Original articles

Forward masking with frequency-following response analyses

Silvana Maria Sobral Griz[1]
https://orcid.org/0000-0002-6759-5964

Denise Costa Menezes[1]
https://orcid.org/0000-0002-2008-4091

Karina Paes Advíncula[1]
https://orcid.org/0000-0002-1314-6100

Maria Aline de Lucena Lima[1]
https://orcid.org/0000-0002-2424-0928

Pedro de Lemos Menezes[2]
https://orcid.org/0000-0003-1999-5055

[1] Universidade Federal de Pernambuco –
UFPE, Departamento de Fonoaudiologia,
Recife, Pernambuco, Brasil.

[2] Universidade Estadual de Ciências
da Saúde de Alagoas – UNCISAL,
Departamento de Fonoaudiologia,
Alagoas, Pernambuco, Brasil.

Conflict of interests: Nonexistent

ABSTRACT

Purpose: to analyze forward masking in normally hearing young people, by using frequency-following responses.

Methods: the synthetic syllable /da/ was used for the recordings of ten individuals, in the following conditions: /da/ with no masking, and /da/ after 4, 16, 32, and 64 milliseconds of masking. F-test (ANOVA) was applied for repeated measures with the Greenhouse-Geisser correction to compare testing conditions. For significant differences, multiple comparisons (between pairs of conditions) and Bonferroni correction were used. Data normality was verified by applying the Shapiro-Wilk test, and statistical significance was used at 5%.

Results: wave latencies of all masking conditions were compared with those of no masking. A latency delay was observed in the transient region of the response (PV and A) in all masking conditions, except for 64 milliseconds. Latency delay also occurred for waves PW, PX, and PY, which corresponded to the sustained region of the response.

Conclusion: forward masking was observed, by using frequency-following responses with /da/ syllable in four intervals (4, 16, 32, and 64 ms) of preceding masking. Forward masking was more evident in the transient region of the response than in the sustained one. This study highlights the importance of electrophysiological testing in temporal processing assessment.

Keywords: Evoked Potentials; Speech; Noise

INTRODUCTION

Auditory processing involves all connections from cochlea to the auditory cortex. It is responsible for sound localization and lateralization, auditory discrimination, auditory pattern recognition, temporal hearing (temporal resolution, temporal masking, temporal integration, and temporal ordering), and auditory performance with competing and degraded acoustic signals[1]. When these abilities are well-developed, adequate auditory processing occurs, whereas difficulties in them may cause problems to understand speech sounds[1,2].

Most social listening situations require the listener to recognize speech in background noise to establish social communication. However, understanding speech in noise – which is related to the auditory temporal processing, specifically temporal masking – is a challenge for some listeners. Aspects involved in understanding speech in noise have been broadly investigated, and such related complaints are frequent, even among normally hearing people[3]. The ability to recognize speech in noise is based on temporal perception of sounds – therefore, it is also related to auditory temporal processes[4].

Temporal processing is the ability of the auditory system to perceive or distinguish different stimuli in a transient temporal sequence. It encompasses four categories: temporal ordering, temporal resolution, temporal masking, and temporal integration[1]. Temporal masking is the change in a sound threshold due to presence of another one. When a target speech and the masking noise are perceived at the same time, masking effect is called simultaneous temporal masking; when the masking noise is perceived a few milliseconds before the target speech, masking effect is called forward masking; and, when the masking noise occurs after the target speech, masking effect is named backward masking[5].

In forward masking, masking noise remains in the auditory system for a few milliseconds after it has physically ceased or decreased in amplitude, leading to a change in forward speech perception[3]. This effect happens when the speech signal and masking noise are separated by different intervals. It may happen because the hair cells, after being stimulated by the masking noise, require a few milliseconds to recover their sensitivity, to then be stimulated by the subsequent speech sound. The magnitude of this recovery depends on several characteristics, such as the interval between noise and speech[3].

