



## OPEN ACCESS

## EDITED BY

Alexis Deighton MacIntyre,  
University of Cambridge, United Kingdom

## REVIEWED BY

Agudemu Borjigin,  
University of Wisconsin-Madison,  
United States  
Yue Zhang,  
Cochlear, Australia

## \*CORRESPONDENCE

Hyun Joon Shim  
✉ eardoc11@naver.com

†These authors share first authorship

RECEIVED 09 September 2024

ACCEPTED 08 April 2025

PUBLISHED 02 May 2025

## CITATION

Choi HJ, Kyong J-S, Won JH and Shim HJ  
(2025) Neural adaptations to temporal cues  
degradation in early blind: insights from  
envelope and fine structure vocoding.  
*Front. Neurosci.* 19:1493641.  
doi: 10.3389/fnins.2025.1493641

## COPYRIGHT

© 2025 Choi, Kyong, Won and Shim. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Neural adaptations to temporal cues degradation in early blind: insights from envelope and fine structure vocoding

Hyo Jung Choi<sup>1,2†</sup>, Jeong-Sug Kyong<sup>3,4†</sup>, Jong Ho Won<sup>5</sup> and  
Hyun Joon Shim<sup>1,2\*</sup>

<sup>1</sup>Department of Otorhinolaryngology-Head and Neck Surgery, Nowon Eulji Medical Center, Eulji University School of Medicine, Seoul, Republic of Korea, <sup>2</sup>Eulji Tinnitus and Hearing Research Institute, Nowon Eulji Medical Center, Seoul, Republic of Korea, <sup>3</sup>Sensory Organ Institute, Medical Research Institute, Seoul National University, Seoul, Republic of Korea, <sup>4</sup>Department of Radiology, Konkuk University Medical Center, Seoul, Republic of Korea, <sup>5</sup>Alston & Bird, LLP, Washington, DC, United States

In our previous study, early-blind individuals have better speech recognition than sighted individuals, even when the spectral cue was degraded using noise-vocoders. Therefore, this study investigated the impact of temporal envelope degradation and temporal fine structure (TFS) degradation on vocoded speech recognition and cortical auditory response in early blind individuals compared to sighted individuals. The study included 20 early-blind subjects ( $31.20 \pm 42.5$  years, M: F = 11:9), and 20 age- and -sex-matched sighted subjects. Monosyllabic words were processed using the Hilbert transform to separate the envelope and TFS, generating vocoders that included only one of these components. The amplitude modulation (AM) vocoder, which contained only the envelope component, had the low-pass filter's cutoff frequency for AM extraction set at 16, 50, and 500 Hz to control the amount of AM cue. The frequency modulation (FM) vocoders, which contained only the TFS component, were adjusted to include FM cues at 50%, 75%, and 100% by modulating the noise level. A two-way repeated measures ANOVA revealed that early-blind subjects outperforming sighted subjects across almost all AM or FM-vocoded conditions ( $p < 0.01$ ). Speech recognition in early-blind subjects declined more with increasing TFS degradation, as evidenced by a significant interaction between group and the degree of TFS degradation ( $p = 0.016$ ). We also analyzed neural responses based on the semantic oddball paradigm using the N2 and P3b components, which occur 200–300 ms and 250–800 ms after stimulus onset, respectively. Significant correlations were observed between N2 and P3b amplitude/latency and behavioral accuracy ( $p < 0.05$ ). This suggests that early-blind subjects may develop enhanced neural processing strategies for temporal cues. In particular, preserving TFS cues is considered important for the auditory rehabilitation of individuals with visual or auditory impairments.

## KEYWORDS

speech intelligibility, temporal degradation, vocoder, temporal envelope, temporal fine structure, N2 and P3b

# 1 Introduction

Auditory temporal resolution refers to the auditory system's ability to detect and process rapid changes in sound over time. Temporal processing allows the auditory system to extract important features, such as pitch, timing, and the rhythmic structure of speech, which are crucial for distinguishing between different speech sounds and understanding speech, especially in noisy environments (Haggard, 1984; McKay et al., 2013; McFarlane and Sanchez, 2024). An acoustic signal in the temporal domain is decomposed into a slowly varying temporal envelope and a rapidly varying temporal fine structure (TFS) (Hilbert, 1912). The temporal envelope cue plays a crucial role in speech recognition in quiet environments environments (Drullman et al., 1994; Shannon et al., 1995). Studies using a noise vocoder, where the bandwidth is divided and the temporal envelope information of each band is preserved, have shown that even when most of the spectral cue of speech is removed, 90% of words were correctly identified through the temporal envelope (Shannon et al., 1995). Smith et al. found that when using 4–16 frequency bands of an “auditory chimera,” in which the envelope from one sound is paired with the TFS of another, the recognition of English speech was dominated by the envelope (Smith et al., 2002), whereas the recognition of tonal languages, such as Mandarin Chinese, relies more on TFS (Xu and Pfingst, 2003; Wang et al., 2015). TFS becomes important in sound localization (Yin and Chan, 1990; Smith et al., 2002; Borjigin et al., 2022), as well as pitch perception through fundamental-frequency (Moore, 1973; Houtsma and Smurzynski, 1990; Qin and Oxenham, 2005) and music perception (Smith et al., 2002; Heng et al., 2011). However, there has been a long-standing debate on whether TFS contributes to masking release through spatial cues and F0 information (Lorenzi et al., 2006; Moore, 2008; Oxenham, 2008; Gnansia et al., 2009; Oxenham and Simonson, 2009). A recent study found that greater TFS sensitivity does not enhance masking release from F0 or spatial cues but aids resilience to reverberation and reduces listening effort, as indicated by faster response times (Borjigin and Bharadwaj, 2025).

But blind individuals rely solely on auditory signals for communication, making it essential to investigate their speech perception abilities in comparison with sighted individuals. This is particularly important for developing rehabilitation programs for visually impaired individuals. Early-blind individuals, who were either blind at birth or became blind within the first year of life, experience compensatory mechanisms in the brain that enhance the processing of non-visual senses such as hearing and touch. This enhancement also extends to their auditory temporal resolution abilities. Several studies have demonstrated that early-blind individuals show advantages in temporal-order judgment ability (Weaver and Stevens, 2006), temporal patterns (Bae et al., 2022), auditory temporal resolution (Muchnik et al., 1991), temporal modulation detection (Shim et al., 2019), and temporal attention for stimulus selection (Röder et al., 2007) over sighted subjects.

Our previous study demonstrated that speech recognition declined as spectral cues were reduced (i.e., with a decreased number of channels) in both blind and sighted individuals. However, early-blind individuals have better speech recognition

than sighted individuals, even when the spectral cue was degraded using noise-vocoders with different numbers of channels (Choi et al., 2024a). Nonetheless, spectral degradation had a greater impact on speech recognition with increasing degradation in early-blind subjects. Therefore, this study focused on temporal resolution to determine whether early-blind subjects have speech recognition advantages over sighted subjects in environments with various levels of degraded temporal resolution. Blind individuals are strongly reliant on auditory cues for communication without visual cues, which could markedly disrupt their daily life, even with minor impairments in temporal cues. However, few studies have examined speech recognition in blind individuals in the context of limited auditory temporal cues.

To investigate the impact of temporal resolution degradation on the speech recognition of early-blind individuals, we used noise-vocoded speech. Monosyllabic words were processed using the Hilbert transform to separate the envelope and TFS, generating vocoders that included only one of these components. Using an amplitude modulation (AM) vocoder, which contained only the envelope component, the low-pass filter's cutoff frequency for AM extraction was set at 16, 50, and 500 Hz to control the envelope cut-off frequency of the AM cue (Shannon et al., 1995). The frequency modulation (FM) vocoders, which contained only the TFS component, were adjusted to include FM cues at amount of 50%, 75%, and 100% by modulating the noise level (Moon et al., 2014).

In addition, we used the “semantic oddball paradigm” to investigate the neural correspondence of speech recognition affected by degradation of the temporal cues in early-blind individuals. We focused on the N2 and P3b components, which are associated with higher-order neural processing for stimulus discrimination and evaluation (Voola et al., 2023). These components likely depend more on top-down processing when temporal speech cues are degraded. The N2 component is a negative deflection starting around 200–300 ms post-stimulus (Folstein and Van Petten, 2008), and is a sensitive index for examining the course of semantic and phonological encoding (Schmitt et al., 2000) or listening to sound with the oddball paradigm (Finke et al., 2016; Voola et al., 2023). P3b, which occurs between 250 and 800 ms, exhibits a variable peak that is dependent on the individual's response, and its amplitudes are typically greater over the parietal electrodes. P3b was measured using the parietal electrodes (CP1, CP2, P3, P4, and Pz), as outlined in Finke et al. (2016). P3b is associated with the judgment of stimulus inconsistency while updating working memory. Prolonged latencies may represent slower stimulus evaluation (Beynon et al., 2005; Henkin et al., 2015). Our previous study using a one-syllable oddball paradigm with animal and non-animal stimuli across varying channel vocoder conditions confirmed that the N2 and P3b responses reflect cortical effects. This indicates that semantic integration is less efficient due to reduced spectral information in speech (Choi et al., 2024b). Therefore, we assessed semantic processing, as represented by the N2 and P3b responses, using the same paradigm with degradation of the envelope and TFS cues, and compared these responses between early-blind and sighted subjects.

