

Caroline Nunes Rocha-Muniz<sup>1,2</sup>   
Eliane Schochat<sup>1</sup> Investigation of the neural discrimination of  
acoustic characteristics of speech sounds  
in normal-hearing individuals through  
Frequency-following Response (FFR)*Investigação da discriminação neural das  
características acústicas dos sons de fala  
em normo-ouvintes por meio do Frequency  
Following Response (FFR)*

## Keywords

Audiology  
Electrophysiology  
Auditory Pathways  
Auditory Perception  
Speech Perception

## Descritores

Audiologia  
Eletrofisiologia  
Vias Auditivas  
Percepção Auditiva  
Percepção de Fala

## Correspondence address:

Caroline Nunes Rocha-Muniz  
Universidade de São Paulo – USP  
Rua Cipotânea, 51, Cidade  
Universitária, São Paulo (SP), Brasil,  
CEP: 05360-160.  
E-mail: carolrocha@usp.br

Received: May 10, 2019

Accepted: March 12, 2020

## ABSTRACT

**Purpose:** To evaluate how the auditory pathways encode and discriminate the plosive syllables [ga], [da] and [ba] using the auditory evoked Frequency-following Response (FFR) in children with typical development. **Methods:** Twenty children aged 6-12 years were evaluated using the FFR for the [ga], [da] and [ba] stimuli. The stimuli were composed of six formants and were differentiated in the F2 to F3 transition (transient portion). The other formants were identical in the three syllables (sustained portion). The latencies of the 16 waves of the transient portion (<70ms) and of the 21 waves of the sustained portion (90-160ms) of the stimuli were analyzed in the neural responses obtained for each of the syllables. **Results:** The transient portion latencies were different in the three syllables, indicating a distinction in the acoustic characteristics of these syllables through their neural representations. In addition, the transient portion latencies progressively increased in the following order: [ga] < [da] < [ba], whereas no significant differences were observed in the sustained portion. **Conclusion:** The FFR proved to be an efficient tool to investigate the subcortical acoustic differences in speech sounds, since it demonstrated different electrophysiological responses for the three evoked syllables. Changes in latency were observed in the transient portion (consonants) but not in the sustained portion (vowels) for the three stimuli. These results indicate the neural ability to distinguish between acoustic characteristics of the [ga], [da] and [ba] stimuli.

## RESUMO

**Objetivo:** Avaliar como as vias auditivas codificam e diferenciam as sílabas plosivas [ga],[da] e [ba], por meio do potencial evocado auditivo *Frequency Following Response* (FFR), nas crianças em desenvolvimento típico. **Método:** Vinte crianças (6-12 anos) foram avaliadas por meio do FFR para estímulos [ga],[da] e [ba]. Os estímulos foram compostos por seis formantes, sendo diferenciados na transição F<sub>2</sub> e F<sub>3</sub> (porção transiente). Os demais formantes foram idênticos nas três sílabas (porção sustentada). Foram analisadas latências de 16 ondas que compõe a porção transiente do estímulo (<70ms) e latências de 21 ondas da porção sustentada (90-160ms) nas respostas neurais obtidas para cada uma das sílabas. **Resultados:** As respostas eletrofisiológicas registradas por meio do FFR demonstraram que as latências da porção transiente da resposta neural foram diferentes nas três sílabas evocadas. Além disso, os valores de latência das ondas da porção transiente foram aumentando progressivamente, sendo [ga]<[da]<[ba]. Já na porção sustentada da resposta, não houve diferenças significantes nas latências das ondas que compõe essa porção. **Conclusão:** O FFR mostrou-se uma ferramenta eficiente na investigação da discriminação subcortical de diferenças acústicas dos sons de fala, uma vez que demonstrou diferentes resposta eletrofisiológica para três sílabas evocadas. Na porção transiente (consoantes) foram observadas mudanças de latência e na porção sustentada (vogal) não houve diferenças entre as latências para os três estímulos. Esses resultados demonstram a capacidade neural de distinção entre características acústicas dos estímulos [ga],[da],[ba].

Study conducted at Departamento de Fisioterapia, Fonoaudiologia e Terapia Ocupacional, Faculdade de Medicina, Universidade de São Paulo – USP - São Paulo (SP), Brasil.

<sup>1</sup> Departamento de Fisioterapia, Fonoaudiologia e Terapia Ocupacional, Faculdade de Medicina, Universidade de São Paulo – USP - São Paulo (SP), Brasil.

<sup>2</sup> Faculdade de Ciências Médicas da Santa Casa de São Paulo - São Paulo (SP), Brasil.

**Financial support:** Fapesp – 2011/23131-8.

**Conflict of interests:** nothing to declare.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

In speech, the hearing perception of vowels can be determined by a small number of frequencies of the first formants, which reflect the resonance properties of the vocal tract<sup>(1)</sup>. Plosive or stop consonants are produced through a temporary obstruction of airflow through the vocal tract in three different phases: total obstruction of the oral cavity, pressure build-up while the oral cavity remains blocked, and sudden release of the air current causing noise, which is also called burst. The acoustic register corresponding to the airflow release refers to the source of transient noise<sup>(2,3)</sup>.

Plosives provide rich acoustic cues that serve as a basis to identify the place of articulation and voicing, such as formant transition, burst spectrum, presence or absence of aspiration, and duration of Voice Onset Time (VOT), which corresponds to the time of voicing start or attack<sup>(4)</sup>.

