



## Original Article

# Effect of Consonant Duration on Formation of Consonant-Vowel Syllable Evoked Auditory Cortical Potentials

Mehmet Yaralı, Süha Yağcıoğlu

Department of Audiology, Hacettepe University School of Medicine, Ankara, Turkey (MY)  
Department of Biophysics, Hacettepe University School of Medicine, Ankara, Turkey (SY)

Cite this article as: Yaralı M, Yağcıoğlu S. Effect of Consonant Duration on Formation of Consonant-Vowel Syllable Evoked Auditory Cortical Potentials. J Int Adv Otol 2018; 14(1): 39-43.

**OBJECTIVE:** To compare P1-N1-P2-N2 response latencies and amplitudes evoked by voiced and unvoiced consonant-vowel syllables (CVS) /bi-/pi/ and /di-/ti/ by analyzing how the cortical responses to consonants and vowels interact during the formation of a syllable-evoked response.

**MATERIALS and METHODS:** Auditory late latency responses were recorded from 12 healthy individuals between the ages of 20 and 40 years with normal hearing while presenting /bi-/pi/ and /di-/ti/ tokens and individual consonant-vowel parts of syllables. Amplitude/latency values of P1-N1-P2-N2 responses were compared between /bi-/pi/ and /di-/ti/ pairs. Formation of CVS-evoked responses by consonant and vowel responses was also investigated.

**RESULTS:** N1-P2-N2 latencies evoked by /bi/ were significantly shorter than /pi/. P2-N2 amplitudes evoked by /di/ were significantly higher and N2 latencies were shorter than /ti/. N1-P2-N2 peaks of /bi/, /pi/, and /di/ seemed to be combinations of respective peaks of consonant and vowel-evoked responses. For /ti/, P1 and N1 seem to be stemming only from the consonant part, P2 from consonant P2 and vowel N1, and N2 from consonant N2 and vowel P2-N2.

**CONCLUSION:** For both CVS pairs, longer consonant durations resulted in lower amplitudes and/or longer latencies, and this sheds light on why voiced-unvoiced CVSs evoke cortical responses with different features. Obtaining evoked responses to each consonant-vowel part of the syllables among listeners with perceptual difficulties and hearing devices might help to reveal which acoustic cues are not well represented in the auditory brain.