## 2 Subjects and methods

### 2.1 Subjects

The study population included a group of 20 early-blind subjects ( $31.20 \pm 4.25$  years, male: female [M: F] = 11:9) and a control group of 20 sighted subjects ( $28 \pm 6.9$  years, male: female [M: F] = 11:9). There was no significant difference in age between the two groups ( $p < 0.05$ ). All of the subjects were right-handed, aged  $< 40$  years, and had normal hearing thresholds in both ears ( $\leq 20$  dB hearing level at 0.25, 0.5, 1, 2, 3, 4, and 8 kHz). They had no other neurological or ontological problems. The early-blind group only included people who were blind at birth or who became blind within 1 year of birth, and were classified in categories 4 and 5 according to the 2006 World Health Organization guidelines for the clinical diagnosis of visual impairment (category 4, “light perception” but no perception of “hand motion”; category 5, “no light perception”). Table 1 provides the characteristics of the blind subjects. The study was conducted by the Declaration of Helsinki and the recommendations of the Institutional Review Board of Nowon Eulji Medical Center, with written informed consent from all subjects. Informed consent was obtained verbally from the blind subjects in the presence of a guardian or third party. The subjects then signed the consent form, and a copy was given to them.

### 2.2 AM- and FM-vocoded speech

Stimuli were recorded by a male speaker reading five lists of 25 Korean monosyllabic words in a soundproof booth using a lapel microphone (BY-WMA4 PRO K3; BOYA, Shenzhen, Hong Kong). All the recorded stimuli were sampled at a rate of 44,100 Hz. The overall root mean square amplitude was normalized to  $-25$  dB relative to full scale using Adobe Audition (Adobe Systems, San Jose, CA, USA), ensuring that the average signal intensity was 25 dB below the maximum possible digital level to maintain consistent stimulus intensity across recordings.

For the amplitude modulation vocoder, the input signal was first filtered into eight frequency bands ranging from 80 to 8,000 Hz, with each band equally spaced on an equivalent rectangular bandwidth scale (Glasberg and Moore, 1990).

The band cutoffs were determined to ensure that the filter bandwidths closely matched those of the auditory filters. The cutoff frequencies of each bandpass filter were determined using a logarithmically spaced frequency range based on the Greenwood function (80, 214, 424, 748, 1,250, 3,234, 5,103, and 8,000 Hz). The cutoff frequency of the low-pass filter for temporal envelope extraction was applied at 16, 50, and 500 Hz. The central frequency of each channel was calculated as the geometric mean between the two corresponding cutoff frequencies associated with that specific channel. The amplitude envelope for each frequency band was then extracted through Hilbert transform. Finally, we summed the sub-band signals to generate the noise-vocoded signals (Shannon et al., 1995; Faulkner et al., 2012; Evans et al., 2014) (Figure 1A).

For the FM-vocoder, the input signal was first filtered using a wideband bandpass filter (80–8,000 Hz; Figure 1B). The Hilbert transform was then applied to each subband signal to decompose

it into its analytic signal, from which the envelope and temporal fine structure were extracted. The TFS component was isolated by retaining only the phase information, represented by the cosine value of the phase of the analytic signal. A separate set of band-limited noise signals was generated and filtered using the same wideband bandpass filter as employed for the input signal. The root mean square of band-limited noise signals was set to that of analytic signal. To vary the amount of FM cues available in the output signals, we used the phase randomization technique of Moon et al. (2014).

$$Y(t) = \text{abs}(X(t)) \times \cos(\text{angle}([(1 - NF) \times X(t)] + [NF \times N(t)])) \quad (1)$$

where  $Y(t)$  is the output stimulus,  $X(t)$  is the analytic signal,  $N(t)$  is the filtered noise in an analytic form, and  $NF$  is a “noise factor” from 0 to 1. We added the weighted random noise component (i.e., analytic signal  $[NF \times N(t)]$ ) to the weighted original analytic signal  $[(1 - NF) \times X(t)]$ . Then, the randomized TFS was obtained by taking the cosine value of the angle of these mixed signals. The randomized TFS was then modulated with the envelope of the 1-band signal. We tested  $NF$  values of 0.5, 0.25, and 0. The  $NF$  value of 0.5 produced the output signal containing 50% of the FM cues for the original signal. The  $NF$  value of 0.25 produced the output signal including 75% of the original FM cues. Finally, the  $NF$  value of 0 preserved the intact (100%) FM cues. Vocoding was performed using a custom MATLAB script (2020a, Mathworks, Inc., Natick, MA, USA), in which the spectra became more blurred as the cut-off frequency of the envelope decreased and as the preserved amount of the TFS decreased, as shown in Figure 2.

### 2.3 Procedures

#### 2.3.1 Behavioral test

Speech recognition using the AM and FM vocoders was compared between early-blind and sighted subjects. The perception of one-syllable words was tested under three different amounts of envelope cues (AM vocoder: 16, 50, and 500 Hz cutoff frequency) and three different amounts of TFS cues (FM vocoder: 50, 75, and 100%) using five lists, each containing 25 Korean monosyllabic words. The participants were asked to repeat the words after they were presented through a loudspeaker placed 1 meter in front of the subject's ear. All tests were conducted in a soundproof room with an audiometer (Madsen Astera 2; GN Otometrics, Taastrup, Denmark), and the stimuli was presented at 70 dB SPL. The word recognition scores were calculated as the percentage of correctly repeated words.

#### 2.3.2 N2 and P3b

The neural response was recorded across 31 AG-Ag/Cl sintered electrodes placed according to the international 10-20 system (Klem, 1999) and referenced at FCz in an elastic 32-channel cap using the actiCHamp Brain Products recording system (BrainVision Recorder Professional, V.1.23.0001, Brain Products GmbH, Munich, Germany). All recordings were made in a dimly

TABLE 1 Clinical characteristics for the early-blind subjects.

| Subject | Age (years) | Onset | Sex | Visual acuity       | Cause of blindness                       |
|---------|-------------|-------|-----|---------------------|------------------------------------------|
| B01     | 26          | Birth | M   | No light perception | Corneal opacity                          |
| B02     | 29          | Birth | F   | No light perception | Retinopathy of prematurity               |
| B03     | 25          | Birth | M   | NO light perception | Cause unknown                            |
| B04     | 39          | Birth | M   | Light perception    | Optic atrophy                            |
| B05     | 26          | Birth | M   | Light perception    | Xanthochromism                           |
| B06     | 25          | Birth | M   | No light perception | Optic atrophy                            |
| B07     | 27          | Birth | M   | No light perception | Retinopathy of prematurity               |
| B08     | 31          | Birth | F   | No light perception | Retinopathy of prematurity               |
| B09     | 37          | Birth | F   | No light perception | Optic atrophy                            |
| B10     | 30          | Birth | F   | No light perception | Glaucoma                                 |
| B11     | 31          | Birth | F   | No light perception | Retinopathy of prematurity               |
| B12     | 34          | Birth | F   | Light perception    | Retinoblastoma                           |
| B13     | 31          | Birth | M   | Light perception    | Cause unknown                            |
| B14     | 36          | Birth | F   | Light perception    | Retinopathy of prematurity               |
| B15     | 28          | Birth | F   | Light perception    | Optic atrophy                            |
| B16     | 31          | Birth | F   | No light perception | Poor eye development                     |
| B17     | 32          | Birth | M   | No light perception | Persistent hyperplastic primary vitreous |
| B18     | 28          | Birth | M   | No light perception | Retinopathy of prematurity               |
| B19     | 27          | Birth | M   | Light perception    | Microphthalmos                           |
| B20     | 37          | Birth | M   | No light perception | Retinopathy of prematurity               |

lit, sound-attenuated, electrically shielded chamber. The electro-oculogram (EOG) and electrocardiogram (ECG) were tagged to trace the subject’s eye movement and heartbeat, respectively. The electroencephalogram (EEG) data were digitized online at a sampling rate of 1,000 Hz. The ground electrode was placed between electrodes Fp1 and Fp2. Software filters were set at low (0.5 Hz) and high (70 Hz) cutoffs. A notch filter at 60 Hz was set to prevent powerline noise, and the impedances of all scalp electrodes were kept below 5 k $\Omega$  using EEG electrode gel throughout the recording, following the manufacturer’s instructions.

2.3.2.1 Oddball paradigm

Based on the semantic oddball paradigm, the subjects listened to animal stimuli or non-animal but meaningful stimuli (Choi et al., 2024b). Overall, 70% of the trials involved animal words (e.g., mouse, snake, bear; all monosyllable in Korean). The remaining 30% consisted of monosyllable non-animal words but belonged to a different semantic category. The subjects sat comfortably in the soundproof booth and listened to the animal or non-animal words in a random order. The researchers told the subjects to expect to hear an animal word and instructed them to press the button as quickly and accurately as possible upon hearing the word. In the cutoff frequency condition (16, 50, and 500 Hz) and the TFS condition (50, 75, and 100%), 210 animal words and 90 non-animal words were presented in six blocks, and the subjects listened to a total of 900 trials in each condition. The inter-stimulus interval was fixed at 2,000 ms, and a jitter of 2–5 ms was allowed.

The order of presentation was randomized within the blocks and the order of blocks was counterbalanced among subjects using E-Prime software (version 3, Psychology Software Tools, Sharpsburg, PA). Each subject had a 5-min break after completing each block. The subjects had a familiarizing session before starting the trials to ensure that they understood the task and that their muscles were relaxed. The intensity of sound was fixed at 70 dB SPL when calibrated at the listener’s head position, 1 m from the loudspeaker.