Electrophysiological measures have been used to study auditory processing[5], by using various stimuli, including speech sounds, to elicit a response[6]. Speech-evoked auditory brainstem response (sABR), by using a syllable as stimulus (usually /da/) generates responses according to the transient and sustained components of the syllable (/d/ and /a/, respectively). A series of positive and negative peaks are observed regarding the transient and sustained components of the syllable. The sustained region of the response is called frequency-following response (FFR) due to its periodic characteristics. However, FFR has been commonly used to refer to speech-evoked auditory brainstem responses, including both transient and sustained portions of the responses[6-8].

Considering /da/ syllable, the consonant /d/ is its initial (onset) and transient component, and the vowel /a/, its sustained component. According to Skoe and Kraus[6] the first positive peak is called V, and the first valley is called A, and both respond to consonant /d/ (transient component). Responses elicited by the vowel /a/ (sustained component) are subsequently recorded valleys (D, E, F). The transition region is valley C and the last valley, named O, is the offset of the responses.

Differently from Skoe and Kraus[6], Hodge et al.[8] described peaks and valleys as PV, A, PW, PX, PY, PZ, and O. Responses of the transient component are PV and A. The following positive peaks (PW, PX, PY, PZ) are responses of the sustained portion of the stimulus, and the offset of the response is also named O. Analysis of the present study is based on the descriptions by Hodge et al.[8].

Despite a great number of electrophysiological studies with speech stimuli, few investigations have used forward masking to influence the response. This study aimed to analyze forward masking in normally hearing young people, by using frequency-following responses (FFR).

METHODS

This cross-sectional observational study was approved by the Research Ethics Committee at the Federal University of Pernambuco (*Universidade Federal de Pernambuco* – UFPE), Brazil, under protocol number 1.727.677.

Ten individuals, six of whom females, participated in this study. They were 18 to 25 years old (mean age 21 years), with normal hearing (pure-tone thresholds ≤ 25 dB HL at frequencies from 250 to 8000 Hz), and no history of speech and/or neurological disorders.

Any concerns regarding auditory processing disorders were investigated in the interview.

All FFR were recorded with Intelligent Hearing System (IHS), with participants inside a sound booth. Their skin was prepared with abrasive paste, and electrodes were placed according to 10-20 International System, as follows: two inverted electrodes at the mastoid (M1 and M2), a non-inverted electrode at Fz, and the ground electrode at Fpz. Both speech stimulus /da/ and masking noise were sent to the right ear via insert earphones (E39). The stimulus rate was 3.77 m/s, the window was set at 70 ms, with high- and low-pass filters at 50 Hz and 3000 Hz, respectively. For each trace, a total of 9,000 sweeps were acquired in three replicable runs of 3,000 sweeps. An ipsilateral channel was used to analyze the waves. In the resulting tracings, the PV (positive peak) and wave A (negative peak) of the transient region were identified and analyzed, as well as the PW, PX, PY, PZ, and O of the sustained component. Absolute latencies of all peaks were determined and analyzed by two audiologists with expertise in electrophysiological exams.

The synthetic syllable /da/ and a speech-shaped noise (SSN) were used. The speech syllable contains a transient component – consonant /d/ – and a sustained component – vowel /a/. The syllable lasted 40 ms and was presented at 75 dB peSPL, in alternated polarity. Masking noise, whose spectrum included Portuguese speech frequencies, was developed at University of North Carolina at Chapel Hill. It was presented at a fixed intensity of 80 dB SPL, lasting 100 ms (10 ms onset/offset ramps).

Each FFR trace was conducted in five conditions: 1) unmasked speech – /da/ alone; 2) masked speech – /da/ presented 4 ms after the noise; 3) masked speech – /da/ presented 16 ms after the noise; 4) masked speech – /da/ presented 32 ms after the noise; 5) masked speech – /da/ presented 64 ms after the noise.

Statistical analysis was performed to compare tested conditions: without noise was labeled "Unmasked", and those with a masking noise were labeled "4ms", "16ms", "32ms", and "64ms", referring to the delay between noise and syllable.

A descriptive analysis was performed for mean, standard deviation, coefficient of variation, minimum, and maximum value. Inferential analysis was performed with F-test (ANOVA) for repeated measures, with the Greenhouse-Geisser correction to compare testing conditions. For significant differences, multiple comparisons (between pairs of conditions) and the Bonferroni correction were used. Data normality was verified by applying the Shapiro-Wilk test.