Studies conducted with animal models have shown that perception of this acoustic information is encoded through many levels of the auditory system and with different neural events. Both peripheral and central structures, such as the auditory nerve and fibers of the cochlear nuclei, are able to synchronize the phases (phase-locking activity) for the harmonics (integer multiples of the fundamental frequency) of a speech stimulus<sup>(5,6)</sup>. In addition, these structures as well as the rostral part of the lower colliculus also show increased activity (discharge rate) for VOT<sup>(7)</sup>.

In humans, neural synchrony in response to acoustic characteristics of speech has been measured using the Frequency-following Response (FFR), an auditory evoked potential also known as Brainstem Auditory Evoked Potential with complex or speech stimuli (BAEPc or BAEPs). This terminology has been changed since mid-2015<sup>(8)</sup> not to limit the concepts involved by this potential, such as the nature integrated (top-down and bottom-up) and related to enriching experiences and stimuli<sup>(9)</sup>.

The FFR reflects a neural response composed of several different types of cells, mainly neural cells, in the rostral portion of the brainstem.

The brainstem responds with a high level of neural synchrony and is exceptionally well tuned to the spectral and temporal characteristics of sound, including speech sounds. However, the mechanisms involved in the neural encoding accuracy of many acoustic cues in speech remain speculative.

A large number of studies have investigated how auditory brainstem potentials respond to the speech sound [da]<sup>(10)</sup>. For this research, a structure that suggests that different neural mechanisms are responsible for encoding different acoustic aspects of speech sounds was proposed<sup>(11)</sup>. Speech sounds consist of three fundamental components: pitch (a source characteristic conveyed by the fundamental frequency); formants (filter characteristics conveyed by the selective enhancement and attenuation of harmonics), and the timing of major acoustic aspects. All of these aspects are important for speech perception and, although they are simultaneously present in the speech signal and its responses, specific components of the brainstem respond separately to each of these components<sup>(4)</sup>.

In a mature auditory system, the basal region of the cochlea is more responsive to high frequencies, while its apical region

responds better to low frequencies. This tonotopic organization is preserved throughout the auditory pathway to the cortex, and it is believed that it can assist with preserving the spectral relationship in the neural activity pattern<sup>(12,13)</sup>.

Studies have shown that perception of differences between phonemes using the cortical auditory evoked potentials (for instance, [da], [ga] and [ba]) is related to the frequencies contained in the formants of the stimuli used<sup>(14-18)</sup>.

Formant transitions are one of the essential cues underlying the identification of plosive consonants<sup>(19)</sup>. Thus, an interesting way to study how neural encoding of this transition occurs in the auditory system would be to assess the stimuli that differ only in the characteristics of filter (or harmonics), as in the case of the [da], [ga] and [ba] syllables. A primary difference between these syllables is the transition of frequencies from the second to third formants (F2 and F3).

Since F2 and F3 are beyond the brainstem phase-locking capacity, differences between these spectral cues can be observed through the latencies of neural responses.

Based on the tonotopic organization of the auditory system, low-frequency sounds, located in the apical portion of the cochlea, generate responses milliseconds later compared with those generated by high-frequency sounds encoded in the basal portion of the cochlea. Thus, the response time to high-frequency stimuli could have lower latency responses than those to low-frequency stimuli. This progression of latency as a function of frequency has been demonstrated in auditory brainstem responses to pure tone stimuli<sup>(20)</sup>.

Thus, the investigation of neural encoding for the distinctive features of the [da], [ga] and [ba] syllables, which occurs in the transition from the F2 formant, through the FFR can assist with assessing the neural encoding for the formants and understanding the processes that underlie the neural differentiation of acoustic contrasts of different speech stimuli, such as plosive consonants.

Aiming to expand knowledge on neural discrimination between different acoustic characteristics, this study assessed how the auditory pathways encode and differentiate the plosive consonant-vowel syllables [ga], [da], and [ba], presented through speech stimuli, using the FFR in children with typical development.

The following hypotheses were considered:

- (1) Because of the tonotopic organization of the auditory system, which promotes faster encoding of high frequencies, the differing F2 and F3 frequencies of the formants of the presented stimuli should manifest themselves as latency shifts. Thus, this change should occur through progressive increase in the latency of the [ga], [da] and [ba] responses (that is, [ga] <[da] <[ba]) as a result of neural synchrony;
- (2) Latency differences should decrease throughout the response until they disappear by the time the three syllables reach their steady state;
- (3) There should be no differences between the latencies of the electrophysiological responses for the three stimuli in the sustained portion.

## METHODS

This study was approved by the Research Ethics Committee of the University of São Paulo Medical School (FMUSP) under protocol no. 109/12. The parents and/or legal guardians of the participating children were informed about the procedures and signed an Informed Consent Form (ICF) prior to study commencement.

### Study sample

The study sample was composed of 20 children with typical development (according to information obtained through interviews with the teachers and parents and/or legal guardians of the participating children), absence of neurological, cognitive and psychiatric disorders, school complaints, and speech and language impairments.

All participants presented thresholds within the normal range ( $\leq 15$  dB HL) for the assessed frequencies (500-4000 Hz), speech recognition with scores  $>90\%$ , normal tympanometric measures, and BAEP with click stimulus within normality. These children also had normal performance in the auditory processing assessment. Changes in auditory processing were ruled out following the criteria recommended by the AAA<sup>(21)</sup> and ASHA<sup>(22)</sup> through the use of a monotic test, a dichotic test, and two temporal tests. If changes associated with auditory, neurological, cognitive or psychiatric aspects were verified, individuals would be excluded from the study and referred to specialized service.