2.3.2.2 Data processing

The data were preprocessed and analyzed with Brain Vision analyzer (version 2.0, Brain Products GmbH) and MATLAB R2019b (Mathworks) using EEGLAB v2021 (Delorme and Makeig, 2004) and Fieldtrip (Oostenveld et al., 2011) toolboxes. EEG was filtered with a high-pass filter at 0.1 Hz (Butterworth filter with a 12 dB/oct roll-off) and a low-pass filter at 50 Hz (Butterworth filter with a 24 dB/oct rolloff). Data were resampled at 256 Hz. Fast independent component analysis (Hyvärinen and Oja, 2000) was used to reject artifacts associated with eye blinks and body movement (average of 4 independent components, range 3–6) and reconstructed (Makeig et al., 1997), with transformation to the average reference. The EEG waveforms were time-locked to each stimulus onset and segmented from 200 ms before the stimulus onset to 1,000 ms after the stimulus onset. Baseline correction was then performed. The epochs with incorrect behavioral responses were excluded from further preprocessing. Before averaging, bad channels were interpolated using a spherical spline function (Perrin

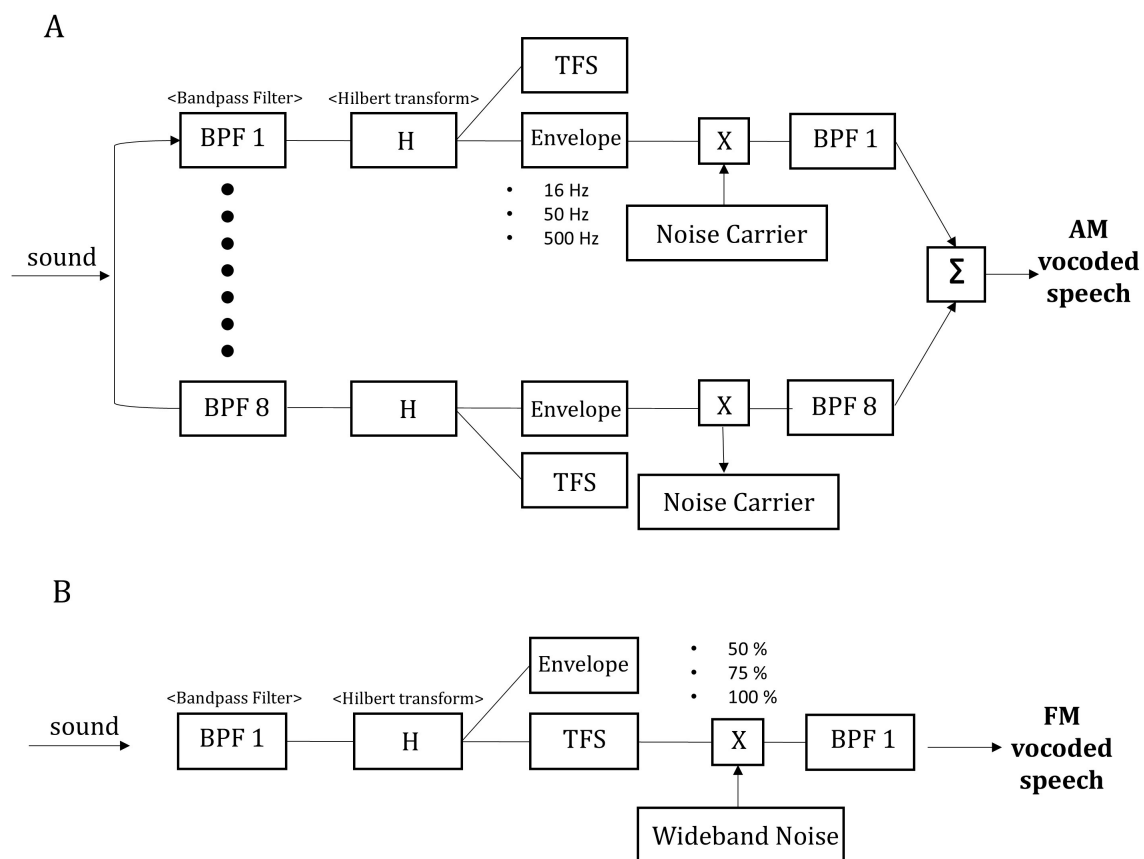


FIGURE 1

Schematic diagram of the amplitude modulation (AM) vocoder (A). The input sounds were divided into eight channel bands using bandpass filters (BPF1 to BPF8), and each filtered sound was subjected to Hilbert transformation (H) to extract the envelope of each band, removing the temporal fine structure (TFS). The temporal envelope cutoff frequencies for AM extraction were set at 16, 50, and 500 Hz. The vocoded speech signal was generated by adding a noise carrier to the envelopes in each channel band. Finally, the signals were passed through each bandpass filter and summed to produce the AM-vocoded speech sound. Schematic diagram of the frequency modulation (FM) vocoder (B). The input sound was passed through a single-frequency bandpass filter (BPF1) and the filtered sound was subjected to H to extract the TFS. The amount of TFS was manipulated by wideband noise (50, 75, and 100%).

et al., 1989), and segments with values  $\pm 70 \mu V$  at any electrode were rejected. All of the subjects had data for at least 150–197 usable standard trials out of 210 trials and 63–87 usable target trials out of 90 trials. An average wave file was generated for each subject for each condition. According to previous studies, the latency ranges for N2 and P3b were determined based on the grand average computed across all conditions and participants. Accordingly, the N2 component in the current study was defined as the periods of 330–650 ms and 350–600 ms post-stimulus onset for AM and FM, respectively. The P3b component was defined as the periods of 560–895 ms and 565–895 ms post-stimulus onset for AM and FM, respectively. The peak latency and peak amplitude were measured by half-area quantification, which may be relatively unaffected by latency jitter (Luck, 2014; Finke et al., 2016). The ERP latency was quantified using the 50% area latency measure. We computed the signed area under the ERP waveform over a given latency range and then defined the time point that divides the area in half. This measure is known to be less affected by single-trial latency jitter and it is relatively insensitive to high-frequency noise (Petermann et al., 2009; Meyer et al., 2011; Luck, 2014). Difference

waveforms were constructed by subtracting the target stimuli from the standard stimuli within each condition (Deacon et al., 1991). The area latency and amplitude of the N2 and P3b difference waveforms were compared between each condition and group. N2 was measured by pooling the signals from the frontocentral electrodes (Fz, FC1, FC2, and Cz), whereas P3b was measured by averaging the signals from the parietal electrodes (CP1, CP2, P3, P4, and Pz), as illustrated in Figure 3 and outlined in Finke et al. (2016).

## 2.4 Statistical analysis

Two-way repeated-measures analysis of variance (RM-ANOVA) was used to analyze the effects of group, AM vocoder, and FM vocoder on monosyllable recognition, as well as the latency and amplitude of the N2 and P3b components. *Post-hoc* paired *t*-tests, significance levels were set at 0.05 for multiple comparisons after applying Bonferroni's correction to the *p*-values. Pearson correlation analyses between AM or FM vocoded speech



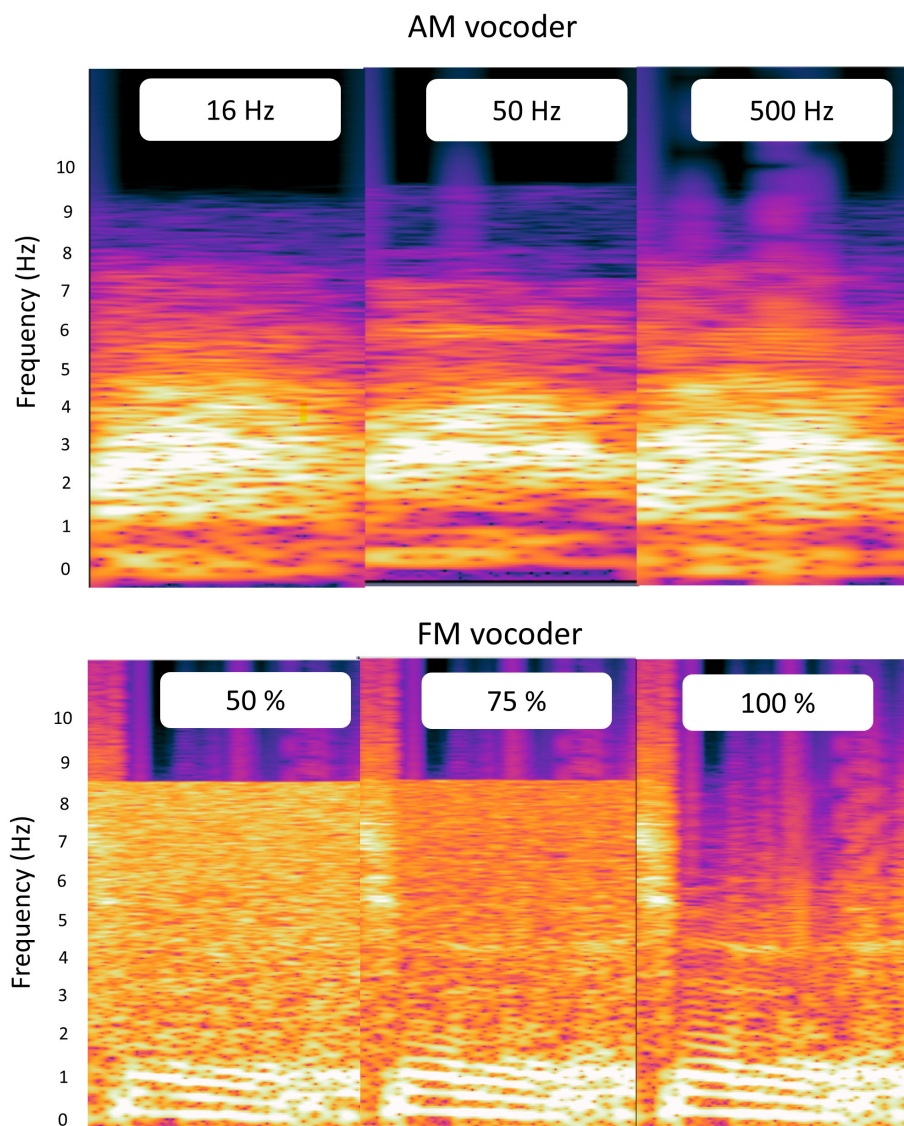


FIGURE 2

Spectrograms of the amplitude modulated (AM) and frequency modulated (FM) vocoder outputs for the word “MAL”. The top row shows the spectrograms for the AM vocoder at three different temporal envelope cutoff frequencies (16, 50, and 500 Hz). The bottom row displays the spectrograms for the FM vocoder at three different amounts of temporal fine structure (TFS; 50, 75, and 100%).

recognition and neural responses of the N2 and P3b components were performed with Bonferroni’s correction ( $\alpha = 0.05/6 = 0.008$ ). All statistical analyses were performed using IBM SPSS software (ver. 25.0; IBM Corp, Armonk, NY, USA).

