

# Prosodic analysis of the speech of a child with cochlear implant

Aline Neves Pessoa-Almeida<sup>1,2</sup>; Alexsandro Meireles<sup>2</sup>; Sandra Madureira<sup>1</sup>; Zuleica Camargo<sup>1</sup>

<sup>1</sup>*Integrated Acoustic Analysis and Cognition Laboratory-LIAAC- LAEL-PUC-SP, Brazil*

<sup>2</sup>*Federal University of Espirito Santo, UFES –Brazil*

[aline.pessoa@ufes.br](mailto:aline.pessoa@ufes.br); [meirelesalex@gmail.com](mailto:meirelesalex@gmail.com); [madusali@pucsp.br](mailto:madusali@pucsp.br); [zcamargo@pucsp.br](mailto:zcamargo@pucsp.br)

## Abstract

According to previous studies [1,6], acoustic and perceptual analysis can be considered useful clinical tools to investigate the speech characteristics of hearing impaired children (HIC). This study aimed at describing voice quality settings in speech samples from a HIC wearing cochlear implants. These samples were collected during speech therapy sessions in three moments: at the time the HIC was 5 years and 1 month old, and at 6 years and 1 month and at 7 years and 1 month. The perceptual analysis of the vocal quality was based on the Vocal Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS - Camargo & Madureira, 2008). The recorded corpus was analyzed by means of the *ExpressionEvaluator script* (Barbosa, 2009) ran by *Praat* software v5.2.10. The measures, which were automatically extracted, comprised the fundamental frequency-f0, first f0 derivative, intensity, spectral slope and long-term mean spectrum. The correlations found between the acoustic and perceptual data are worth considering in rehabilitation programmes.

**Index terms:** Vocal quality; Auditory Perception; Acoustic Analysis; Cochlear Implant;

## 1. Introduction/Background:

This research focuses on the perception and production of voice quality settings and dynamic speech aspects [1-8] and it concerns the speech therapy and audiology settings since the analysis takes into account speech productions of a HIC and cochlear implant user (CI), collected during speech therapy sessions [6].

As indicated in a previous study [6], authors reinforce the demand for evaluating not only standardized speech tasks, but also semi-spontaneous speech data collected in the therapeutic environment.

Perceptual and acoustic analysis of voice quality settings, prosodic features and of temporal organization aspects are useful clinical tools to investigate the speech characteristics of HIC [5,6,8].

The description of vocal quality settings (supralaryngeal, laryngeal and tension) and voice dynamics related aspects such as pitch, loudness, continuity and rate can provide useful means for clinic evaluation and intervention on HIC and CI individuals.

Speech segmental characteristics may undergo changes under the influence of voice quality settings. These interactions between segmental and prosodic levels [7,8] must be considered in relation to the physiological, acoustic, auditory and cognitive mechanisms involved in the production and perception of speech [8-10,16].

The analysis of the perceptual and acoustic correlates of vocal quality and voice dynamics makes it possible to identify changes in speech production which demonstrate the oral language development in children who use CI. [21-23]. 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 two kinds of prosodic aspects: vocal quality settings and voice dynamics elements.

The vocal quality settings are taken as the result from the combined actions of the larynx and the supralaryngeal vocal tract [10-16]. Furthermore, such description aims at revealing the long-term tendencies that characterize vocal quality settings, which can be regarded as products of the respiratory, laryngeal/phonatory, supralaryngeal/articulatory systems, as well as muscular tension conditions. For the voice dynamics evaluation the BP-VPAS, the model provides the possibility of evaluating pitch and loudness parameters, the use of pauses, speech rate and respiratory support.

Speaker:		Date of recording:		Judge:		Recording ID:					
		FIRST PASSES		SETTINGS		SECOND PASSES					
		Neutral	Non-neutral			1	2	3	4	5	6
<b>A. VOCAL TRACT FEATURES</b>											
1. Labial				Lip rounding/protrusion							
				Lip spreading							
				Lip lateralization							
				Minimized range							
			Extensive range								
2. Mandibular				Close jaw							
				Open jaw							
				Protruded jaw							
			Extensive range								
			Minimized range								
3. Lingual tipblade				Advanced tipblade							
4. Lingual body				Retracted tipblade							
				Fronted tongue body							
				Backed tongue body							
				Raised tongue body							
			Lowered tongue body								
			Extensive range								
			Minimized range								
5. Pharyngeal				Pharyngeal constriction							
				Pharyngeal expansion							
6. Velopharyngeal				Adequate nasal escape							
				Nasal							
				Denasal							
7. Larynx height				Raised larynx							
				Lowered larynx							
<b>B. OVERALL MUSCULAR TENSION</b>											
8. Vocal tract tension				Tense vocal tract							
				Lax vocal tract							
9. Laryngeal tension				Tense larynx							
				Lax larynx							
<b>C. PHONATION FEATURES</b>											
10. Voicing type				Neutral	Non-neutral						
				Moderate	Extreme	1	2	3	4	5	6
11. Laryngeal friction				Whisper							
				Whispery							
				Harsh							
12. Laryngeal irregularity				Tremor							