**KEYWORDS:** Consonant duration, consonant-vowel syllables, auditory evoked cortical potentials

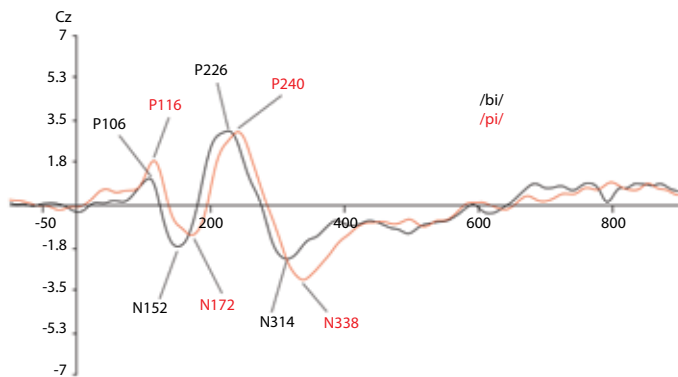
## INTRODUCTION

One of the methods for investigating cortical-level auditory processing is analyzing the cortical P1-N1-P2 responses evoked by short-duration stimuli such as tone bursts, clicks, and speech tokens. Although these responses are actually evoked by stimuli onsets, they are also evoked by acoustic changes. For example, frequency changes in continuous tones<sup>[1-3]</sup>, amplitude and spectrum changes in speech sounds<sup>[4]</sup>, embedded silent gaps in noise<sup>[5,6]</sup>, and changes in periodicity of ongoing signals<sup>[7]</sup> evoke P1-N1-P2 response times that are locked to the acoustic changes in the stimuli. Effects of acoustic changes can also be observed in consonant-vowel syllable (CVS)-evoked cortical potentials. For example, N1 latency<sup>[8]</sup> and the presence or absence of two N1 responses corresponding to the onset of the vowel part of a CVS is affected by the voice onset time (VOT)<sup>[9]</sup>. VOT is defined as the time lag between release of the consonant and periodic low frequency glottal pulsations<sup>[10]</sup>. In a series of studies with monkeys, stimuli with short VOT evoked single-peak responses, whereas double-peak response times locked to consonant and vowel onsets were observed for long VOT stimuli<sup>[11,12]</sup>. Similarly, double-peak N1 responses for consonant-vowel stimuli having long VOT and single-peak N1 responses for stimuli with short VOT was observed in human subjects<sup>[13,14]</sup>. Not only N1 responses, but also P1 and N2 responses were found to be affected by VOT, namely, P1-N2 response latencies were prolonged in children for stimuli with long VOT<sup>[15]</sup>. VOT also has implications for evoked potential amplitudes. For example, the amplitudes of N1 and P2 responses are higher for voiced (short VOT) syllables than unvoiced (long VOT) syllables<sup>[11]</sup>. A similar effect was observed for N2 responses in children, with larger N2 amplitudes being observed for short VOT<sup>[16]</sup>.

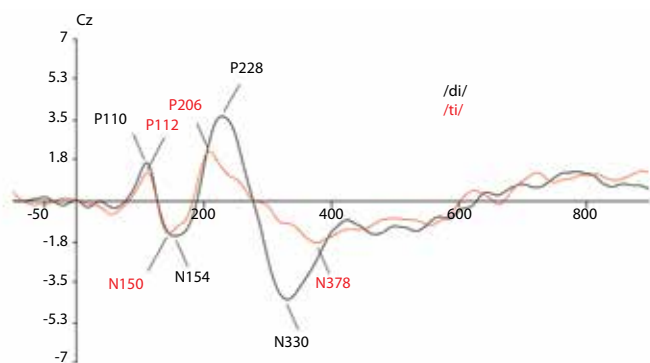
**Corresponding Author:** Mehmet Yaralı; mhmyrly@yahoo.com.tr

**Submitted:** 15.12.2016 • **Revision Received:** 20.06.2017 • **Accepted:** 07.07.2017 • **Available Online Date:** 02.11.2017

©Copyright 2018 by The European Academy of Otology and Neurotology and The Politzer Society - Available online at [www.advancedotology.org](http://www.advancedotology.org)



**Figure 1.** Cortical responses evoked by /bi/ and /pi/  
Cz: central zero electrode



**Figure 2.** Cortical responses evoked by /di/ and /ti/  
Cz: central zero electrode

Keeping in mind the effect of VOT (a parameter in which the consonant and vowel parts of a syllable interact with each other) on syllable-evoked cortical potentials, one possible way to investigate how each of the consonant and vowel parts combine to form the syllable-evoked response is to record responses evoked by each constituent part in addition to the whole syllable response. This was investigated by Ostroff et al. [17] who recorded cortical potentials evoked by the word /sei/ and /s/, /ei/, in other words, the word and the consonant-vowel parts. The positive peak of /sei/ was reported to consist of the P2 evoked by the consonant portion and the P1 evoked by the vowel, and P2 primarily seemingly consist of the P2 response to the vowel or to the combination of offset response to /s/ and onset response to /ei/. The observation that the first positive peak evoked by /sei/ consists of the consonant-evoked P2 and vowel-evoked P1 might be related to the relatively long duration of the consonant. In that study, it could be observed that different peaks (for example, P2 of the consonant and P1 of the vowel) are combining to form the first positive peak of /sei/ are combining to form the syllable-evoked response due to the relatively long duration of the consonant part. That is, if the consonant duration were shorter, the P1 peaks evoked by each part would be combining to form the whole syllable-evoked response. One idea at this point might be to analyze how responses to CVSs with different consonant durations are formed by the combination of responses to their respective consonant and vowel parts and to observe the formation of CVS-evoked responses by the consonant and vowel parts with different consonant durations. This would allow for the determination of the consonant duration effect on the formation of syllable-evoked responses.