## 3 Results

### 3.1 Behavioral data

We measured the recognition of vocoded speech with the temporal envelope and TFS, each degraded at three different levels. A mixed two-way RM-ANOVA (two groups  $\times$  envelope cutoff frequency) revealed significant main effects of group ( $F_{(1,38)} = 9.734, p = 0.003$ ) and envelope cutoff frequency ( $F_{(1.568,59.566)} =$

$69.151, p < 0.001$ ). However, there was no significant interaction between group and envelope cutoff frequency ( $F_{(1.568,59.566)} = 0.954, p = 0.372$ ). *Post-hoc* tests using Bonferroni correction indicated that early-blind subjects outperformed sighted subjects in AM-vocoded speech recognition across all cutoff frequencies (16 Hz:  $p = 0.002$ ; 50 Hz:  $p = 0.004$ ; 500 Hz:  $p = 0.008$ ; Table 2, Figure 4A).

For TFS, the RM-ANOVA (two groups  $\times$  amount of TFS) showed significant main effects of group ( $F_{(1,38)} = 6.301, p = 0.016$ ) and amount of TFS ( $F_{(2,76)} = 393.653, p < 0.001$ ), and a significant interaction between these two factors ( $F_{(2,76)} = 4.363, p = 0.016$ ). In the *post-hoc* tests using Bonferroni correction revealed that early-blind subjects showed better FM-vocoded speech recognition than sighted subjects, except at a TFS of 50% (50% TFS:  $p = 0.639$ ; 75% TFS:  $p = 0.016$ ; 100% TFS:  $p = 0.017$ ; Table 3, Figure 4B).

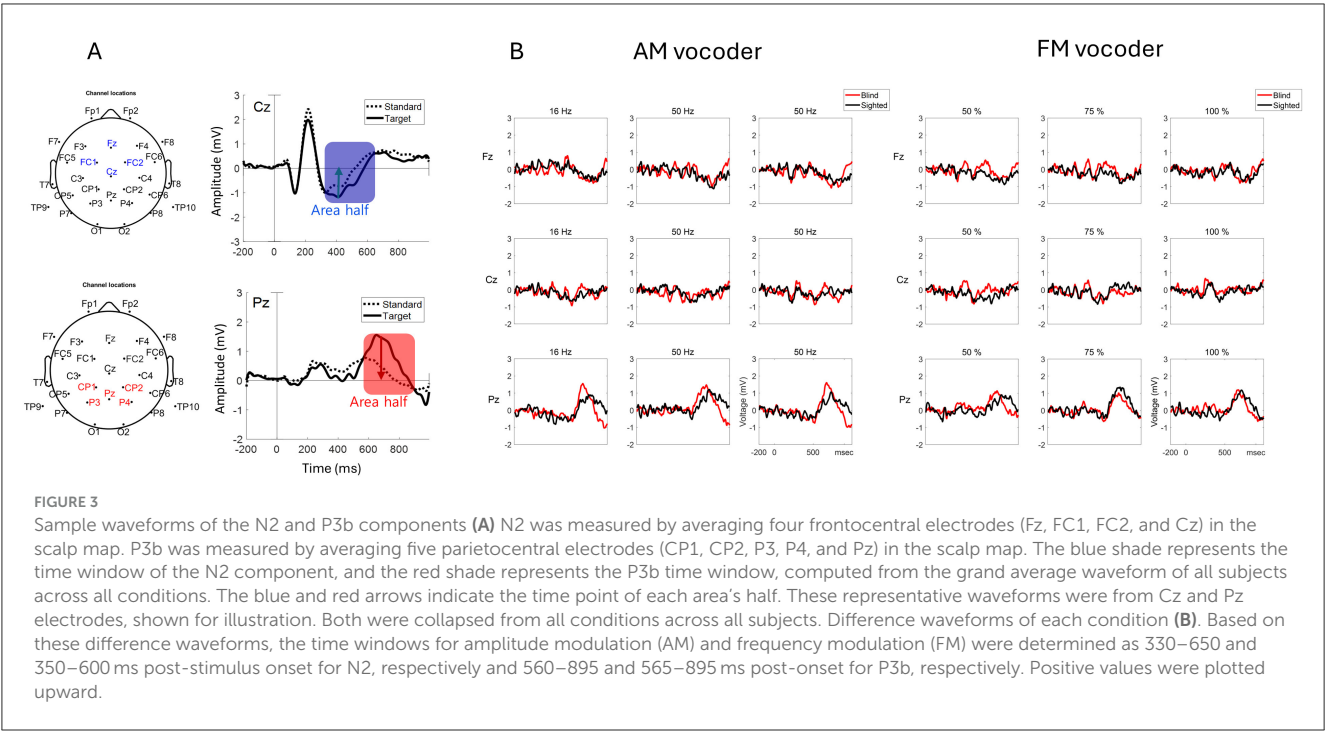


TABLE 2 Statistical summary of envelope cutoff frequency (amplitude modulated) vocoded speech.

|                   | Sum of square          | df     | Mean square     | F              | p         | $\eta^2_G$ |
|-------------------|------------------------|--------|-----------------|----------------|-----------|------------|
| Cut-off frequency | 211.517                | 1.568  | 134.936         | 69.151         | <0.001*** | 0.645      |
| Group             | 300.833                | 1      | 300.833         | 9.734          | 0.003**   | 0.204      |
| AM*Group          | 2.917                  | 1.568  | 1.861           | 0.954          | 0.372     | 0.024      |
| Residual          | 116.233                | 59.566 | 1.951           |                |           |            |
| <i>Post-hoc</i>   |                        |        |                 |                |           |            |
|                   |                        |        | Mean difference | Standard error | p         |            |
| 16 Hz             | Ealy-blind vs. sighted |        | −2.75           | 0.842          | 0.002**   |            |
| 50 Hz             | Ealy-blind vs. sighted |        | −3.25           | 1.065          | 0.004**   |            |
| 500 Hz            | Ealy-blind vs. sighted |        | −3.50           | 1.246          | 0.008**   |            |

\*\* $p < 0.01$ , \*\*\* $p < 0.001$ .  
 $\eta^2_G = \frac{SS_{effect}}{SS_{effect} + SS_{subjects} + SS_{error}}$  (Small: < 0.01, Medium: 0.01~0.06, Large: 0.14).

Overall, the results indicate that early-blind subjects showed superior recognition compared with sighted subjects, even under conditions with degradation of the auditory temporal envelope and TFS. Speech recognition in early-blind subjects declined more with increasing TFS degradation, as evidenced by a significant interaction between group and the degree of TFS degradation. However, there was no difference between the groups regarding the impact of temporal envelope degradation on speech recognition.

3.2 EEG data

The effect of envelope cutoff and group on the latency and amplitude of N2 and P3b was examined using mixed two-way

RM-ANOVA (two groups  $\times$  envelope cutoff frequency). The analysis revealed a significant effect of envelope cutoff frequency for the N2 amplitude ( $F_{(1.549,58.881)} = 7.244$ ,  $p = 0.003$ ) and P3b latency ( $F_{(2,76)} = 14.238$ ,  $p < 0.001$ ). The group effect for the P3b amplitude showed a trend toward significance ( $F_{(1,38)} = 4.081$ ,  $p = 0.050$ ), although the result did not reach the conventional threshold for statistical significance ( $p < 0.05$ ; Table 4, Figure 5).

For TFS, the RM ANOVA (two groups  $\times$  amount of TFS) showed a significant effect of the amount of TFS on N2 latency ( $F_{(2,76)} = 8.400$ ,  $p < 0.001$ ) and amplitude ( $F_{(2,76)} = 7.812$ ,  $p < 0.001$ ), as well as P3b latency ( $F_{(2,76)} = 8.734$ ,  $p < 0.001$ ) and amplitude ( $F_{(2,76)} = 15.868$ ,  $p < 0.001$ ). However, significant group effects were not found for the latency or amplitude of N2 or P3b (Table 5, Figure 5).