		Neutral	SETTING	Moderate	Extreme				
				1	2	3	4	5	6
<b>D. PROSODIC FEATURES</b>									
13. Pitch	Mean		High						
	Range		Minimized range						
	Variability		Extensive range						
14. Loudness	Mean		High						
	Range		Extensive range						
	Variability		Minimized range						
<b>E. TEMPORAL ORGANIZATION</b>									
15. Continuity			Interrupted						
16. Rate			Fast						
			Slow						
<b>F. OTHER FEATURES</b>									
17. Respiratory support			Adequate						
			Inadequate						
18. Dysphonia			Absent						
			Present						

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].

It is very difficult to analyze speech development in hearing impaired children due to methodological difficulties and the number of variables involved: degree of hearing loss; age at the beginning of amplification; involvement of the hearing impaired children's families especially at the beginning of the therapeutical care concerning the use of hearing aids and participation of the family in the hearing rehabilitation process; and changes in articulatory and phonatory patterns.

In the speech of children, the phonation system is oscillating and, due to the non-linear relationships between the elements, it can feature patterns of great variability. A model which allows the description of articulatory, phonatory and tense settings can be applied to evaluate changes in the adopted patterns that define the maturation of the mechanism.

The phonetic model [12] of voice quality settings description used to analyze the speech samples in this study does not make a distinction between normal and pathological speech. The focus is on the articulatory, laryngeal and tense settings which interact with the speech segments affecting their quality. In this way, the inherent quality of speech segments is modified. If, for example, an [i] is produced with a lip rounding setting of voice quality, its inherent characteristic of lip spreading will be modified. Since the vocal quality settings are described according to the principles of susceptibility to the speech segments, that is, an oral speech segment, for example, will be considered more susceptible to a nasalized voice quality setting than a nasal one, it follows that settings are defined and described in relation to key speech segments.

Considering the variability of speech patterns and the complex interactions between perception and production, a phonetic model [12] which provides the description of long term muscular adjustments related to the production of voice quality settings enables the analysis and comparison of interactions between prosody and segments.

The language acquisition process is very complex and descriptions of speech characteristics based on normal parameters are not helpful since speech characteristics are not stable. From the earliest babbling—especially in children with hearing impairment—babbling may show evidences of learning the relationships between motor gestures and acoustic characteristics.

This study aimed at describing voice quality settings in speech samples from a HIC who wears cochlear implants. These samples were collected during speech therapy sessions in three moments: at the time the HIC was 5 years and 1 month old and at 6 years and 1 month and at 7 years and 1 month.

## 2. Methods

Speech samples comprised audio recordings from semi-spontaneous speech of a HI and user of CI child within the chronological age range of 5 years and 1 month and 7 years and 1 month (Table 1). The recording of the *corpus* took place in a therapeutic context, in a speech therapy room. The unit of analysis is long-term, recurring features throughout speech production. The analysis includes all long-term trends of speech production featuring a speaker in particular: the moments used speech samples analyzed are 5-15 seconds.

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) at 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	Hearing aids usage (chronological age: years; month)	IC –age at the time of surgery (chronological age: years; month)	Audio-recording sessions (chronological age/ auditory age: years; month)
Y.	0;7	2;7/ unilateral CI	Moment 1 (5;1/2;4), Moment 2 (6;1/3;4), Moment 3 (7;1/4;4)

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

The acoustic analysis was developed by using the *Expression Evaluator script* [17,19] running in the software *Praatv5.2.10*. The script generates f0 (median, interquartile semi-amplitude, 99,5% quartile and skewness), first f0 derivate (mean, standard deviation (SD) and skewness), as well as intensity (skewness), spectral slope (mean, SD and skewness) and LTAS (SD) measures [17,19].