The motivation in the current study was to investigate cortical responses evoked by voiced-unvoiced consonant-vowel stimuli pairs with different consonant durations and different constituent consonant-vowel parts in order to reveal the effect of consonant duration on the formation of the syllable-evoked response. For this aim, voiced-unvoiced /bi-/pi/ and /di-/ti/ pairs, with different consonant durations between the voiced-unvoiced parts of each pair, were utilized for evoked potential testing. The consonant duration difference between /di/ and /ti/ is larger than that of /bi/ and /pi/. Shorter consonant durations are expected to lead to the same polarity responses forming the CVS-evoked response, whereas longer consonant durations will result in different combinations of polarity responses.

**MATERIALS and METHODS**

Approval from the university ethics committee was obtained for this research with decision code GO 16/146-35, and all participants gave written informed consent. CVS-evoked auditory late latency responses were obtained from 12 healthy individuals between the ages of 20 and 40 years. All participants had pure tone thresholds of 20 dB or better at frequencies of 250-8000 Hz. Participants had no history of neurological-psychiatric disorders or use of related medications. The stimuli for auditory evoked potential recording were /bi-/pi/ and /di-/ti/ syllables prepared via an online text to speech translator (AT&T Text to Speech, <http://www2.research.att.com>) [18] and processed via Praat Software [19]. The stimuli durations were 275, 303, 290, and 334 ms for /bi/, /pi/, /di/, and /ti/, respectively. In order to see the effects of consonant and vowel parts of the evoked responses on the CVS-evoked response, individual consonant and vowel parts were also used in evoked potential recording by replacing either the vowel or consonant part with silence. In each recording block, the same stimulus was presented 120 times with stimulus onset asynchrony of 1100 ms while participants were watching a subtitled muted movie. The order of blocks was randomized for each participant. Twenty-channel electroencephalography (EEG) recordings were obtained with a Neuroscan 4.3. system (Neuroscan; Compumedics, Charlotte, USA) while subjects were seated in a comfortable armchair watching a subtitled muted movie of their choice. EEG data were analyzed with the EEG Lab [20] and ERP Lab [21] software packages. Raw EEG recordings were band-pass filtered at 0.5-30 Hz and notch-filtered at 50 Hz with artifact rejection at  $\pm 100 \mu\text{V}$  and epoched between -100 ms and 900 ms. The artifact rejection threshold was  $\pm 100 \mu\text{V}$ . The linked ear reference was used.

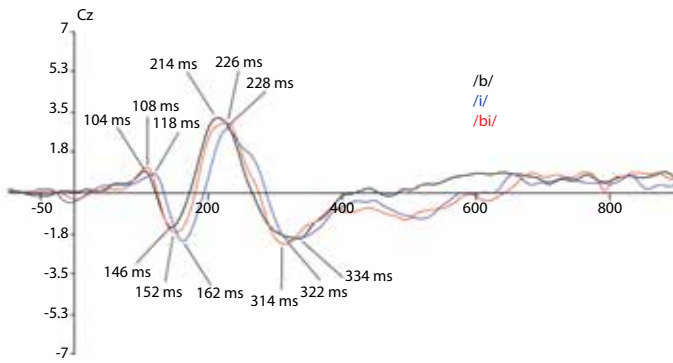
**Statistical Analysis**

Statistical Package for Social Sciences (Version 13.0, SPSS Inc.; Chicago, USA) was utilized for comparing cortical potentials evoked by syllables /bi-/pi/ and /di-/ti/. P1, N1, P2 and N2 response latencies and amplitudes were compared within the syllable pairs of /bi-/pi/ and /di-/ti/ with independent samples t-test.

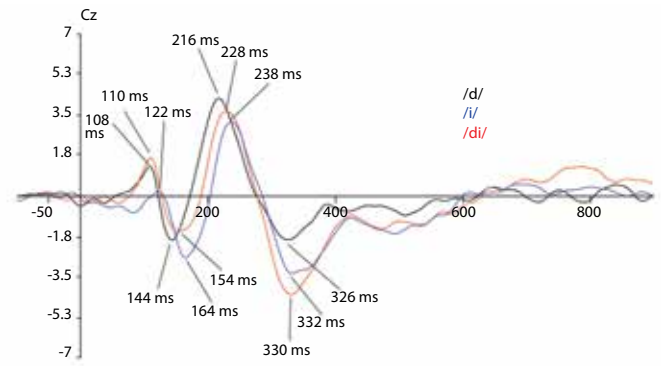
**RESULTS**

/bi-/pi/ and /di-/ti/ speech tokens evoked P1-N1-P2-N2 responses with similar morphologies at the central zero (Cz) electrode (Figure 1, 2). For the sake of clarity, peaks are named together with their latencies.