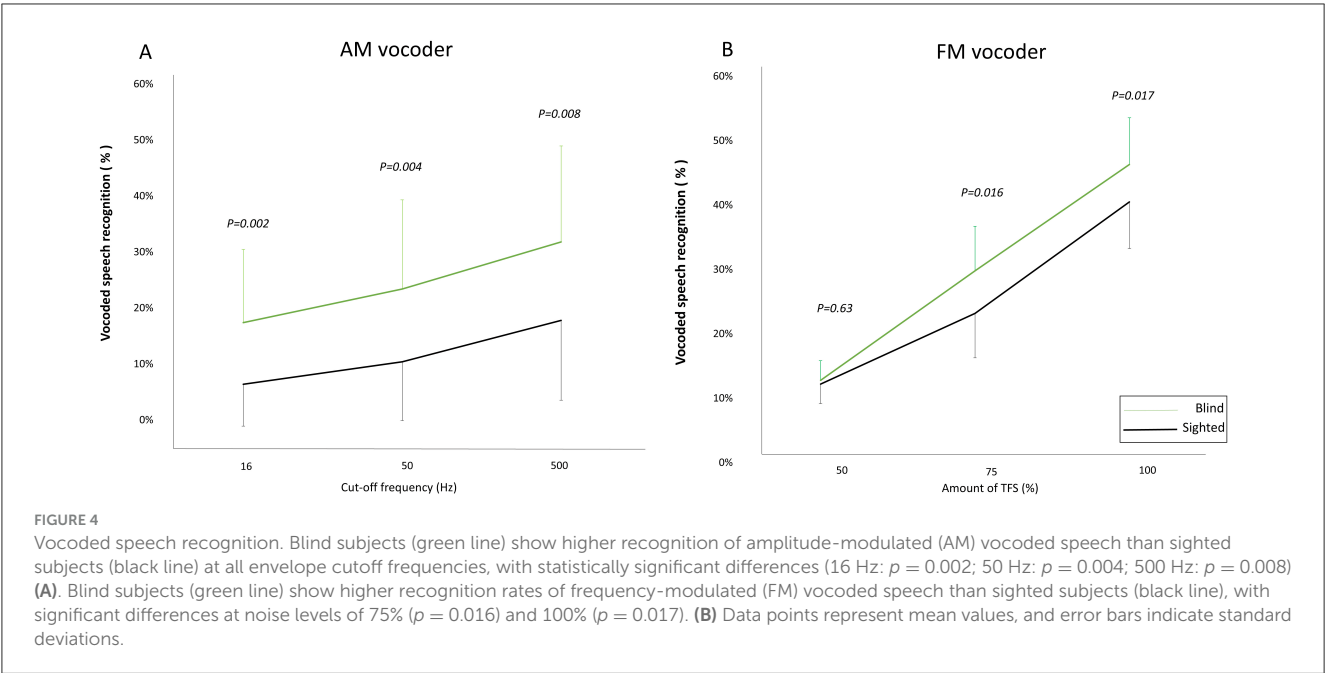


TABLE 3 Statistical summary of the amount of temporal fine structure (TFS; frequency modulated) vocoded speech.

|               | Sum of square          | df | Mean square     | F              | p         | $\eta^2_G$ |
|---------------|------------------------|----|-----------------|----------------|-----------|------------|
| Amount of TFS | 4,788.067              | 2  | 2,394.653       | 393.653        | <0.001*** | 0.912      |
| Group         | 140.833                | 1  | 140.833         | 6.301          | 0.016*    | 0.142      |
| FM*Group      | 53.067                 | 2  | 26.533          | 4.363          | 0.016*    | 0.103      |
| Residual      | 462.200                | 76 | 6.082           |                |           |            |
| Post-hoc      |                        |    |                 |                |           |            |
|               |                        |    | Mean difference | Standard error | p         |            |
| 50%           | Ealy-blind vs. sighted |    | −0.3            | 0.634          | 0.639     |            |
| 75%           | Ealy-blind vs. sighted |    | −3.3            | 1.308          | 0.016*    |            |
| 100%          | Ealy-blind vs. sighted |    | −2.9            | 0.057          | 0.017*    |            |

\* $p < 0.05$ , \*\*\* $p < 0.001$ .  
 $\eta^2_G = \frac{SS_{effect}}{SS_{effect} + SS_{subjects} + SS_{error}}$  (Small: < 0.01, Medium: 0.01~0.06, Large: 0.14).

3.3 Correlation of neural response with behavioral data

We determined the correlations between AM or FM vocoded speech recognition and neural responses of the N2 and P3b components regarding latency and amplitude. A significant correlation was observed between the P3b latency and behavioral accuracy in AM vocoded speech recognition ( $r = -0.316$ ,  $p < 0.001$ ; Figure 6, left panel). Significant correlations were found between the N2 amplitude and behavioral accuracy in FM vocoded speech perception ( $r = 0.294$ ,  $p = 0.001$ ). Likewise, the P3b peak latency and amplitude exhibited significant correlations with behavioral accuracy (latency:  $r = -0.315$ ,  $p < 0.001$ ; amplitude:  $r = 0.293$ ,  $p = 0.001$ ; Figure 6, right panel).

4 Discussion

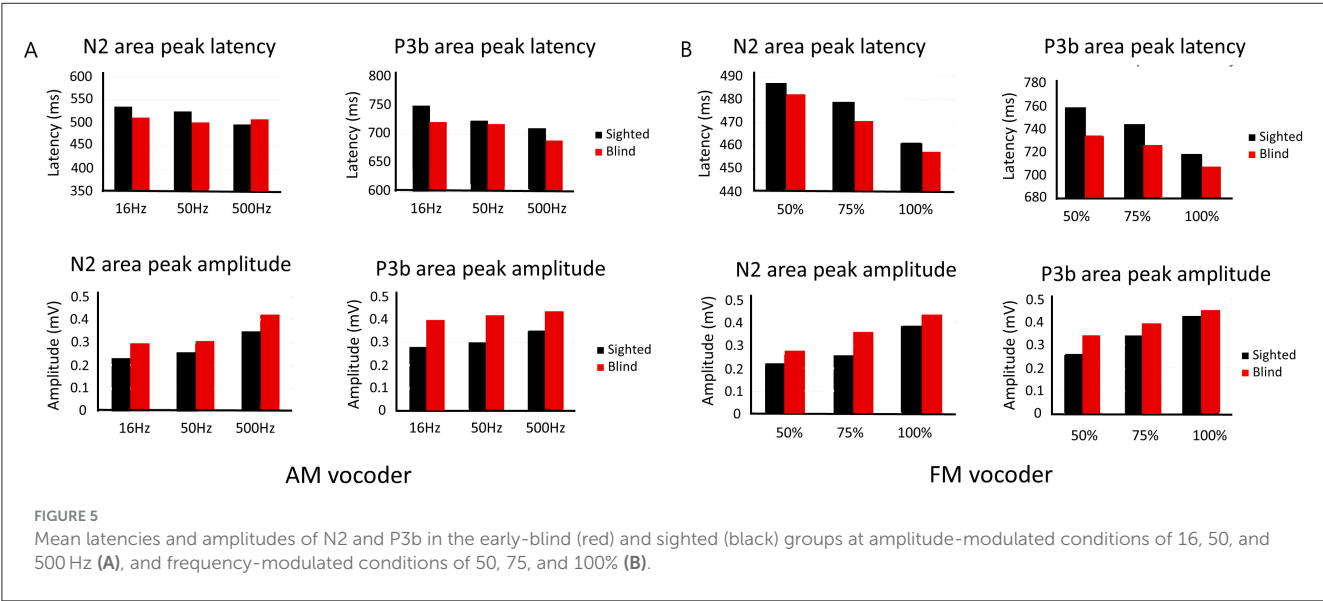
We investigated the effects of degraded temporal cues on speech recognition and semantic processing in early-blind individuals compared with sighted subjects. Our findings showed that early-blind participants demonstrated better speech recognition performance across almost all conditions, even with degradation of the temporal envelope and TFS, which is less detrimental for early-blind individuals. Furthermore, the P3b responses indicated that early-blind individuals may have enhanced cortical mechanisms for semantic processing in the case of degraded temporal cues. Supporting this notion, several studies have reported that early-blind individuals better utilize temporal cues compared with sighted individuals, including the processing of temporal-order judgment (Weaver and Stevens, 2006), temporal modulation



TABLE 4 Statistical summary of the effect of envelope cutoff frequency on the latency and amplitude of N2 and P3b components.

|                 | Sum of square | df     | Mean square | F      | p         | $\eta^2_G$ |
|-----------------|---------------|--------|-------------|--------|-----------|------------|
| N2              |               |        |             |        |           |            |
| Latency         | 10,181.562    | 1.703  | 5,979.021   | 2.930  | 0.069     | 0.072      |
| Group           | 4,940.833     | 1      | 4,940.833   | 1.363  | 0.250     | 0.035      |
| latency*Group   | 8,167.917     | 1.703  | 4,796.527   | 2.351  | 0.111     | 0.058      |
| Residual        | 1,32,047.396  | 64.709 | 2,040.619   |        |           |            |
| Amplitude       | 0.321         | 1.549  | 0.207       | 7.244  | 0.003*    | 0.160      |
| Group           | 0.119         | 1      | 0.119       | 2.610  | 0.114     | 0.064      |
| Amplitude*Group | 0.002         | 1.549  | 0.002       | 0.055  | 0.908     | 0.001      |
| Residual        | 1.683         | 58.881 | 0.029       |        |           |            |
| P3b             |               |        |             |        |           |            |
| Latency         | 25,494.200    | 2      | 12,747.100  | 14.238 | <0.001*** | 0.273      |
| Group           | 10,849.008    | 1      | 10,849.008  | 3.027  | 0.090     | 0.074      |
| Latency*Group   | 3,030.067     | 2      | 1515.033    | 1.692  | 0.191     | 0.043      |
| Residual        | 68,041.067    | 76     | 895.277     |        |           |            |
| Amplitude       | 0.056         | 1.520  | 0.037       | 1.752  | 0.180     | 0.044      |
| Group           | 0.331         | 1      | 0.331       | 4.081  | 0.050     | 0.097      |
| Amplitude*Group | 0.008         | 1.520  | 0.005       | 0.238  | 0.727     | 0.006      |
| Residual        | 1.205         | 57.759 | 0.021       |        |           |            |

\* $p < 0.05$ , \*\*\* $p < 0.001$ .