The perceptual and acoustic results were statistically analyzed by means of *XIStat* software [16]. There were two elements under analysis: the perceptual judgments and acoustic measures results. Canonical correlation analysis between perceptual and acoustic data and the discriminant analysis of the acoustic correlates were performed taking into account the speech productions of the child at 5 years and 1 month, 6 years and 1 month and 7 years and 1 month.

It is important to highlight of methodological adequacy and consistency of the statistical analysis (canonical and discrimination analysis) adopted to consider the correlations between qualitative (perceptual evaluation of the settings) and quantitative (acoustic measures) in this study. This kind of analysis can be applied to semi-spontaneous and spontaneous speech excerpts and does not require the use of standardized speech samples

The issue of adopting relative measures instead of absolute measures allows correlation between perceptual evaluation and acoustic measures in the characterization of each speaker profile. This procedure does not require the labeling of vowel and consonantal segments, which may not be well delimited in certain children speech productions in the earlier stages of language development and in those productions considered altered for the age bracket. Even the speech of hearing children in the earlier stages of language development can not be described according to norm parameters.

This research was approved by the Ethics Committee at PUC-SP (#135/2009).

## 3. Results

The circular diagrams derived from the canonical correlation analysis between perceptual and acoustic data, considering moments 1, 2 and 3, are presented in Figures: 1, 2 and 3.

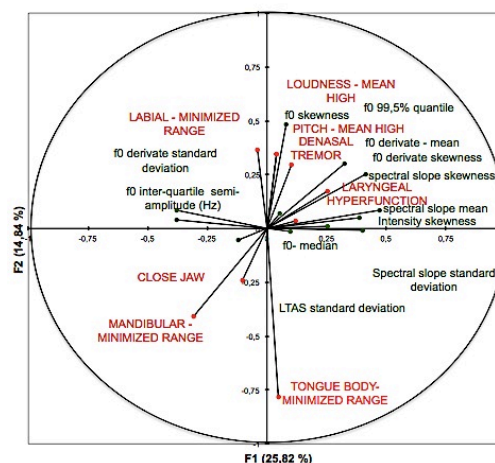


Figure 2: Circular diagrams from canonic correlation analysis: correlations between acoustic and perceptual data from a CI user in moment 1.

The correlations presented in Figure 2 comprised the most frequent vocal quality settings and voice dynamics parameters: f0-median and usual *pitch* high (90,5%) and usual *loudness mean* high (90,5%), f0-median associated with tremor (64,4%), f0-derivate skewness with labial minimized range (61%) and spectral slope - standard deviation associated with labial minimized range (50,5%).

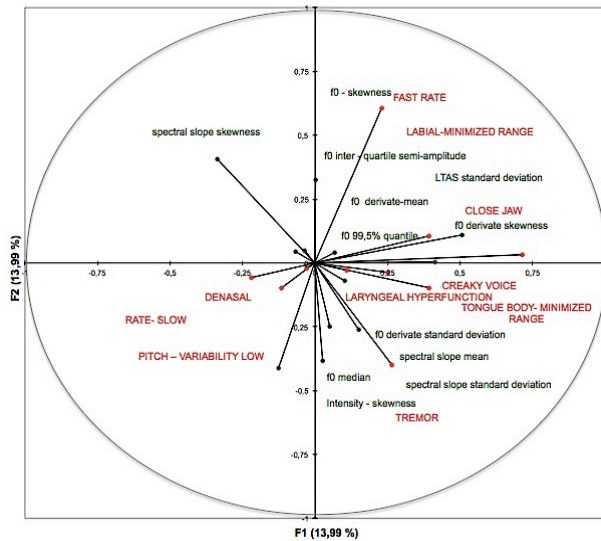


Figure 3: Circular diagrams from canonic correlation analysis: correlations between acoustic and perceptual data from a CI user in moment 2.