Statistical significance was used at 5%. Statistical analyses were performed, by using the Statistical Package for the Social Sciences (SPSS), version 23.

RESULTS

Forward masking was analyzed by comparing latency of PV, A, PW, PX, PY, PZ, and O, with and without noise. The unmasked was compared to masked responses at 4, 16, 32, and 64 ms. Latency values at the unmasked condition were established as a reference to analyze changes in latency for the four maskied conditions.

Absolute latency (PV, A, PW, PX, PY, PZ, and O) for each testing condition (unmasked and masked responses at 4, 16, 32, and 64 ms) are shown in Table 1. Results indicate that the mean latencies for PV, A, PY, PZ, and O were lower for the unmasked test. PW latency was lower in the masked test at 64 ms as compared to the unmasked condition (22.03 and 22.05 ms, respectively). PX latency had the same mean values for unmasked and masked condition at 64 ms (30.65 ms).

Table 1. Electrophysiological responses in PV, A, PW, PX, PY, PZ, and O peaks, according to the testing conditions

Wave	Test condition	Mean	SD	CV	Minimum	Maximum
PV	Unmasked	7.18 (A)	0.37	5.15	6.63	7.95
	4ms	8.28 (B)	0.51	6.16	7.10	9.07
	16ms	8.22 (B)	0.31	3.77	7.75	8.80
	32ms	8.13 (B)	0.31	3.81	7.75	8.75
	64ms	7.90 (B)	0.42	5.32	7.48	8.75
P-value	p (1) = 0.001*					
A	Unmasked	8.46 (A)	0.72	8.51	7.35	9.73
	4ms	9.72 (B)	0.53	5.45	9.15	10.65
	16ms	9.66 (B)	0.32	3.31	9.28	10.25
	32ms	9.44 (B)	0.72	7.63	8.55	10.80
	64ms	9.17 (AB)	0.51	5.56	8.35	9.88
P-value	p (1) < 0.001*					
PW	Unmasked	22.05 (A)	0.37	1.68	21.63	22.88
	4ms	22.50 (B)	0.48	2.13	21.95	23.60
	16ms	22.45 (B)	0.42	1.87	21.83	23.33
	32ms	22.41 (B)	0.48	2.14	21.83	23.53
	64ms	22.03 (AB)	0.73	3.31	20.30	23.15
P-value	p (1) = 0.030*					
PX	Unmasked	30.65 (A)	0.44	1.44	30.10	31.65
	4ms	31.30 (B)	0.38	1.21	30.68	32.13
	16ms	31.13 (B)	0.46	1.48	30.53	32.13
	32ms	31.25 (B)	0.47	1.50	30.53	32.25
	64ms	30.65 (AB)	1.23	4.01	27.70	31.80
P-value	p (1) = 0.005*					
PY	Unmasked	39.27 (A)	0.54	1.38	38.45	40.45
	4ms	39.92 (B)	0.46	1.15	39.40	41.10
	16ms	39.75 (B)	0.31	0.78	39.25	40.25
	32ms	39.83 (B)	0.41	1.03	39.33	40.78
	64ms	39.62 (AB)	0.56	1.41	38.75	40.65
P-value	p (1) < 0.001*					
PZ	Unmasked	47.89	0.52	1.09	47.28	49.10
	4ms	48.34	0.49	1.01	47.60	49.43
	16ms	48.19	0.57	1.18	47.70	49.48
	32ms	48.01	0.77	1.60	46.45	49.63
	64ms	48.18	0.53	1.10	47.65	49.43
P-value	p (1) = 0.138					
O	Unmasked	49.06 (A)	0.76	1.55	48.03	50.60
	4ms	49.59 (B)	0.58	1.17	48.88	50.88
	16ms	49.47 (AB)	0.58	1.17	48.77	50.48
	32ms	49.19 (AB)	0.80	1.63	47.77	50.68
	64ms	49.12 (A)	0.56	1.14	48.58	50.48
P-value	p (1) = 0.021*					

Captions: SD = standard deviation; CV = coefficient of variation
(*) 5% significant difference.
(1) Applying the F-test (ANOVA) for repeated measures, by using Bonferroni pairwise comparisons.
Note: If all the letters in parentheses are different, there is a significant difference between the corresponding test conditions.