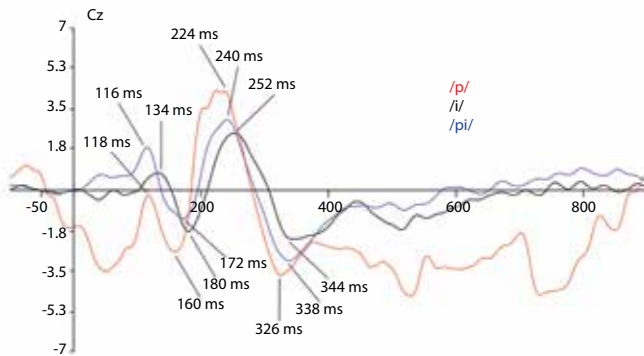
Comparison between responses evoked by /bi-/pi/ and /di-/ti/ pairs showed that N1, P2, and N2 response latencies were significant-



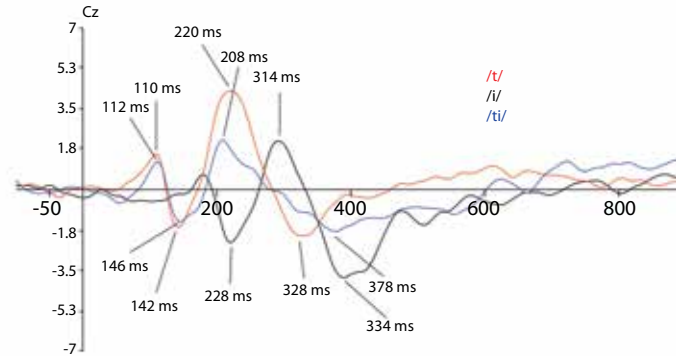
**Figure 3.** Cortical responses evoked by /b/, /i/, and /bi/ Cz: central zero electrode



**Figure 5.** Cortical responses evoked by /d/, /i/, and /di/ Cz: central zero electrode



**Figure 4.** Cortical responses evoked by /p/, /i/, and /pi/ Cz: central zero electrode



**Figure 6.** Cortical responses evoked by /t/, /i/, and /ti/ Cz: central zero electrode

ly shorter ( $t(9)=-3.48, p=.007$ ;  $t(8)=-4, p=.01$ ; and  $t(11)=-4, p=.002$ , respectively) for /bi/ ( $M=151.6, SD=9.88$ ;  $M=220.44, SD=15.35$ ; and  $M=311.67, SD=36.09$ , respectively) than for /pi/ ( $M=168, SD=14.85$ ;  $M=233.11, SD=14.36$ ; and  $M=344.33, SD=34.44$ , respectively). P2 and N2 amplitudes were significantly higher ( $t(7)=3.1, p=.017$  and  $t(9)=-2.37, p=.042$ , respectively) for /di/ ( $M=5.48, SD=2.84$  and  $M=-5.12, SD=3.87$ , respectively) than for /ti/ ( $M=3.37, SD=1.55$  and  $M=-2.47, SD=1.39$ , respectively). N2 latencies of /di/ ( $M=333.2, SD=9.81$ ) were significantly shorter ( $t(9)=-3.03, p=.014$ ) than those of /ti/ ( $M=359.4, SD=25.96$ ).