### Stimuli and response capture parameters

The FFR was obtained through the presentation of acoustic speech stimuli - plosive consonant-vowel syllables [da], [ga], and [ba]. The speech stimuli were synthesized<sup>(23)</sup> at 20 kHz frequency, 16-bit resolution, and 170 ms duration. The stimuli were composed of six formants, differentiated in the onset frequencies (initial portion of the stimulus), in the transition from the second (F<sub>2</sub>) to the third (F<sub>3</sub>) formants (Table 1). These stimuli were the same used by Johnson et al.<sup>(4)</sup>

### Procedures

The stimuli were presented only to the right ear using an electroneuromyograph (SmartEP model) equipped with the cABR module (Intelligent Hearing Systems, Miami, FL, USA) at a speed of 4.35 stimuli/sec and an intensity of 80 dBnHL.

The electrophysiological responses generated by the [da], [ga] and [ba] stimuli were processed with a 50-3000Hz filter (70-2000 Hz offline filter). The artifact rejection criterion was  $\pm 35 \mu V$ .

The FFR was captured through surface electrodes in the positions Cz, M2 (right mastoid), and Fpz as ground with an analysis window of 230 ms (45 and 185 ms corresponding to the pre- and post-stimulus periods).

Two 2000-stimulus scans were performed for each syllable presented with alternating polarity. The two waves generated by the scans were calculated by weighted sum and the resulting final wave, with 4000 stimuli, was analyzed.

### Analysis of the responses

#### Formant transition period

The formant transition period was defined as the portion of the response corresponding to the onset and formant transition periods of the stimuli (0-50ms). Based on the first hypothesis, latency differences between the stimuli were expected in this portion of the response.

In this portion, a total of 16 transient peaks were recorded, with six positive peaks and 10 negative peaks (Figure 1) in the initial 70 ms of the electrophysiological response.

#### Sustained response period

Sustained response was defined as the portion corresponding to the steady part of the stimulus (51-170 ms).

In this portion, a total of 21 transient peaks were recorded, with seven positive peaks and 14 negative peaks between 90 and 160 ms of the electrophysiological response.

### Data analysis

#### Qualitative analysis

The data were quantitatively analyzed using the Cross-phaseogram technique<sup>(24)</sup>. This technique calculates the wave phase differences between two electrophysiological responses as a function of time and frequency and illustrates the differences in the transient portion in [ga] vs. [ba], [da] vs. [ba] and [ga] vs. [da] comparisons.

When the response to the [ga] stimulus leads in phase relative to that to the [ba] stimulus, the graphical representation

**Table 1.** Values (in Hz) of the fundamental frequency and the six formant frequencies of each stimulus

	Stimulus Formant Frequencies										
	F <sub>0</sub>		F <sub>1</sub>		F <sub>2</sub>		F <sub>3</sub>		F <sub>4</sub>	F <sub>5</sub>	F <sub>6</sub>
	Flat	Onset	Steady-state	Onset	Steady-state I	Onset	Steady-state	Flat	Flat	Flat	
[ga]	100	400	720	<b>3000</b>	1240	<b>3100</b>	2500	3300	3750	4900	
[da]	100	400	720	<b>1700</b>	1240	<b>2580</b>	2500	3300	3750	4900	
[ba]	100	400	720	<b>900</b>	1240	<b>2400</b>	2500	3300	3750	4900	

Adapted from Johnson et al.<sup>(4)</sup>

consists of shades of yellow, orange and red, with the largest differences represented in dark red. If the opposite occurs, i.e., the [ba] response leads the [ga] response, the representation is in blue shades, and when there is no difference between the phases, the plot appears green.

### Quantitative analysis

A grand mean (GM) latency for the peaks obtained across the three stimuli was computed to normalize these values for all peaks - 16 in the transient portion and 21 in the sustained portion - so that they could be described on the same scale. After that, this GM was subtracted from each individual peak latency ( $\text{Latency}_{\text{Individual}} - \text{Latency}_{\text{GM}}$ ). Thus, earlier peaks are negative numbers, later peaks are positive numbers, and peaks near the GM are close to zero<sup>(4)</sup>.

Multivariate analysis of variance with repeated measures (repeated-measures MANOVA) was performed to compare the test averages across the three studied stimuli<sup>(25)</sup>. In the repeated-measures MANOVA, the *p*-value and the *F* ratio, which is used to test the global difference between groups, were analyzed using the Wilks' Lambda ( $\lambda$ ) test.

To complement the descriptive analysis, confidence interval (CI) was used to assess the extent to which the average could vary at a certain level of confidence. The CI established for data analysis was 95%, with a significance level of 0.05 (5%).

## RESULTS

The latency peaks resulting from the stimuli with the [da], [ga] and [ba] syllables were analyzed according to the transient (16 peaks) and sustained (21 peaks) portions.

It was hypothesized that during the formant transition period (transient portion) the difference in frequencies of the F2 and F3 formants would be shown in the latencies of their electrophysiological responses, with responses to the stimuli progressively increasing their latency values ([ga] < [da] < [ba]) due to the progression between the frequency differences of the sounds. As for the sustained portion of the stimulus (90-170 ms), it was hypothesized that there would be no significant

differences in the electrophysiological responses between the three stimuli used.

In the qualitative analysis carried out using the Cross-phaseogram<sup>(24)</sup> technique, it was found that the greatest discrimination occurred between [ga] and [ba], followed by between [da] and [ba] (Figure 2). This discrimination is represented in shades of yellow, orange, and red. Smaller discrimination occurred for the pair with the least difference, that is, [ga] and [da], represented in Figure 2 by most predominant green. The differences occurred only in the transient portion, with difference between F2 and F3 for the three syllables (10-50 ms). In the sustained portion (similar in the three syllables), no difference between the response phases was observed, shown in green in Figure 2.