Regarding the masked conditions, latency of all waves was higher in 4 ms test and lower in 64 ms test (except for wave PZ). Variability expressed by the coefficient of variation was low (< 33.3%).

Significant differences in latency values were observed for most of the masked conditions when compared with unmasked latencies, except for the PZ wave, which had no significant differences in any of the masking conditions.

Multiple comparison tests showed (A) a significant difference between unmasked condition and: (Ai) all masking conditions for PV wave latency; (Aii) masking conditions at 4, 16, and 32 ms for the A, PW, PX, and PY wave latencies; (Aiii) masking conditions at 4 ms for O wave latency; and (B) in all masking conditions there were no significant differences between the mean latencies of the waves, except for O, which showed a significant difference at 4 ms.

Figure 1 corresponds to the grand average response for all testing conditions.
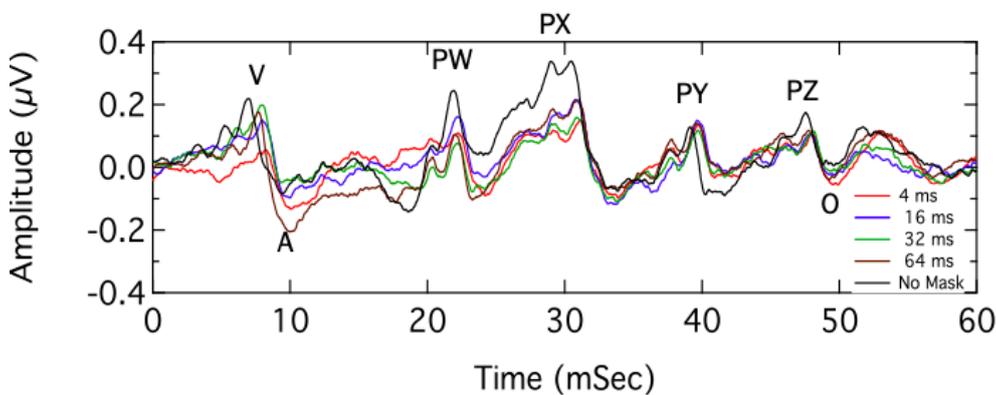
Figure 1. Frequency-following response in relation to noise delay in normally hearing young people

A latency delay was observed in the transient region of the response (represented by waves PV and A), which decreased in the sustained region (represented by waves PW to O). Waves of the sustained region (/a/) are PW, PX, PY, PZ, and O, in this order. It was observed that latency values were similar to unmasked condition. There was no difference in PZ wave latencies between unmasked and all masked conditions.

DISCUSSION

The ability to recognize speech in noisy environments requires the auditory system to distinguish target sound from the noise, especially when the signal-to-noise ratio is small[3], which can cause changes in speech perception. In noisy situations, the auditory system might not process transitory elements of the speech, as it is a complex sound with variations in frequency and amplitude[9]. Speech properties start to be processed at subcortical levels, at the brainstem nuclei – hence, recording the brainstem electrical responses provides accurate data on the word processing[10].

For the unmasked condition, latency responses of the transient component (PV and A) were similar to those previously reported. In the present study, latency for PV was 7.1 ms, and for A, 8.4 ms, as compared to wave V latency of 6.53 ms, wave A latency of 8.0 ms, wave V latency of 6.61 ms, and wave A latency of 7.5 ms[11]. This suggests an established brainstem response pattern for the transient region of the response of a speech stimulus.

In forward masking (i.e., masking preceding the speech stimulus), there was a significant delay in wave latencies elicited by the transient region of the stimulus (PV and A), when the /da/ syllable was presented 4, 16, 32, and 64 ms after the noise as compared to the unmasked responses of the abovementioned waves (except at 64 ms for wave A).

A greater simultaneous masking influence on the transient region of the response has been reported[11]. This may be due to the transient characteristics of the consonants, which exhibit more vulnerable components when exposed to noise. In noisy conditions, amplitude responses are smaller and do not show a high temporal periodicity[7]. Moreover, wave responses

are temporally closer to the end of masking noise, and this may also contribute to a greater interference of the masking[9].