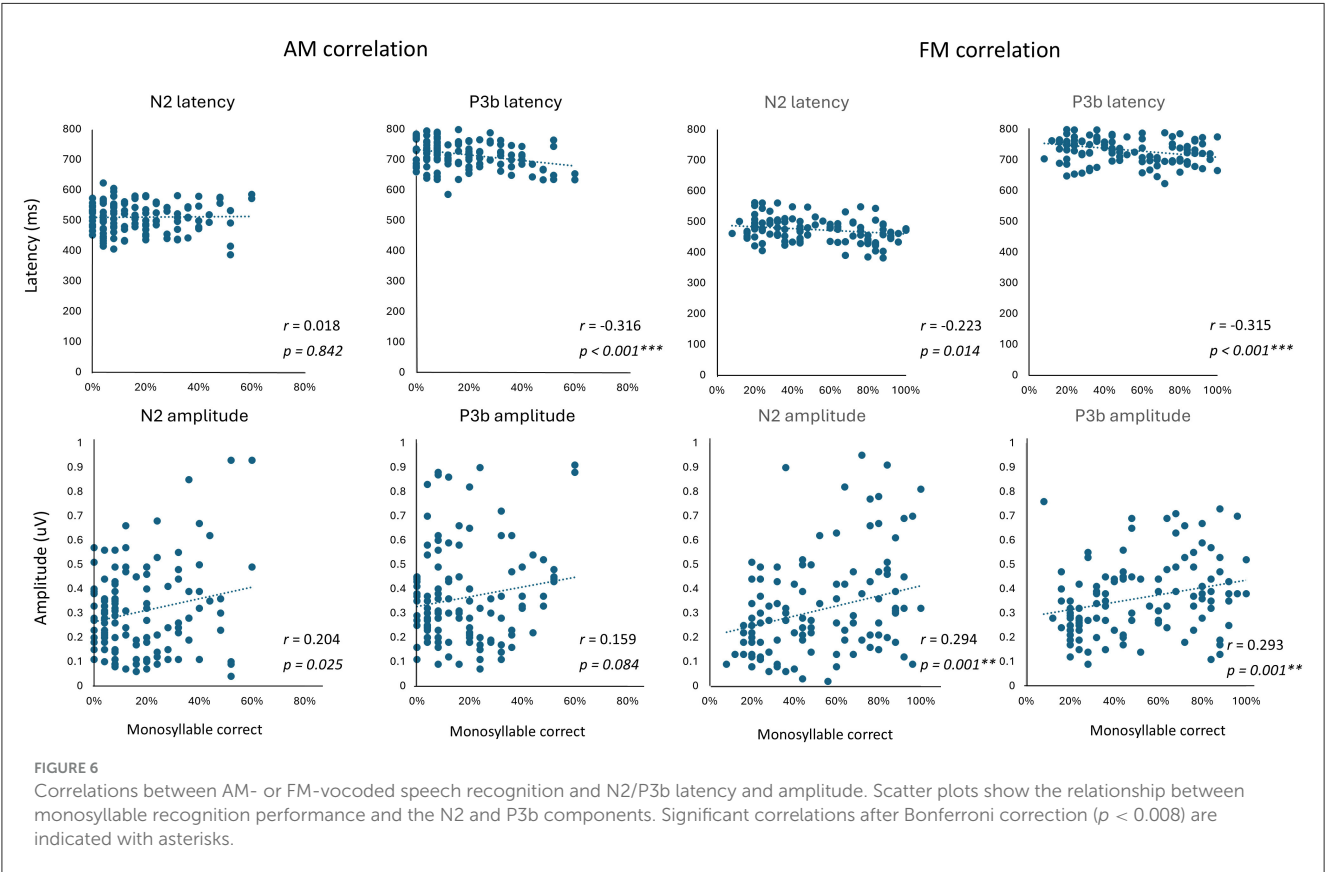
detection (Shim et al., 2019), temporal patterns (Bae et al., 2022), and temporal resolution ability using gap detection (Muchnik et al., 1991). However, some studies found no difference in the gap detection threshold (Weaver and Stevens, 2006; Boas et al., 2011) and temporal bisection (Vercillo et al., 2016; Campus et al., 2019; Gori et al., 2020) between blind and sighted individuals. Several studies have demonstrated that early-blind participants were better at comprehending ultrafast speech (time-compressed speech) than sighted individuals, which underscores the adaptation

of their auditory system to improve the encoding of temporal aspects of acoustic signals (Moos and Trouvain, 2007; Dietrich et al., 2013; Hertrich et al., 2013). Furthermore, both early- and late-blind individuals can acquire enhanced ability for ultrafast speech comprehension (Hertrich et al., 2013) and temporal modulation detection (Shim et al., 2019). Early-blind individuals prioritize temporal information in multidimensional selection tasks, initially selecting events based on timing rather than location, followed by a parallel selection incorporating both temporal and spatial

TABLE 5 Statistical summary of the effect of the amount of temporal fine structure on the latency and amplitude of N2 and P3b components.

|                 | Sum of square | df    | Mean square | <i>F</i> | <i>p</i>  | $\eta^2_G$ |
|-----------------|---------------|-------|-------------|----------|-----------|------------|
| N2              |               |       |             |          |           |            |
| Latency         | 13,051.926    | 2     | 6,525.963   | 8.400    | <0.001*** | 0.181      |
| Group           | 952.033       | 1     | 952.033     | 0.356    | 0.554     | 0.009      |
| latency*Group   | 103.082       | 2     | 51.541      | 0.066    | 0.936     | 0.002      |
| Residual        | 59043.367     | 76    | 776.886     |          |           |            |
| Amplitude       | 0.508         | 2     | 0.254       | 7.812    | <.001***  | 0.171      |
| Group           | 0.133         | 1     | 0.133       | 2.739    | 0.106     | 0.067      |
| Amplitude*Group | 0.019         | 1.922 | 0.010       | 0.287    | 0.751     | 0.007      |
| Residual        | 2.469         | 76    | 0.032       |          |           |            |
| P3b             |               |       |             |          |           |            |
| Latency         | 24,050.317    | 2     | 12,025.158  | 8.734    | <0.001*** | 0.187      |
| Group           | 9,275.208     | 1     | 9,275.208   | 3.202    | 0.082     | 0.078      |
| Latency*Group   | 967.117       | 2     | 483.558     | 0.351    | 0.705     | 0.009      |
| Residual        | 1,04,633.900  | 76    | 1,376.762   |          |           |            |
| Amplitude       | 0.331         | 2     | 0.165       | 15.868   | <0.001*** | 0.295      |
| Group           | 0.081         | 1     | 0.081       | 2.014    | 0.164     | 0.050      |
| Amplitude*Group | 0.022         | 2     | 0.11        | 1.060    | 0.351     | 0.027      |
| Residual        | 0.792         | 76    | 0.010       |          |           |            |

\*\*\**p* < 0.001.



attributes (Röder et al., 2007). The superior utilization of temporal cues in the brain by early-blind individuals compared with sighted individuals is presumed to be a result of compensatory plasticity due to long-term visual deprivation. Numerous neuroimaging studies have shown that blind individuals recruit the visual cortex to perform auditory functions (Leclerc et al., 2000; Gougoux et al., 2009; Collignon et al., 2011; Voss and Zatorre, 2012) and have a thicker visual cortex than sighted individuals (Voss and Zatorre, 2012). In addition, cross-modal plasticity occurs through the enhancement of pre-existing audiovisual connections (Beer et al., 2011; Collignon et al., 2013; Pelland et al., 2017) or the development of new audiovisual connections following the loss of vision (Karlen et al., 2006; Chabot et al., 2008). Synchronization of neuronal populations to the temporal dynamics of speech was observed in the primary visual cortex of early-blind individuals, along with functional connectivity between the temporal and occipital cortices (Van Ackeren et al., 2018). These findings suggest that the brain of blind individuals may adopt an architecture that enables them to track temporal cues, and the cerebrum appears to play a key role in temporal sound processing (Schulze and Langner, 1997; Eggermont, 2002; Bao et al., 2004).

The significant interaction between group and amount of TFS indicates that, while early-blind subjects may exhibit an overall advantage, the effectiveness of TFS shows a greater decrease with the level of degradation, thereby emphasizing the complexity of auditory processing in this population. In contrast, the impact of the envelope on speech recognition did not differ between the two groups, consistent with our previous results (Choi et al., 2024a). Earlier studies used two cutoff frequencies for the envelope cue (50 and 500 Hz), whereas the current study involved three cutoff frequencies (16, 50, and 500 Hz). However, the results were the same among the studies. The sensitivity of early-blind individuals to the reduction of TFS cues underlying the deterioration of speech recognition suggests that their ability to perceive speech in noise may be significantly compromised as they age or develop hearing loss. This is because the efficient use of TFS cues is severely limited with aging and hearing impairment (Lorenzi et al., 2006; Moore et al., 2006; Hopkins and Moore, 2007; Hopkins et al., 2008).