Since no differences between the stimuli were identified in the sustained portion, statistical analysis was performed only in the transient portion.

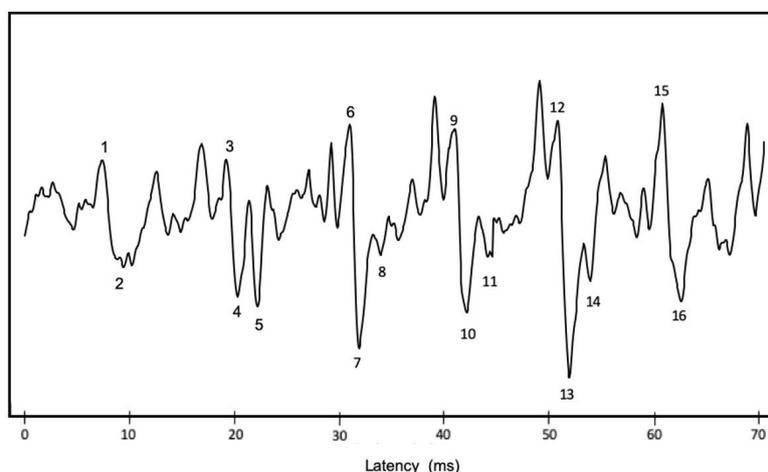
The latencies of the 16 waves that compose the transient portion (0-70 ms) in each of the stimuli were analyzed. Table 2 shows the descriptive analysis of the latency measures of the 16 waves of all children with typical development. Since some individuals did not present all 16 waves, the acronym 'N' was inserted to specify the number of participants who presented that wave and, consequently, the number of participants used for the other analyses.

Figure 3 shows the GM of the electrophysiological responses obtained by the 20 individuals for the three stimuli ([ga], [da], and [ba]).

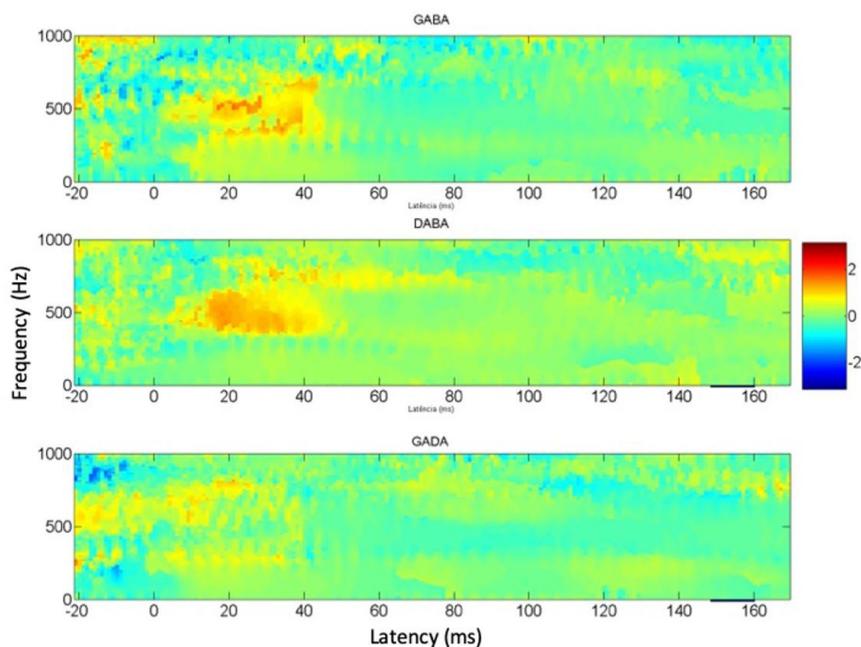
Figure 4 shows the result of the GM subtracted from each of the 16 peaks of the transient portion ( $\text{Latency}_{\text{Individual}} - \text{Latency}_{\text{GM}}$ ). The earlier peaks are negative numbers [ga], later peaks are positive numbers [ba], and peaks near the GM are close to zero [da].

Figure 5 shows the CIs for the 16 waves of the transient portion.

Repeated-measures MANOVA was conducted to determine whether there were differences between the three syllables studied. Results of the analyses were divided into four parts: a) latency of the onset peaks (1,2); b) latency of the major



**Figure 1.** Representation of the 16 waves recorded in the transient portion of the electrophysiological response generated by the stimulus /da/



**Figure 2.** Difference in average responses for the syllables [ga] and [ba]; [da] and [ba]; [ga] and [da] in the 20 children evaluated using the Cross-phaseogram analysis technique

**Table 2.** Descriptive values of the absolute latencies of each peak of the FFR to all stimuli

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
		Peak latency (ms)															
[ga]	Mean	8.10	9.00	21.27	22.48	24.26	31.53	32.70	34.50	41.79	43.11	44.58	51.75	53.19	54.82	62.10	63.39
	SD	0.43	0.78	0.49	0.66	0.79	0.45	0.56	0.69	0.50	0.71	0.84	0.66	1.21	0.90	0.82	0.97
	N	20	19	20	20	18	20	20	14	20	20	15	20	20	15	20	20
[da]	Mean	8.03	9.25	21.78	22.81	24.57	31.78	32.96	34.90	41.72	42.91	44.78	51.83	53.12	55.02	61.94	63.23
	SD	0.38	0.59	0.82	0.82	1.19	0.51	0.59	0.87	0.51	0.55	0.93	0.84	0.97	0.91	0.97	1.10
	N	20	20	20	20	17	20	20	16	20	20	15	20	20	18	20	20
[ba]	Mean	8.24	9.57	22.04	23.19	25.07	32.21	33.43	35.23	42.00	43.31	45.24	52.00	53.24	55.28	62.01	63.39
	SD	0.50	0.49	0.99	0.86	0.91	0.47	0.47	0.85	0.48	0.56	0.81	1.00	1.06	1.10	0.98	1.09
	N	20	20	18	19	15	20	20	13	20	20	19	20	20	17	20	20

SD - Standard Desviation; N: Number of children

peaks (3,4,6,7,9,10,12,13,15,16); c) latency of the minor peaks (5,8,11,14); latency of the end-point peaks (15,16).