The correlations shown in Figure 3 concern the most frequent vocal quality settings and voice dynamics parameters: fast rate with f0-skewness (46,2%), creaky voice associated with f0-interquartile semi-amplitude (45,1%), slow rate with f0-interquartile semi-amplitude (41,5%) and tense larynx associated with intensity skewness(41,4%)

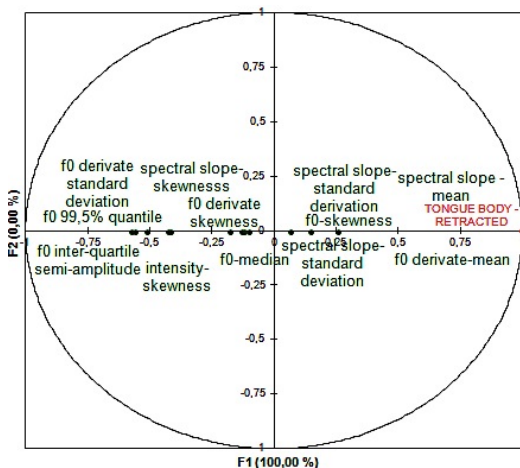


Figure 4: Circular diagrams from canonic correlation analysis: correlations between acoustic and perceptual data from a CI user in moment 3.

The correlations shown in Figure 4 revealed the most frequent vocal quality settings and voice dynamics parameters: f0-standard deviation derivative and minimized range of tongue body (93,9%) and harsh voice (93,9%) associated with falsetto and f0-median (91,3%).

#### 4. Discussion and Conclusions

The results presented in terms of perceptual and acoustic data from a cochlear implant user during a period of speech therapy 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. These tendencies reveal the interaction between articulatory mechanisms and laryngeal (phonatory) events in language acquisition. In the articulatory domain, lingual, jaw and velopharyngeal settings were related to loudness, pitch and speech rate elements. In the phonatory domain, the tension (laryngeal hyperfunction) and laryngeal (harsh voice and whisper) settings groups were found to be very productive.

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

Taking into account the perceptual and acoustic data distribution in canonical correlation analysis (figures 3, 4 and 5), the interaction between some supralaryngeal mobilizations, especially tongue body and spectral slope and LTAS measures, is highlighted[10,14]. In the analyzed samples, these data can be interpreted as revealing some vocal loading in order to achieve some articulatory targets and, again, reinforce the complex interactions between supralaryngeal settings, voice dynamics elements and spectral measures in language acquisition for HI children [2,5-6, 15]

The findings also 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 values concerning habitual f0 at moment 1 (auditory age: 2y4m) were higher than those at moment 2 (auditory age: 3y4m), whereas at moment 3 (auditory age: 4y4m) these values increased 406,4 Hz. For this speaker, the influence of acoustic data of spectral slope and f0-skewness and intensity-skewness, in relation to laryngeal hyperfunction associated with aperiodicity are noteworthy. *Pitch* and *loudness* variability decreased. The phonetic literature refers to extreme and abrupt pitch variations not only for AASI users, but also for CI users [8,22,25,28]. Such findings indicate the influence of laryngeal hyperfunction and also reinforce the complex interactions of f0 control mechanisms in HIC, particularly in terms of the association of voice quality settings -vocal tract settings (minimized range – tongue, jaw and lip), laryngeal (tremor) and vocal tract hyperfunction. The findings concerning f0 variations in the listener population speech indicate that speech intelligibility and phonological discrimination can be affected if vocal aperiodicity characteristics are present. In relation to the vocal dynamics elements, the slow rate at moment 1 turned into a fast one at the moment 3, when the analysis detected a higher influence of the retracted tongue body setting of vocal quality. From the acoustic point of view, differentiation of evolutionary stages occurred mainly due to the distribution of f0 and spectral slope measures, which are supported by descriptions of perceptual basis. The measurements of median f0 increased over time (auditory age from 2y4m to 4y4m), as expected, because body and laryngeal development would be lowering. However, the retracted tongue body setting can change raised larynx setting and interfere with the characteristic vibration of the vocal folds. In addition, the resonance imbalances widely reported in the literature [1, 2, 6, 8, 15, 22, 23, 25, 26] may be related to vocal tract adjustments reported in this speaker (body of tongue, jaw and velopharyngeal mobility). The collected data likewise strengthen aspects of reduced movement of articulators. The collection revealed by the three