Responses evoked by consonant and vowel parts and by whole responses evoked by the CVSs are shown for /bi/, /pi/, /di/, and /ti/ in Figures 3, 4, 5, and 6 respectively. Similar to what was reported by Ostroff et al. [17], responses to CVSs are combinations of individual waveforms evoked by each part. For example, N1-P2-N2 peaks of /bi/ and /pi/ seem to be combinations of respective peaks of consonant and vowel-evoked responses, and the N1-P2-N2 peaks occurred at latencies in-between the respective peaks of consonant and vowel-evoked responses. When investigated in detail, it could be observed that the reason for early latencies for N1-P2 peaks of /bi/ seem to be related to earlier latencies of /b/-evoked N1-P2 peaks. The first negative and second positive peaks of /b/ were 146 ms and 214 ms, respectively, which were earlier than those of /p/ (160 ms and 224 ms, respectively). Moreover, because the time lag between consonant and vowel parts is longer in /pi/, the peaks evoked by /i/ have prolonged latency for this syllable, leading to a shift in latency of the /pi/-evoked peaks compared to /bi/. Here it

can be observed that consonant duration leads to a shift in latencies of /pi/ in addition to a longer latency for the consonant-evoked part compared to /b/. For N2 latency, although the /b/-evoked N2 latency is prolonged compared to /p/ (334 and 326 ms, respectively), the time lag between the onset of /p/ and /i/ leads to a large difference in terms of /i/-evoked N2 latencies of /bi/ and /pi/ (314 ms and 338 ms, respectively).

The combination is slightly different for /ti/, and P1 and N1 seem to be stemming only from the consonant part- P2 from consonant P2 and vowel N1, and N2 from consonant N2 and vowel P2-N2. The N2 latency difference between /di/ and /ti/ seems to be stemming from the longer latency of the /i/-evoked response belonging to /ti/ (390 ms) compared to the /i/-evoked N2 peak latency for /di/ (332 ms). Although /d/ and /t/-evoked N2 latencies are very close to each other (326 and 328 ms, respectively), unlike the case for /b/ and /p/, the /i/ part of the syllable /ti/ seems to be leading to a shift in N2 latency of the /ti/ compared to /di/.

In terms of amplitudes, the time lag between /t/ and /i/ seems to be leading to a reduction in P2 and N2 amplitudes due to different polarity responses combining to form the peaks. In Figure 6 it can be observed that P2 of /ti/ seems to be composed of the second positive peak of /t/ and first negative peak of /i/, leading to a reduction in the P2 amplitude of the syllable. This also holds for N2, which is composed of /t/-evoked N2 peak and /i/-evoked P2-N2 peaks. Here the combination of different polarity responses seems to be leading to a reduction in N2 amplitudes similarly to the formation of

P2. For /di/, the same polarity peaks belonging to the consonant and vowel parts are forming the syllable-evoked response, leading to shorter latencies and larger amplitudes compared to different polarity response combinations for /ti/. Contrary to /b/ and /p/, /d/ and /t/-evoked responses have similar latencies of N2 (326 ms and 328 ms, respectively). For /di/ and /ti/, the differences in latencies of the evoked responses seems to be solely dependent on consonant duration, unlike for /bi/ and /pi/ for which the voiced consonant-evoked response peak latencies of the CVS /bi/ are earlier and the prolongation of latencies for the unvoiced token /pi/ is further increased due to the time lag between the consonant and vowel parts as discussed before.

## DISCUSSION

In the current work, cortical potentials evoked by /bi-/pi/ bilabial stops and alveolar stops /di-/ti/ were recorded, and the evoked responses were different within the pairs. The N1-P2-N2 response latencies were significantly shorter for voiced /bi/ than for unvoiced /pi/, and this is in line with findings of the relation between VOT and N1<sup>[9, 10]</sup>, N2<sup>[15]</sup>, and P2<sup>[8]</sup>. However, /bi/ was not found to be evoking larger N1-P2 responses than /pi/, contrary to the findings of Tremblay et al.<sup>[16]</sup>, probably due to the higher VOT difference between pairs in the study by Tremblay et al.<sup>[16]</sup> (around 61 ms) compared to ours (around 17 ms.) Similarly, Han et al.<sup>[22]</sup> did not observe consistent changes in P2 amplitudes with 0-50 ms VOT changes of 10 ms steps in-between the /ba-/pa/ continuum.