### Analysis of onset peak latencies

Repeated-measures MANOVA showed no multivariate difference in latency measures between electrophysiological responses to the [da], [ga] and [ba] stimuli [ $F(16.4)=1.90$ ;  $p=0.16$ ].

### Analysis of major peak latencies

Repeated-measures MANOVA showed multivariate difference in latency measures between the electrophysiological responses to the [da], [ga] and [ba] stimuli [ $F(16.4)=92.05$ ;  $p<0.001$ ; partial  $\eta^2=0.99$ ; Wilks  $\lambda=0.99$ ].

Univariate analysis was used to evaluate the relative contributions of each latency measure of the analyzed waves in the difference found. These analyses indicated statistically significant differences between the three stimuli studied for

the following waves: 3 ( $p<0.001$ ), 4 ( $p<0.001$ ), 6 ( $p<0.001$ ), 7 ( $p<0.001$ ), 9 ( $p<0.001$ ), and 10 ( $p<0.001$ ).

For paired comparison, the paired  $t$ -test was applied to verify the differences between the stimuli (Table 3).

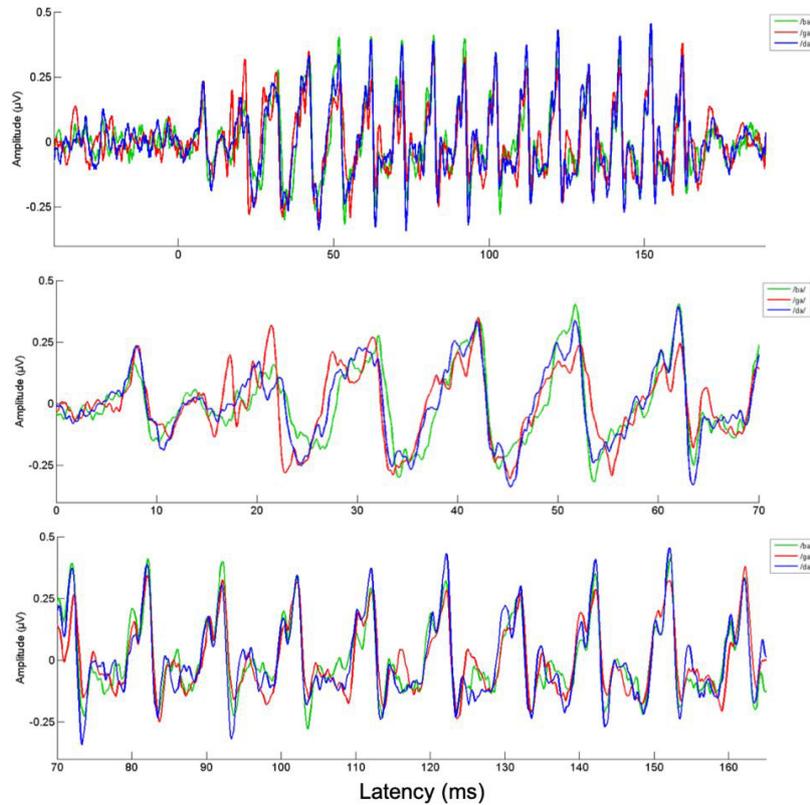
### Analysis of minor peak latencies

Repeated-measures MANOVA could not be used to assess the latency values of minor peaks since the number of absences found in these waves (Table 2) hindered the application of this type of analysis. This finding demonstrates greater inconstancy of these waves compared with peaks of the major, onset and end-point waves.

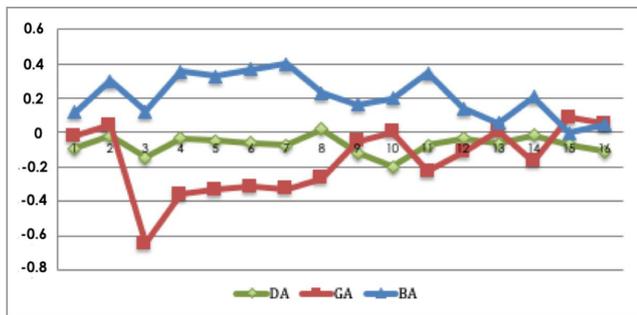
Thus, only the paired  $t$ -test was used to verify the differences between the stimuli (Table 4).

### Analysis of end-point peak latencies

Repeated-measures MANOVA showed multivariate difference in latency measures between the electrophysiological responses



**Figure 3.** (A) Average of the electrophysiological responses obtained by the FFR with the /ga/ (green), /da/ (red), and /ba/ (blue) stimuli in the 20 participants; (B) Transient portion of the electrophysiological response; (C) Sustained portion of the electrophysiological response



**Figure 4.** Grand mean (GM) subtracted from each of the 16 peaks of the transient portion ( $\text{Latency}_{\text{Individual}} - \text{Latency}_{\text{GM}}$ ) for the three studied stimuli

to the [da], [ga] and [ba] stimuli [ $F(16.4)=3.37; p=0.035$ ; partial  $\eta^2=0.45$ ; Wilks  $\lambda=0.54$ ]

For paired comparison, the paired *t*-test was applied to verify the differences between the stimuli (Table 5).