samples shows the evolution of meaningful maintenance vocal tract adjustments. The discriminant analysis of the findings using the VPAS-PB, in terms of moments of the recordings (moments 1, 2 and 3), revealed segregation of the variables at the rate of 74.16%. When individually analyzed, total segregation was higher for emission moment 3, with 100%, whereas moment 2 presented 75%, and moment 1, 66.67%. The most influential factors for this differentiation referred to the combinations of minimized range of jaw (grade 1), laryngeal hyperfunction (grade 1), decreased pitch variability (grade 1), and retracted tongue body (grade 1), represented by 80,18% combined with minimized range of tongue body (grade 2), jaw (grade 2) settings and the absence of laryngeal hyperfunction, at the rate of 19,82%. The discriminant analysis of the acoustic measures by *ExpressionEvaluator Script*, in terms of moments of the recordings (1, 2, and 3), revealed total segregation of the variables at 64.52%. Considering each moment in therapy, total segregation was higher for moment 2 emissions, by 80%, whereas moment 3 presented 72,22%, and 1 moment, 37.50%. The most influential factors for such differentiation referred to the combinations of f0-derivate SD, f0-mediana and interquartile (99,5%), represented by 88,21%, combined measures off0-semi-amplitude interquartile (11,79%). The f0 values-mean shows smoother variation at moments 1 and 2 than at the third moment, in which abrupt variations were found. In moment 3, f0-semi-amplitude interquartile arises - which reflects greater change compared to times 1 and 2. At moment 3, the f0 measure at 99.5% quartile showed higher values than at the other two moments. Similar characteristics in moments 2 and 3 were found, leading us to consider that a greater improvement took place between the first and second moments. The second and the third moments were characterized by more accurate productions of speech sounds. These aspects show the dynamic nature of speech and the complex interactions that take place between segmental and prosodic elements. At moment 3, a better delimitation of prosodic groups coincided with a period of gradual refinement of the ability in terms of articulatory productions [7]. It is worth pointing out that, besides the diagnosis and early intervention being important for the prognosis [3,4,21], specific rehabilitation procedures concerning the oral sensorial-motor system, voice and speech, seem to be crucial for a good verbal-oral language development and for the acoustic feedback [1,3,6]. Such information leads to the possibility to detail the articulatory maneuvers adopted by the children in developmental language process, and enhances the reliance of therapy, which indicates probable strategies in trying to attain the acoustic-articulatory targets in speech production. The findings reinforce some correlations between the acoustic and perceptual data, which are relevant to be considered in rehabilitation processes.

## 5. Acknowledgements

We acknowledge Plínio Barbosa from UNICAMP, for the revised version of the *ExpressionEvaluator Script*. This study was sponsored by a FAPESP grant (2009/10644-7).

## 6. References:

- [1] Pessoa, A.N.; Pereira, L.K.; Camargo, Z.A.; Madureira, S.; Novaes, B.C.A.C. An analysis of voice quality and voice dynamics in the speech production of a cochlear implant user. In: 13th Meeting of the International Clinical Linguistics and Phonetics Association - ICPLA, Oslo-Norway, June 23-26, 2010, p-286
- [2] 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
- [3] Boothroyd, A. Auditory development of the hearing child. *Scandinavian Audiology*, 26(Suppl. 46), 9-16. 1997.
- [4] Flexer, C. *Facilitating Hearing and Listening in Young Children* (2nd ed.). San Diego, CA: Singular Publishing Group. 1999.
- [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 Linguistics*. 1:417-28, 2002
- [6] Pessoa, A.N.; Novaes, BCAC; Madureira, S; Camargo, Z. Perceptual and acoustic correlates of a speech in a bilateral cochlear implant user. In: *Abstract Book Speech Prosody 2012*, 6th International Conference, Qiuwu Ma, Hongwei Ding and Daniel Hirst (eds.), Tongji University Press, Shanghai, China, May 22- 25, ISBN 978-7-5608-4869-3, v2, p51-54
- [7] Albano E, Barbosa P, Gama-Rossi A, Madureira S, and Silva A. "A interface fonética-fonologia e a interação prosódica-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; 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; Madureira S. "Avaliação vocal sob a perspectiva fonética: investigação preliminar". *São Paulo: Distúrbios da Comunicação*, Abr. 20(1): 77-96, 2008.
- [10] Camargo Z; 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.; Mackenzie-Beck, J. "Vocal Profile Analysis scheme-VPAS". Edinburgh: QMUC, Speech Science Research Centre; 2007.
- [14] Hammaberg B.; Gauffin J. "Perceptual and acoustics characteristics of quality differences in pathological voices as related to physiological aspects". In: Fujimura O, Hirano M. *Vocal fold physiology*. 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.; 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: Cengage 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] Lee, KY, Tong, MC, Van Hasselt, CA. The tone production performance of children receiving cochlear implants at different ages. *Ear Hear*, v.28, n.2 Suppl, p.34S-37S, 2007.