The EEG results provide further insights into the neural correlates of these behavioral findings. The significant effects of the envelope cutoff frequency on the N2 amplitude and P3b latency suggest that the degradation of temporal resolution influences higher-order cognitive processes involved in speech recognition and semantic integration. The amount of TFS showed significant main effects on the amplitude and latency of the N2 and P3b components. In the correlation analysis of neural responses with behavioral data, only one significant correlation was found for AM-vocoded speech, whereas three significant correlations were observed for FM-vocoded speech. This result might reflect a clearer effect of the condition, as observed in the RM ANOVA analysis of the latency and amplitude of N2 and P3b for FM-vocoded speech compared to AM-vocoded speech. The N2 component is associated with lexical information and semantic categorization (Schmitt et al., 2000; Van den Brink and Hagoort, 2004), whereas the P3b component is related to attention and updating working memory (Beynon et al., 2005; Henkin et al., 2015). The N2 and P3b results indicate that, with the degradation of the temporal

envelope or TFS cues, there is an increased reliance on top-down processing for speech recognition. Similar patterns of N2/P3b utilizing the same speech oddball paradigm were observed in the context of degraded auditory spectral cues, which corresponded to reduced semantic integration with spectral degradation (Choi et al., 2024a,b). In adverse listening environments, the brain retrieves word meanings from our mental lexicon, which involves circuits for categorizing words based on their meanings. This process is reflected by a delayed latency and greater amplitude of the P3b component, which varies with the intensity of background noise (Henkin et al., 2008; Finke et al., 2016; Balkenhol et al., 2020). Other studies have also shown that individuals tend to depend more on top-down processing when spectral or temporal information is compromised or in the case of adverse listening conditions (Davis et al., 2005; Peelle and Davis, 2012).

The observed trend toward a significant difference in the P3b amplitude between the early-blind and sighted individuals hints at underlying differences in cognitive processing strategies between these groups, although this finding warrants further exploration with larger sample sizes. This finding could suggest that the brains of blind individuals may react more robustly to higher-order processing, including working memory. In a magnetoencephalography study, enhanced neural synchronization to acoustic fluctuations in early-blind individuals was observed in the theta range (corresponding to the syllabic rate) in the primary visual cortex (Van Ackeren et al., 2018). Furthermore, N2 and P3b were prolonged in cochlear implant users compared with subjects with normal hearing, implicating a slower stimulus evaluation in the former, indicating slower access to lexical information and prolonged word evaluation. This finding highlights the impact of auditory processing on cognitive function (Henkin et al., 2008, 2015; Finke et al., 2016).

To our knowledge, this study is the first to compare speech recognition and relevant cortical-evoked potentials between early-blind and sighted individuals in listening environments involving degradation of the auditory temporal envelope and TFS. The results indicate that preserving TFS is crucial for speech recognition in visually impaired individuals with hearing impairment, thereby providing insights into the auditory rehabilitation of people with visual/auditory impairment. A limitation of this study is that we used vocoded speech to simulate degradation of the temporal envelope and TFS cues in young participants with normal hearing rather than in people with actual temporal resolution deficits. Future research should focus on elderly individuals with both visual and hearing impairments.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving humans were approved by Institutional Review Board of Nowon Eulji Medical Center. The studies were

conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

HC: Writing – original draft, Writing – review & editing, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization. J-SK: Writing – original draft, Writing – review & editing, Formal analysis, Methodology, Software, Validation, Visualization. JW: Writing – review & editing, Methodology, Software. HS: Supervision, Writing – original draft, Writing – review & editing, Conceptualization, Funding acquisition, Validation.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was

supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1I1A3071587).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Bae, E. B., Jang, H., and Shim, H. J. (2022). Enhanced dichotic listening and temporal sequencing ability in early-blind individuals. *Front. Psychol.* 13:840541. doi: 10.3389/fpsyg.2022.840541
- Balkenhol, T., Wallhauser-Franke, E., Rotter, N., and Servais, J. J. (2020). Changes in speech-related brain activity during adaptation to electro-acoustic hearing. *Front. Neurol.* 11:161. doi: 10.3389/fneur.2020.00161
- Bao, S., Chang, E. F., Woods, J., and Merzenich, M. M. (2004). Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. *Nat. Neurosci.* 7, 974–981. doi: 10.1038/nn1293
- Beer, A. L., Plank, T., and Greenlee, M. W. (2011). Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Exp. Brain Res.* 213, 299–308. doi: 10.1007/s00221-011-2715-y
- Beynon, A. J., Snik, A. F., Stegeman, D. F., and van den Broek, P. (2005). Discrimination of speech sound contrasts determined with behavioral tests and event-related potentials in cochlear implant recipients. *J. Am. Acad. Audiol.* 16, 42–53. doi: 10.3766/jaaa.16.1.5
- Boas, L. V., and Muniz, L. Neto, S.d.S.C., and Gouveia, M.d.C.L. (2011). Auditory processing performance in blind people. *Braz. J. Otorhinolaryngol.* 77, 504–509. doi: 10.1590/S1808-86942011000400015
- Borjigin, A., and Bharadwaj, H. M. (2025). Individual differences elucidate the perceptual benefits associated with robust temporal fine-structure processing. *Proc. Natl. Acad. Sci.* 122:e2317152121. doi: 10.1073/pnas.2317152121
- Borjigin, A., Hustedt-Mai, A. R., and Bharadwaj, H. M. (2022). Individualized assays of temporal coding in the ascending human auditory system. *Eneuro* 9:ENEURO.0378-21.2022. doi: 10.1523/ENEURO.0378-21.2022
- Campus, C., Sandini, G., Amadeo, M. B., and Gori, M. (2019). Stronger responses in the visual cortex of sighted compared to blind individuals during auditory space representation. *Sci. Rep.* 9:1935. doi: 10.1038/s41598-018-37821-y
- Chabot, N., Charbonneau, V., Laramée, M.-E., Tremblay, R., Boire, D., Bronchti, G., et al. (2008). Subcortical auditory input to the primary visual cortex in anophthalmic mice. *Neurosci. Lett.* 433, 129–134. doi: 10.1016/j.neulet.2008.01.003
- Choi, H. J., Kyong, J.-S., Lee, J. H., Han, S. H., and Shim, H. J. (2024a). The impact of spectral and temporal degradation on vocoded speech recognition in early-blind individuals. *Eneuro* 11. doi: 10.1523/ENEURO.0528-23.2024
- Choi, H. J., Kyong, J.-S., Won, J. H., and Shim, H. J. (2024b). Effect of spectral degradation on speech intelligibility and cortical representation. *Front. Neurosci.* 18:1368641. doi: 10.3389/fnins.2024.1368641
- Collignon, O., Dormal, G., Albouy, G., Vandewalle, G., Voss, P., Phillips, C., et al. (2013). Impact of blindness onset on the functional organization and the connectivity of the occipital cortex. *Brain* 136, 2769–2783. doi: 10.1093/brain/awt176
- Collignon, O., Vandewalle, G., Voss, P., Albouy, G., Charbonneau, G., Lassonde, M., et al. (2011). Functional specialization for auditory-spatial processing in the occipital cortex of congenitally blind humans. *Proc. Natl. Acad. Sci.* 108, 4435–4440. doi: 10.1073/pnas.1013928108
- Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *J. Exp. Psychol. Gen.* 134:222. doi: 10.1037/0096-3445.134.2.222
- Deacon, D., Breton, F., Ritter, W., and Vaughan Jr, H. G. (1991). The relationship between N2 and N400: Scalp distribution, stimulus probability, and task relevance. *Psychophysiology* 28, 185–200. doi: 10.1111/j.1469-8986.1991.tb00411.x
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Dietrich, S., Hertrich, I., and Ackermann, H. (2013). Ultra-fast speech comprehension in blind subjects engages primary visual cortex, fusiform gyrus, and pulvinar—a functional magnetic resonance imaging (fMRI) study. *BMC Neurosci.* 14, 1–15. doi: 10.1186/1471-2202-14-74
- Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064. doi: 10.1121/1.408467
- Eggermont, J. J. (2002). Temporal modulation transfer functions in cat primary auditory cortex: separating stimulus effects from neural mechanisms. *J. Neurophysiol.* 87, 305–321. doi: 10.1152/jn.00490.2001
- Evans, S., Kyong, J., Rosen, S., Golestani, N., Warren, J., McGettigan, C., et al. (2014). The pathways for intelligible speech: multivariate and univariate perspectives. *Cereb. Cort.* 24, 2350–2361. doi: 10.1093/cercor/bht083
- Faulkner, A., Rosen, S., and Green, T. (2012). Comparing live to recorded speech in training the perception of spectrally shifted noise-vocoded speech. *J. Acoust. Soc. Am.* 132, EL336–EL342. doi: 10.1121/1.4754432
- Finke, M., Büchner, A., Ruigendijk, E., Meyer, M., and Sandmann, P. (2016). On the relationship between auditory cognition and speech intelligibility in cochlear implant users: an ERP study. *Neuropsychologia* 87, 169–181. doi: 10.1016/j.neuropsychologia.2016.05.019
- Folstein, J. R., and Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170. doi: 10.1111/j.1469-8986.2007.00602.x
- Glasberg, B. R., and Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 47, 103–138. doi: 10.1016/0378-5955(90)90170-T