## DISCUSSION

Due to the importance of neural processing in the transition of acoustic elements overtime for the integrity of speech processing, there is great interest in understanding how the central auditory system encodes this information in a normal auditory nervous system so that what occurs when this encoding is broken, or is still under development, can be understood.

This study aimed to understand how the auditory pathways located in the brainstem reflect subtle acoustic differences

existing between the plosive consonant-vowel syllables [ga], [da] and [ba], which differ only in the transition from the  $F_2$  to  $F_3$  frequencies.

The results confirmed the first hypothesis of this study, that is, the differing  $F_2$  and  $F_3$  frequencies were manifested in the neural processing of the acoustic characteristics of the studied stimuli. In other words, changes in the latency time of the electrophysiological response have been demonstrated with a progressive increase in the response latency for the [ga], [da] and [ba] stimuli (i.e.,  $[ga] < [da] < [ba]$ ).

This difference between latency period and the stimuli was evident mainly for the latencies of the major and minor peaks. However, it was noticed that the major peaks had a clearer and more steady morphology and were present in all participants, unlike the minor peaks (Table 2).

According to the theory presented by Johnson et al.<sup>(4)</sup>, this distinction in the responses between major and minor peaks supports the idea that separate neural mechanisms are responsible for encoding different acoustic aspects of speech sounds. The major peaks would represent the fundamental frequency ( $F_0$ ) and correspond to the glottal pulse in the stimulus, thus transmitting information about the pitch. In contrast, the minor peak latencies reflect the stimulus transition formants, which vary between the [ga], [da] and [ba] syllables and are expressed in the time domain in the electrophysiological response, because variation of these frequencies is beyond the phase-locking capacity of the auditory system.

**Table 3.** Paired *t*-test results for each stimulus contrast for the major peaks

Waves	<i>p</i> -value					
	/ga/ vs. /ba/		/da/ vs. /ba/		/ga/ vs. /da/	
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
3	-4.54	<0.001*	-1.49	0.154	4.35	<0.001*
4	-4.71	<0.001*	-2.40	0.002*	3.17	0.005*
6	-8.852	<0.001*	-6.981	<0.001*	0.432	0.007*
7	-8.713	<0.001*	-3.708	0.001*	2.358	0.03*
9	-3.203	0.005*	-5.653	<0.001*	-1.172	0.26*
10	-1.478	0.16	-4.693	<0.001*	-1.803	0.09#
12	-1.355	0.19	-2.492	0.02*	0.481	0.64
13	-0.223	0.83	-1.39	0.18	-0.302	0.77

\*Statistically significant; #Marginally significant; *t* = *t*-value; *p*= *p*-value

**Table 4.** Paired *t*-test results for each stimulus contrast for the minor peaks

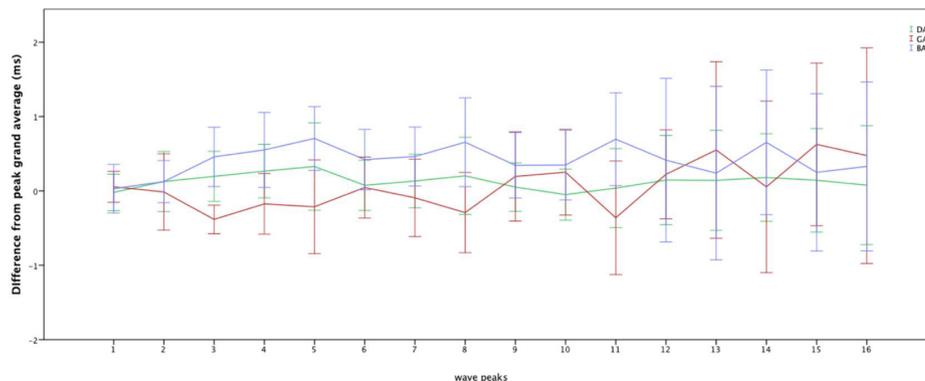
Waves	<i>p</i> -value					
	/ga/ vs. /ba/		/da/ vs. /ba/		/ga/ vs. /da/	
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
5	-4.71	<0.001*	-2.40	0.027*	3.17	0.005*
8	-5.01	0.001*	-1.42	0.19	1.74	0.1
11	-4.45	0.001*	-2.51	0.026*	0.88	0.40
14	-2.73	0.019*	-3.34	0.005*	1.91	0.08

\*Statistically significant; *t* = *t*-value; *p*= *p*-value

**Table 5.** Paired *t*-test for each contrast between the stimuli at the end-point peaks

Waves	<i>p</i> -value					
	/ga/ vs. /ba/		/da/ vs. /ba/		/ga/ vs. /da/	
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
15	1.68	0.11	-2.38	0.03*	-2.811	0.01*
16	0.03	0.97	-2.33	0.03*	-2.58	0.02*

\*Statistically significant; *t* = *t*-value; *p*= *p*-value

**Figure 5.** Confidence interval (95% CI) of normalized values for each of the 16 peaks for the three studied stimuli

Since major peaks reflect the stimulus  $F_0$ , it would be expected that these peaks be identical across the neural responses obtained in all the syllables used in this study; however, differences between latencies were also observed for the major peaks.

One hypothesis for this observed difference is that the major peaks are influenced by the patterns observed in the minor peaks. Another factor to be considered would be that, in natural articulations, pitch perturbations caused by articulatory movements in the vocal tract could be present. In this case, the systematic pattern observed in the minor peaks could also be evidenced in the major peaks.