Greater PV latency delays are also reported when compared with unmasked responses, suggesting a greater masking effect. Fogerty et al.[12] conducted a study using FFR with 18 syllables (combining consonants b, d, g, j with vowels a, i, u to obtain consonant-vowel and vowel-syllable syllables) in the following conditions: unmasked, with simultaneous masking, and masking preceding the stimulus (forward masking) at 10, 40, and 100 ms. They reported that the mean latencies also increased in both parts of the responses, though greater for the transient region when compared to the sustained region.

To understand forward masking effect on PV latencies, Walton et al.[13] used tone-burst stimuli (at frequencies of 1000, 4000, and 8000 Hz), in the following conditions: unmasked and masked preceding speech stimulus at 2, 4, 8, 16, 32, and 64 ms. Latencies increased at 4, 8, and 16 ms a compared to the unmasked condition. PV wave latency returned to values similar to unmasked condition when masking preceding the stimulus was at 64 ms – as found in the present study. Similar results were also described[5,7,11,14] with the /da/ syllable in unmasked and masked conditions; a significant latency delay was found in the transient region of the response when tested with background noise.

The consonants (which are the transient component of the syllable) seem to be more susceptible to the masking noise effect, as they have low-intensity acoustic cues and do not have high temporal periodicity[11]. A latency delay on the transient region of the response (waves PV and A) suggests a forward masking effect.

Psychoacoustic studies[3,5,9,15] have also demonstrat forward masking effect with an increase in hearing thresholds for a masked condition, preceding a short target stimulus – the smaller the intervals between masking noise and target speech, the higher the thresholds.

Grose et al.[15] have also investigated forward and backward masking effects on young and middle-aged people and found higher thresholds in all masked conditions in relation to age. This may explain why it is more difficult for older adults to understand speech in environmental noise.

Dubno et al.[16] presented a masking noise both simultaneously with and before target stimulus at 10,

20, 50, and 100 ms. Thresholds have also increased when both masking conditions were tested, evidencing the forward masking effect.

Forward masking has been widely documented by psychometric and electrophysiological studies, showing greater interference on the transient region of the responses, with minor effect on the sustained part of the FFR responses[17].

In the present study, changes were observed in wave latencies related to the sustained region (PW-O). However, these changes were less significant and more inconsistent. It may be difficult to identify a forward masking effect in this region of the response because of the interaction between the masking effect of the preceding noise and a masking effect caused by the stimulus itself (due to its more complex characteristics)[18,19].

Latencies of the sustained region of the response in unmasked condition had been previously demonstrated demonstrated[13]. In the present study, when the stimuli were presented at different intervals (4, 16, 32, and 64 ms) after the masking noise, latencies differed for four masked conditions. However, no significant differences were found in the sustained masked responses when compared with the unmasked condition – perhaps due to sweep differences between those two studies, while here a great sweeps were used.

Furthermore, changes in fundamental frequency can cause changes in the sustained region. Fundamental frequency of the stimulus plays an important role in subcortical coding, facilitating (or confusing) perception of the sustained component of the syllable. However, it is known that vowels (sustained component) contain intense acoustic cues with higher periodicity and are less influenced by noise[18,19].

Speech coding at subcortical level suggests that speech perception may be influenced by decoding temporal aspects of speech and that this ability is already perceived at brainstem level, which is essential for good comprehension of complex sounds such as speech[10].

In summary, results of the present study reinforce that the transient component of the stimulus (i.e., consonant) is more susceptible to noise. This was observed in wave PV and A latency delay. On the other hand, the sustained component, represented by PW to O latencies, seems to be less influenced by noise, possibly because vowels are more intense and have higher temporal periodicity, which favors their perception during forward masking. In other

words, although a latency delay has been observed in the sustained region after noise, forward masking effect was stronger on the transient component of the stimulus.

FFR seems to be an objective measure to understand forward masking in different populations, especially when using a 4 ms delay between the syllable and the noise.