- Gnansia, D., Péan, V., Meyer, B., and Lorenzi, C. (2009). Effects of spectral smearing and temporal fine structure degradation on speech masking release. *J. Acoust. Soc. Am.* 125, 4023–4033. doi: 10.1121/1.3126344
- Gori, M., Amadeo, M. B., and Campus, C. (2020). Temporal cues trick the visual and auditory cortices mimicking spatial cues in blind individuals. *Hum. Brain Mapp.* 41, 2077–2091. doi: 10.1002/hbm.24931
- Gougoux, F., Belin, P., Voss, P., Lepore, F., Lassonde, M., Zatorre, R. J., et al. (2009). Voice perception in blind persons: a functional magnetic resonance imaging study. *Neuropsychologia* 47, 2967–2974. doi: 10.1016/j.neuropsychologia.2009.06.027
- Haggard, M. (1984). “Temporal patterning in speech: The implications of temporal resolution and signal-processing”, in *Time Resolution in Auditory Systems: Proceedings of the 11th Danavox Symposium on Hearing Gamle Avernæs, Denmark, August 28–31, 1984* (Berlin; Heidelberg: Springer), 215–237. doi: 10.1007/978-3-642-70622-6\_13
- Heng, J., Cantarero, G., Elhilali, M., and Limb, C. J. (2011). Impaired perception of temporal fine structure and musical timbre in cochlear implant users. *Hear. Res.* 280, 192–200. doi: 10.1016/j.heares.2011.05.017
- Henkin, Y., Tetin-Schneider, S., Hildesheimer, M., and Kishon-Rabin, L. (2008). Cortical neural activity underlying speech perception in postlingual adult cochlear implant recipients. *Audiol. Neurotol.* 14, 39–53. doi: 10.1159/000153434
- Henkin, Y., Yaar-Soffer, Y., Steinberg, M., and Muchnik, C. (2015). Neural correlates of auditory-cognitive processing in older adult cochlear implant recipients. *Audiol. Neurotol.* 19, 21–26. doi: 10.1159/000371602
- Hertrich, I., Dietrich, S., and Ackermann, H. (2013). How can audiovisual pathways enhance the temporal resolution of time-compressed speech in blind subjects? *Front. Psychol.* 4:530. doi: 10.3389/fpsyg.2013.00530
- Hilbert, D. (1912). “Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen,” in *Integralgleichungen und Gleichungen mit unendlich vielen Unbekannten* (Springer), 8–171. doi: 10.1007/978-3-322-84410-1\_1
- Hopkins, K., and Moore, B. C. (2007). Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *J. Acoust. Soc. Am.* 122, 1055–1068. doi: 10.1121/1.2749457
- Hopkins, K., Moore, B. C., and Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *J. Acoust. Soc. Am.* 123, 1140–1153. doi: 10.1121/1.2824018
- Houtsma, A. J., and Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.* 87, 304–310. doi: 10.1121/1.399297
- Hyvärinen, A., and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw.* 13, 411–430. doi: 10.1016/S0893-6080(00)00026-5
- Karlen, S., Kahn, D., and Krubitzer, L. (2006). Early blindness results in abnormal corticocortical and thalamocortical connections. *Neuroscience* 142, 843–858. doi: 10.1016/j.neuroscience.2006.06.055
- Klem, G. H. (1999). The ten-twenty electrode system of the international federation. The international federation of clinical neurophysiology. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 52, 3–6.
- Leclerc, C., Saint-Amour, D., Lavoie, M. E., Lassonde, M., and Lepore, F. (2000). Brain functional reorganization in early blind humans revealed by auditory event-related potentials. *Neuroreport* 11, 545–550. doi: 10.1097/00001756-200002280-00024
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc. Natl. Acad. Sci.* 103, 18866–18869. doi: 10.1073/pnas.0607364103
- Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.
- Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D., and Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci.* 94, 10979–10984. doi: 10.1073/pnas.94.20.10979
- McFarlane, K. A., and Sanchez, J. T. (2024). Effects of temporal processing on speech-in-noise perception in middle-aged adults. *Biology* 13:371. doi: 10.3390/biology13060371
- McKay, C. M., Lim, H. H., and Lenarz, T. (2013). Temporal processing in the auditory system: insights from cochlear and auditory midbrain implantations. *J. Assoc. Res. Otolaryngol.* 14, 103–124. doi: 10.1007/s10162-012-0354-z
- Meyer, M., Elmer, S., Ringli, M., Oechslin, M. S., Baumann, S., Jancke, L., et al. (2011). Long-term exposure to music enhances the sensitivity of the auditory system in children. *Eur. J. Neurosci.* 34, 755–765. doi: 10.1111/j.1460-9568.2011.07795.x
- Moon, I. J., Won, J. H., Park, M.-H., Ives, D. T., Nie, K., Heinz, M. G., et al. (2014). Optimal combination of neural temporal envelope and fine structure cues to explain speech identification in background noise. *J. Neurosci.* 34, 12145–12154. doi: 10.1523/JNEUROSCI.1025-14.2014
- Moore, B. C. (1973). Frequency difference limens for short-duration tones. *J. Acoust. Soc. Am.* 54, 610–619. doi: 10.1121/1.1913640
- Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *J. Assoc. Res. Otolaryngol.* 9, 399–406. doi: 10.1007/s10162-008-0143-x
- Moore, B. C., Glasberg, B. R., and Hopkins, K. (2006). Frequency discrimination of complex tones by hearing-impaired subjects: evidence for loss of ability to use temporal fine structure. *Hear. Res.* 222, 16–27. doi: 10.1016/j.heares.2006.08.007
- Moos, A., and Trouvain, J. (2007). “Comprehension of ultra-fast speech-blind vs. “normally hearing” persons”, in *Proceedings of the 16th International Congress of Phonetic Sciences* (Germany: Saarland University Saarbrücken Germany), 677–680.
- Muchnik, C., Efrati, M., Nemeth, E., Malin, M., and Hildesheimer, M. (1991). Central auditory skills in blind and sighted subjects. *Scand. Audiol.* 20, 19–23. doi: 10.3109/01050399109070785
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comp. Intell. Neurosci.* 2011, 1–9. doi: 10.1155/2011/156869
- Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. *Trends Amp.* 12, 316–331. doi: 10.1177/1084713808325881
- Oxenham, A. J., and Simonson, A. M. (2009). Masking release for low-and high-pass-filtered speech in the presence of noise and single-talker interference. *J. Acoust. Soc. Am.* 125, 457–468. doi: 10.1121/1.3021299
- Peelle, J. E., and Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* 3:320. doi: 10.3389/fpsyg.2012.00320
- Pelland, M., Orban, P., Dansereau, C., Lepore, F., Bellec, P., Collignon, O., et al. (2017). State-dependent modulation of functional connectivity in early blind individuals. *Neuroimage* 147, 532–541. doi: 10.1016/j.neuroimage.2016.12.053
- Perrin, F., Pernier, J., Bertrand, O., and Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* 72, 184–187. doi: 10.1016/0013-4694(89)90180-6
- Petermann, M., Kummer, P., Burger, M., Lohscheller, J., Eysholdt, U., Döllinger, M., et al. (2009). Statistical detection and analysis of mismatch negativity derived by a multi-deviant design from normal hearing children. *Hear. Res.* 247, 128–136. doi: 10.1016/j.heares.2008.11.001
- Qin, M. K., and Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear Hear.* 26, 451–460. doi: 10.1097/01.aud.0000179689.79868.06
- Röder, B., Krämer, U. M., and Lange, K. (2007). Congenitally blind humans use different stimulus selection strategies in hearing: an ERP study of spatial and temporal attention. *Restorat. Neurol. Neurosci.* 25, 311–322. doi: 10.3233/RNN-2007-253413
- Schmitt, B. M., Münte, T. F., and Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology* 37, 473–484. doi: 10.1111/1469-8986.3740473
- Schulze, H., and Langner, G. (1997). Periodicity coding in the primary auditory cortex of the Mongolian gerbil (*Merionesunguiculatus*): two different coding strategies for pitch and rhythm? *J. Comp. Physiol. A* 181, 651–663. doi: 10.1007/s003590050147
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304. doi: 10.1126/science.270.5234.303
- Shim, H. J., Go, G., Lee, H., Choi, S. W., and Won, J. H. (2019). Influence of visual deprivation on auditory spectral resolution, temporal resolution, and speech perception. *Front. Neurosci.* 13:1200. doi: 10.3389/fnins.2019.01200
- Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416, 87–90. doi: 10.1038/416087a
- Van Ackeren, M. J., Barbero, F. M., Mattioni, S., Bottini, R., and Collignon, O. (2018). Neuronal populations in the occipital cortex of the blind synchronize to the temporal dynamics of speech. *elife* 7:e31640. doi: 10.7554/eLife.31640
- Van den Brink, D., and Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *J. Cognit. Neurosci.* 16, 1068–1084. doi: 10.1162/0898929041502670
- Vercillo, T., Burr, D., and Gori, M. (2016). Early visual deprivation severely compromises the auditory sense of space in congenitally blind children. *Dev. Psychol.* 52:847. doi: 10.1037/dev0000103
- Voola, M., Wedekind, A., Nguyen, A. T., Marinovic, W., Rajan, G., Tavora-Vieira, D., et al. (2023). Event-related potentials of single-sided deaf cochlear implant users: using a semantic oddball paradigm in noise. *Audiol. Neurotol.* 28, 280–293. doi: 10.1159/000529485
- Voss, P., and Zatorre, R. J. (2012). Occipital cortical thickness predicts performance on pitch and musical tasks in blind individuals. *Cereb. Cortex* 22, 2455–2465. doi: 10.1093/cercor/bhr311



Wang, S., Dong, R., Liu, D., Wang, Y., Liu, B., Zhang, L., et al. (2015). The role of temporal envelope and fine structure in Mandarin lexical tone perception in auditory neuropathy spectrum disorder. *PLoS One* 10:e0129710. doi: 10.1371/journal.pone.0129710

Weaver, K. E., and Stevens, A. A. (2006). Auditory gap detection in the early blind. *Hear. Res.* 211, 1–6. doi: 10.1016/j.heares.2005.08.002

Xu, L., and Pfingst, B. E. (2003). Relative importance of temporal envelope and fine structure in lexical-tone perception (L). *J. Acoust. Soc. Am.* 114, 3024–3027. doi: 10.1121/1.1623786

Yin, T. C., and Chan, J. C. (1990). Interaural time sensitivity in medial superior olive of cat. *J. Neurophysiol.* 64, 465–488. doi: 10.1152/jn.1990.64.2.465