Since the minor peak latencies are neural representations of the formants of these stimuli, the smallest difference found between the [ga] and [da] electrophysiological responses suggests a similar neural encoding between the stimuli. The difference between the formants of the [ga] and [da] acoustic stimuli is more discreet compared with that between [ga] and [ba], which is the pair with the greatest distinction in their formants. Thus, neural encoding through the latency measures of the 16 peaks of the transient portion showed that the [da] and [ga] electrophysiological representation is more similar in  $F_2$  and  $F_3$  and that [ga] and [ba] shows a greater difference

between  $F_2$  and  $F_3$ . These contrasts in the neural representation of electrophysiological responses were also verified through the Cross-phaseogram analysis, as shown in Figure 2.

Therefore, the presented results corroborate the hypothesis of Johnson et al.<sup>(4)</sup> and Hornickel et al.<sup>(26)</sup>, demonstrating that neural encoding for the different acoustic elements is manifested in a different, independent way and can be studied through the FFR.

The second hypothesis of this study was also confirmed. Table 2 and Figure 3 show that the difference between the mean latencies obtained across the three stimuli decreased during the response, until disappearing when the three syllables reach their steady state (vowel) (Figure 2).

However, no statistically significant differences were observed between the mean latencies in the initial or onset part of the response (waves 1 and 2). These findings corroborate those reported by Johnson et al.<sup>(4)</sup>, who related the neural response onset to the initial burst of the voiced plosive syllable stimulus, similar in the three studied stimuli.

The third hypothesis of this study has also been demonstrated, since no differences between the latencies of the electrophysiological responses were observed across the three stimuli in the sustained portion (vowel). This result was already expected since the acoustic properties in this portion are identical across the studied stimuli.

Thus, the different electrophysiological representations of the acoustic characteristics of the transient and sustained portions of the speech stimuli in the brainstem in children with typical development show that different neural mechanisms, mediated by neural synchrony or phase-locking, have separately encoded these acoustic cues.

This study contributes to the understanding of the subcortical neural mechanisms that underlie formant transition encoding. The results showed that the electrophysiological responses in the first 70 ms of responses were responsible for differentiating between spectral cues that assist with distinguishing between consonants. This suggests that different neurons have specific responses to different acoustic aspects, that is, high-frequency stimuli present earlier latency responses than low-frequency stimuli. This progression in latency time as a function of frequency has already been demonstrated in the pure-tone brainstem response<sup>(20)</sup>. In a mature auditory system, the basal region of the cochlea is more responsive to high frequencies and the apical region is more responsive to low frequencies. This tonotopic organization is preserved along with the neural auditory pathways, which would assist with preserving spectral information in neural encoding activity<sup>(12,13)</sup>.

Although this study has contributed new information regarding the representation of transient and sustained acoustic cues in the subcortical auditory pathways, there is still much to be investigated. Regarding the normal encoding of acoustic characteristics, it is hoped that future studies will add a wider repertoire of syllables, including consonants with different places of articulation.

Finally, it is believed that the FFR with speech stimuli (or other complex stimuli), together with other measures and clinical assessments, can inform processes that underlie the biological nature of auditory processing and speech and language changes,

assist with therapeutic strategies, and promote an objective index of therapeutic evolution. For example, some populations may present deficits in neural encoding for specific elements of onset and/or end-point, or specific for formant rapid transition encoding. In contrast, there is also a possibility that some children present deficits in neural encoding, both in transient and sustained information.

Thus, the results of this study enable perception that such populations could be more precisely identified and that more accurate therapeutic programs and strategies could be developed to suit the specific area of difficulty.

## CONCLUSION

The Frequency-following Response (FFR) proved to be an efficient tool to investigate the subcortical discrimination of acoustic differences in speech sounds, since the data demonstrate that the electrophysiological responses present differences relevant to each of the three evoked syllables. In the transient portion (consonants), latency shifts were observed, whereas no differences between the latencies across the three stimuli were found in the sustained portion (vowel). In other words, different neural representations for the different acoustic characteristics of the [ga], [da] and [ba] syllables could be observed.

Considering the existing knowledge on the encoding of acoustic characteristics of speech sounds, these data assist with understanding how the brainstem encodes the important perceptual differences in speech through the FFR. It is believed that this study has significance in expanding the knowledge on how the neural encoding of these acoustic differences occurs in clinical populations.

## ACKNOWLEDGEMENTS

The authors are grateful to the Sao Paulo Research Foundation (FAPESP) for the funding provided to this study (process number 2011/23131-8).

## REFERENCES

1. Hillenbrand J, Gayvert RT. Vowel classification based on fundamental frequency and formant frequencies. *J Speech Hear Res.* 1993;36(4):694-700. <http://dx.doi.org/10.1044/jshr.3604.694>. PMID:8377482.
2. Ladefoged P, Maddieson I. *The sounds of the world's languages.* Oxford: Blackwell. 1996.
3. Johnson K. *Acoustic and auditory phonetics.* Malden, MA: Blackwell; 2003.
4. Johnson KL, Nicol T, Zecker SG, Bradlow AR, Skoe E, Kraus N. Brainstem encoding of voiced consonant-vowel stop syllables. *Clin Neurophysiol.* 2008;119(11):2623-35. <http://dx.doi.org/10.1016/j.clinph.2008.07.277>. PMID:18818121.
5. Sachs MB, Young ED. Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *J Acoust Soc Am.* 1979;66(2):470-9. <http://dx.doi.org/10.1121/1.383098>. PMID:512208.
6. Young ED, Sachs MB. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *J Acoust Soc Am.* 1979;66(5):1381-403. <http://dx.doi.org/10.1121/1.383532>. PMID:500976.