Differences between pairs were also observed for /di-/ti/-evoked responses; N2 responses were larger in amplitude and shorter in latency for /di/ compared to /ti/, and this is compatible with previous work with children<sup>[22]</sup>. The P2 amplitude was also found to be larger for voiced /di/, but N1-P2 latencies were not different, and this is similar to previous work on voiced/unvoiced pairs of /bi-/pi/<sup>[16]</sup>. Han et al.<sup>[22]</sup> also showed no shortening of the N1 latency for /ba/ compared to /pa/ in VOT steps of 10 ms between 0 to 50 ms. In our study, the /di-/ti/ pair had a VOT difference of about 50 ms, which is equal to the maximum VOT difference between /bi-/pi/ pairs in Han et al.<sup>[22]</sup>. A longer VOT difference might lead to prolongation in N1 latencies.

At this point, the possible effect of consonant duration on latency and amplitude differences between voiced-unvoiced CVS-evoked cortical responses might come to mind. Previously, it has been found that voiced-unvoiced CVS-evoked cortical response parameters differ, but the reason for this has not been well clarified. The suspected reason for this finding, namely, the consonant differences between CVSs, was also investigated in our current work in order to shed light on the reason for differences between cortical responses evoked by voiced and unvoiced CVSs.

Although interpretation of these results might be at the limits of speculation, obtaining the individual consonant and vowel-evoked responses was valuable because these give us insights into the way responses to syllables are formed. Previously, the study by Ostroff et al.<sup>[17]</sup> investigated the cortical potentials evoked by parts of a word (/sei/) and discussed how these interact to make up the whole word-evoked potential. In this study, we have applied this decomposition to a set of CVSs and observed that P1-N1-P2-N2

responses evoked by CVSs were composed of respective peaks evoked by consonant and vowel parts as expected. For all speech tokens, the P1-N1-P2-N2 peaks seemed to be forming at latencies between the respective peaks of the consonant and vowel parts. When comparing /bi/ and /pi/, for which the same polarity responses belonging to consonant and vowel parts are combining to form the syllable-evoked response, the early latency values for responses evoked by /bi/ compared to /pi/ seemed to be related to early latencies evoked by the voiced consonant part and to relatively shorter consonant duration for /b/ compared to /p/. Combinations of responses to consonant-vowel parts were slightly different for /ti/, and the voltage deflections with different polarities belonging to the /t/ and /i/- evoked responses while the /ti/-evoked response is forming seemed to be leading to some kind of cancellation resulting in lower P2-N2 amplitudes for /ti/ than for /di/. In terms of N2 latencies, although the voiced and unvoiced /d/ and /t/-evoked peak latencies were close to each other, the longer consonant duration of /ti/ seemed to be leading to prolonged latency for the /i/-evoked response and to a shift in the /ti/-evoked N2 response compared to /di/.

These observations about how consonant and vowel-evoked responses combine to form the syllable-evoked potentials show how the consonant part of the syllable affects the response latencies and amplitudes. To sum up, it can be stated that for /bi-/pi/, response latencies are affected by both consonant duration and the consonant-evoked response itself, and for /di-/ti/ the response latencies and amplitudes seemed to be mostly dependent on consonant duration. To gain more insight into how responses from consonant and vowel parts combine to form the syllable-evoked responses, manipulation of consonant durations and forming different CV syllables (in order to see the effect of the consonant part on the syllable-evoked response) and observing the effects of these manipulations on cortical responses is suggested.

The recommendation at this point is what Tremblay et al.<sup>[16]</sup> has suggested; to create more syllables that reflect different acoustic characteristics and to use them in cortical response testing. Auditory cortical responses evoked by CVSs having different consonant durations can be recorded in order to test the possible systematic effect of consonant duration on evoked potential parameters. Moreover, obtaining evoked responses for each consonant and vowel part of the syllables among listeners with perceptual difficulties might help reveal which acoustic cues are not well represented in the auditory brain. Also, cortical responses evoked by CVSs and individual parts can be measured among hearing aid and cochlear implant users with different perceptual capabilities in order to gain some insight into which acoustic cues are not represented robustly among users with poor perceptual capabilities. Programming hearing devices and providing auditory rehabilitation can be based on these poorly represented acoustic features.

**Ethics Committee Approval:** Ethics committee approval was received for this study from the ethics committee of Hacettepe University.

