approach”. J Voice, vol. 25, n 6, 2011, p. 683-691, 2011.

[27] Benninguer, MS. “Quality of the Voice Literature: What is There and What is Missing”. Journal of Voice. Nov; 25(6): 647-52, 2011.

[28] Andrade, FV. “Análise de parâmetros espectrais da voz em crianças saudáveis de 4 a 8 anos”. Mestrado em Fonoaudiologia. Universidade Veiga de Almeida, Rio de Janeiro, 2009.