7. Chen GD, Nuding SC, Narayn SS, Sinex DG. Responses of single neurons in the chinchilla inferior colliculus to consonant-vowel syllables differing in voice-onset time. *Aud Neurosci*. 1996;3:179-98.
8. White-Schwoch T, Woodruff Carr K, Thompson EC, Anderson S, Nicol T, Bradlow AR, et al. Auditory processing in noise: a preschool biomarker for literacy. *PLoS Biol*. 2015;13(7):e1002196. <http://dx.doi.org/10.1371/journal.pbio.1002196>. PMID:26172057.
9. Kraus N, White-Schwoch T. Unraveling the biology of auditory learning: a cognitive-sensorimotor-reward framework. *Trends Cogn Sci*. 2015;19(11):642-54. <http://dx.doi.org/10.1016/j.tics.2015.08.017>. PMID:26454481.
10. Banai K, Hornickel J, Skoe E, Nicol T, Zecker SG, Kraus N. Reading and subcortical auditory function. *Cereb Cortex*. 2009;19(11):2699-707. <http://dx.doi.org/10.1093/cercor/bhp024>. PMID:19293398.
11. Johnson KL, Nicol T, Kraus N. Brain stem response to speech: a biological marker of auditory processing. *Ear Hear*. 2005;26(5):424-34. <http://dx.doi.org/10.1097/01.aud.0000179687.71662.6e>. PMID:16230893.
12. Langner G. Neural processing and representation of periodicity pitch. *Acta Otolaryngol Suppl*. 1997;532(sup532):68-76. <http://dx.doi.org/10.3109/00016489709126147>. PMID:9442847.
13. Merzenich MM, Reid MD. Representation of the cochlea within the inferior colliculus of the cat. *Brain Res*. 1974;77(3):397-415. [http://dx.doi.org/10.1016/0006-8993\(74\)90630-1](http://dx.doi.org/10.1016/0006-8993(74)90630-1). PMID:4854119.
14. McGee T, Kraus N, King C, Nicol T, Carrell TD. Acoustic elements of speech like stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. *J Acoust Soc Am*. 1996;99(6):3606-14. <http://dx.doi.org/10.1121/1.414958>. PMID:8655792.
15. Sharma A, Dorman M. Cortical Auditory evoked potential correlates of categorical perception of voice-onset time. *J Acoust Soc Am*. 1999;106(2):1078-83. <http://dx.doi.org/10.1121/1.428048>. PMID:10462812.
16. Tremblay K, Piskosz M, Souza P. Effects of age and age related hearing loss on the neural representation of speech cues. *Clin Neurophysiol*. 2003;114(7):1332-43. [http://dx.doi.org/10.1016/S1388-2457\(03\)00114-7](http://dx.doi.org/10.1016/S1388-2457(03)00114-7). PMID:12842732.
17. Korczak P, Stapells DR. Effects of various articulatory features of speech on cortical event-related potentials and behavioral measures of speech-sound processing. *Ear Hear*. 2010;31(4):491-504. <http://dx.doi.org/10.1097/AUD.0b013e3181d8683d>. PMID:20453651.
18. Elangovan S, Stuart A. A cross-linguistic examination of cortical auditory evoked potentials for categorical voicing contrast. *Neurosci Lett*. 2011;490(2):140-4. <http://dx.doi.org/10.1016/j.neulet.2010.12.044>. PMID:21193015.
19. Blumstein SE, Isaacs E, Mertus J. The role of the gross spectral shape as a perceptual cue to place articulation in initial stop consonants. *J Acoust Soc Am*. 1982;72(1):43-50. <http://dx.doi.org/10.1121/1.388023>. PMID:7108042.
20. Gorga M, Abbas P, Worthington D. Stimulus calibration in ABR measurements. In Jacobsen J, editor. *The auditory brainstem response*. San Diego: College-Hill Press; 1985. p. 49-62.
21. AAA: American Academy of Audiology. Diagnosis, treatment, and management of children and adults with central auditory processing disorder [Internet]. Reston: AAA; 2010 [citado em 2019 Maio 10]. Disponível em: <https://www.audiology.org/publications-resources/document-library/central-auditory-processing-disorder>
22. ASHA: American Speech and Hearing Association. (Central) auditory processing disorders. Technical report [Internet]. Washington: ASHA; 2005 [citado em 2019 Maio 10]. Disponível em: <https://www.asha.org/policy/TR2005-00043/>
23. Klatt DH. Software for a cascade/parallel formant synthesizer. *J Acoust Soc Am*. 1980;67(3):971-95. <http://dx.doi.org/10.1121/1.383940>.
24. Skoe E, Nicol T, Kraus N. Cross-phaseogram: objective neural index of speech sound differentiation. *J Neurosci Methods*. 2011;196(2):308-17. <http://dx.doi.org/10.1016/j.jneumeth.2011.01.020>. PMID:21277896.
25. Dancey CP, Reidy J. *Estatística sem matemática para psicologia*. Porto Alegre: Artmed; 2006.
26. Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proc Natl Acad Sci USA*. 2009;106(31):13022-7. <http://dx.doi.org/10.1073/pnas.0901123106>. PMID:19617560.

### Author contributions

*CNRM: proposition of the original idea, collection and analysis of the data, literature review, writing and review of the manuscript; ES: proposition of the original idea, coordination of the study, review of the manuscript.*