**Informed Consent:** Written informed consent was obtained from subjects who participated in this study.

**Peer-review:** Externally peer-reviewed.



**Author Contributions:** Concept - M.Y.; Design - M.Y.; Supervision - S.Y.; Resources - M.Y., S.Y.; Data Collection and/or Processing - M.Y., S.Y.; Analysis and/or Interpretation - M.Y., S.Y.; Literature Search - M.Y.; Writing Manuscript - M.Y.; Critical Review - S.Y.

**Conflict of Interest:** No conflict of interest was declared by the authors.

**Financial Disclosure:** The authors declared that this study has received no financial support.

## REFERENCES

- Dimitrijevic A, Michalewski HJ, Zeng FG, Pratt H, Starr A. Frequency changes in a continuous tone: auditory cortical potentials. *Clin Neurophysiol* 2008; 119: 2111-24. [CrossRef]
- Wang WJ, Tan CT, Martin BA. Auditory evoked responses to a frequency glide following a static pure tone. *Proc Meet Acoust Acoust Soc Am* 2013; 19: 1-7. [CrossRef]
- Arlinger SD, Jerlvall LB, Ahren T, Holmgren EC. Slow evoked cortical responses to linear frequency ramps of a continuous pure tone. *Acta Physiol Scand* 1976; 98: 412-24. [CrossRef]
- Martin BA, Boothroyd A. Cortical, auditory, evoked potentials in response to changes of spectrum and amplitude. *J Acoust Soc Am* 2008; 107: 2155-61. [CrossRef]
- Lister JJ, Maxfield ND, Pitt GJ. Cortical evoked response to gaps in noise: within-channel and across-channel conditions. *Ear Hear* 2007; 28: 862-78. [CrossRef]
- Palmer SB, Musiek FE. N1-P2 recordings to gaps in broadband noise. *J Am Acad Audiol* 2013; 24: 37-45. [CrossRef]
- Martin BA, Boothroyd A. Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope. *Ear Hear* 1999; 20: 33-44. [CrossRef]
- Tremblay KL, Piskosz M, Souza P. Aging alters the neural representation of speech cues. *Neuroreport* 2002; 13: 1865-70. [CrossRef]
- Sharma A, Marsh CM, Dorman MF. Relationship between N1 evoked potential morphology and the perception of voicing. *J Acoust Soc Am* 2000; 108: 3030-5. [CrossRef]
- Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J Neurophysiol* 1999; 82: 2346-57. [CrossRef]
- Steinschneider M, Schroeder CE, Arezzo JC, Vaughan HG. Speech-evoked activity in primary auditory cortex: effects of voice onset time. *Electroencephalogr Clin Neurophysiol* 1994; 92: 30-43. [CrossRef]
- Steinschneider M, Schroeder CE, Arezzo JC, Vaughan HG. Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: temporal response patterns. *Brain Lang* 1995; 48: 326-40. [CrossRef]
- Sharma A, Dorman MF. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J Acoust Soc Am* 1999; 106: 1078-83. [CrossRef]
- Sharma A, Marsh CM, Dorman MF. Relationship between N1 evoked potential morphology and the perception of voicing. *J Acoust Soc Am* 2000; 108: 3030-35. [CrossRef]
- King KA, Campbell J, Sharma A, Martin K, Dorman M, Langran J. The representation of voice onset time in the cortical auditory evoked potentials of young children. *Clin Neurophysiol* 2008; 119: 2855-61. [CrossRef]
- Tremblay KL, Friesen L, Martin BA, Wright R. Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear Hear* 2003; 24: 225-32. [CrossRef]
- Ostroff JM, Martin BA, Boothroyd A. Cortical evoked response to acoustic change within a syllable. *Ear Hear* 1998; 19: 290-7. [CrossRef]
- AT&T text to Speech, Available from: <http://www2.research.att.com>
- Boersma P. Praat, a system for doing phonetics by computer. *Glott Int* 2001; 5: 341-5.
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Meth* 2004; 134: 9-21. [CrossRef]
- Lopez-Calderon J, Luck SJ. ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front Hum Neurosci* 2014; 8: 1-14. [CrossRef]
- Han JH, Zhang F, Kadis DS, Houston LM, Samy RN, Smith ML, et al. Auditory cortical activity to different voice onset times in cochlear implant users. *Clin Neurophysiol* 2016; 127: 1603-17. [CrossRef]