CONCLUSION

Forward masking was observed, by using frequency-following responses (FFR) with the /da/ syllable at four intervals (4, 16, 32, and 64 ms) of preceding masking. Forward masking was most evident in the transient region of the response (which corresponds to consonant /d/). In the sustained region (corresponding to the vowel /a/), forward masking was noticed in the 4 and 16 ms testing conditions, while less evident in the 32 and 64 ms ones. Therefore, no clear pattern of forward masking on the sustained portion of FFR responses was found.

REFERENCES

1. American Speech-Language-Hearing Association (ASHA). Central auditory processing: current status of research and implications for clinical practice. Am J Audiol. 1996;5(2):41-54.

2. Pienkowski M. On the etiology of listening difficulties in noise despite clinically normal audiograms. Ear Hear. 2017;38(2):135-48.

3. Advíncula KP, Menezes DC, Pacífico FA, Costa MLG, Griz SMS. Age effects in temporal auditory processing: modulation masking release and forward masking effect. Audiol., Commun. Res. 2018;23(e1861):1-6.

4. Guimaraes ACF, Santos JN, Rabelo ATV, Magalhaes MC. A Influência do ruído na habilidade auditiva de ordenação temporal para sons verbais. Rev. CEFAC. 2015;17(1):209-15.

5. Tanner AM, Spitzer ER, Hyzy JP, Grose JH. Masking release for speech in modulated maskers: electrophysiological and behavioral measures. Ear Hear. 2019;40(4):1009-15.

6. Skoe E, Kraus N. Auditory brainstem response to complex sounds: a tutorial. Ear Hear. 2010;31(3):302-24.

7. Mehraei G, Gallardo AP, Shinn-Cunningham BG, Dau T. Auditory brainstem response latency in forward masking, a marker of sensory deficits in listeners with normal hearing thresholds. Hearing research. 2017;346:34-44.

8. Hodge SA, Menezes DC, Brown KD, Grose JH. Forward masking of the speech-evoked auditory brainstem response. Otol Neurotol. 2018;39(2):150-7

9. Grose JH, Menezes DC, Porter HL, Griz S. Masking period patterns & forward masking for speech-shaped noise: age-related effects. Ear Hear. 2016;37(1):48-54.

10. Bildeman GM. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. NeuroImage. 2018;15(175):56-69.

11. Necciari T, Laback B, Savel S, Ystad S, Balazs P, Meunier S et al. Auditory time-frequency masking for spectrally and temporally maximally-compact stimuli. PloS one. 2016;11(11):e0166937.

12. Forgety D, Bologna WJ, Ahlstrom JB, Dubno JR. Simultaneous and forward masking of vowels and stop consonants: effects of age, hearing loss, and spectral shaping. J Acoust Soc Am. 2017;141(2):1133-43.

13. Walton J, Orlando M, Burkard R. Auditory brainstem response forward-masking recovery functions in older humans with normal hearing. Hear Res. 1999;127(1-2):86-94.

14. Necciari T, Laback B, Savel S, Ystad S, Balazs P, Meunier S et al. Auditory time-frequency masking for spectrally and temporally maximally-compact stimuli. PloS one. 2016;11(11):e0166937.

15. Grose JH, Griz SMS, Pacífico FA, Advíncula KP, Menezes DC. Modulation masking release using the Brazilian-Portuguese HINT: psychometric functions and the effect of speech time compression. Int J Audiol. 2015;54(4):274-81.

16. Dubno JR, Horwitz AR, Ahlstrom JB. Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal hearing. J Acoust Soc Am. 2003;113:2084-94.

17. Griz SMS, Menezes DC, Angelo VLG, Fonsêca NH, Nascimento TO, Araújo AKL et al. Effect of forward masking on frequency following response as a function of age. J Am Acad Audiol. 2020;31(5):317-23.

18. Ferreira L, Gubiani MB, Keske-Soares M, Skarzynski PH, Sanfins MD, Biaggio EPV. Analysis of the components of frequency-following response in phonological disorders. Int J Pediatr Otorhinolaryngol. 2019;122:47-51.

19. Ananthakrishnan S, Luo X, Krishnan A. Human frequency following responses to vocoded speech. Ear Hear. 2017;38(5):e256-e267.