How Do Native and Non-native Listeners Differ? Investigation with Dominant Frequency Bands in Auditory Evoked Potential

Yifan Zhou[1], Md Rakibul Hasan[2], Md Mahbub Hasan[3], Ali Zia[1], and Md Zakir Hossain[1,2](✉)

[1] Australian National University, Canberra, ACT 2600, Australia
{Ali.Zia,Zakir.Hossain}@anu.edu.au
[2] Curtin University, Perth, WA 6102, Australia
{Rakibul.Hasan,Zakir.Hossain1}@curtin.edu.au
[3] Khulna University of Engineering & Technology, Khulna 9203, Bangladesh
mahbub01@eee.kuet.ac.bd

Abstract. EEG signal provides valuable insights into cortical responses to specific exogenous stimuli, including auditory and visual stimuli. This study investigates the evoked potential in EEG signals and dominant frequency bands for native and non-native subjects. Songs in different languages are played to subjects using conventional in-ear phones or bone-conducting devices. Time-frequency analysis is performed to characterise induced and evoked responses in the EEG signal, focusing on the phase synchronisation level of the evoked potential as a significant feature. Phase locking value (PLV) and weighted phase lag index (WPLI) are used to assess the phase synchrony between the EEG signal and sound signal, while the frequency-dependent effective gain is analysed to understand its impact. The results demonstrate that native subjects experience higher levels of evoked potential, indicating more complex cognitive neural processes compared to non-native subjects. Dominant frequency windows associated with higher levels of evoked potential are identified using a peak-picking algorithm. Interestingly, the choice of playing device has minimal influence on the evoked potential, suggesting similar outcomes with both in-ear phones and bone-conducting devices. This study provides valuable insights into the neural processing differences between native and non-native subjects and highlights the potential impact of playing devices on the evoked potential.

Keywords: EEG · phase locking value · weighted phase lag index · dominant frequency

Y. Zhou and M. R. Hasan—These authors contributed equally to this work.

B. Luo et al. (Eds.): ICONIP 2023, CCIS 1963, pp. 350–361, 2024.
https://doi.org/10.1007/978-981-99-8138-0_28

1 Introduction

The human brain possesses a remarkable capacity to perceive external stimuli, such as auditory and visual cues. This perceptual process manifests itself through discernible patterns of neural activity, which can be effectively characterised through electroencephalogram (EEG) signals [5]. EEG recordings capture cortical oscillatory activities, which can be classified into evoked, induced and ongoing responses [6,16]. Among these, auditory evoked potentials have emerged as a valuable tool for scrutinising the intricacies of the human auditory system, particularly within the hearing-impaired population [9]. These potentials provide insights into the auditory process by analysing EEG signals in response to auditory stimuli. Auditory evoked potentials can be categorised as early, middle and late latency responses, which are determined by the temporal relationship between the cessation of auditory stimuli and the onset of corresponding brain waves. It has been observed that early and middle latency responses are devoid of cognitive involvement, whereas late responses exhibit engagement in cognitive processes such as discriminating pure tones from complex acoustic signals like speech [1]. In addition to evoked potentials, induced potentials that capture more intricate cognitive processes within the brain can also be discerned through EEG recordings. The utilisation of mathematical methodologies to analyse EEG signals plays a crucial role in exploring neural activities within the human brain in response to external stimuli [17].

The characterisation of evoked and induced responses within EEG signals involves comparing the amplitude at a specific time relative to the baseline amplitude. An evoked response is identified when the amplitude surpasses the baseline level subsequent to the stimulus presentation and consistently exhibits a fixed latency delay relative to the stimulus onset. In contrast, induced responses are distinguished by segments of time following the stimulus that display higher amplitudes compared to the baseline without adhering to a fixed temporal latency. It is crucial to account for different frequency bands as evoked and induced responses exhibit more pronounced manifestations within specific frequency ranges, contingent upon the nature of the stimulus. Thus, time-frequency analysis [14] assumes a critical role as an initial step in identifying and examining evoked and induced responses. This paper leverages time-frequency analysis for subsequent signal processing and to determine the frequency bands associated with evoked responses.

Investigating the differences in neural activity patterns between native and non-native listeners has garnered significant attention in neuroscience and has implications for developing machine-brain interfaces [11,19]. This study analyse EEG signals recorded from subjects while playing various sound signals, along with the corresponding sound signal itself. We employ mathematical analysis methods, including calculating frequency based on phase locking value (PLV) [3], weighted phase lag index (WPLI) [10], and effective gain, to identify the dominant frequency band for evoked potentials in native listeners listening to either native or non-native songs. The PLV quantifies the phase difference variability between two signals at a specific time and frequency, enabling the measurement of synchronicity. It is computed by evaluating the phase angle difference derived from the Short-Time Fourier Transform (STFT) outcomes of the two signals. WPLI assess the asymmetry of the phase difference distribution between the signals. The frequency resolution of the STFT is contingent upon the length of the sampling window, necessitating appropriate adjustments to attain the desired resolution. Consequently, a *downsampling factor* is computed based on the desired and modified frequency resolutions, ensuring that the modified window length for both EEG and sound data is an integer value. An algorithm is proposed to determine the optimal *downsampling factor*.

2 Materials and Methods

2.1 Sound and EEG Data

We utilised a publicly available dataset [2], sourced from PhysioNet [7]. The dataset consists of more than 240 two-minute EEG recordings obtained from 20 subjects. The recordings were conducted using an OpenBCI Ganglion Board, where four channels (T7, F8, Cz and P4) were sampled at a rate of 200 Hz.

The EEG signals were recorded in two different conditions: resting state and auditory stimuli. During the resting state, the EEG signals were captured with both eyes open and eyes closed. For auditory stimuli, the study consisted of two types of auditory devices (in-ear phones and bone-conducting devices) and three categories of songs (native, non-native and neutral). Accordingly, EEG signals were recorded with six experimental conditions involving the combinations of these devices and songs. For the non-native song specification, Italian subjects listened to Arabic songs, while non-Italian subjects listened to Italian songs.

2.2 Data Processing and Analysis

The data processing for this study involves several steps conducted using Python. In Step 1, STFT is applied to both the EEG signal and the corresponding sound signal for all experimental conditions and trials, resulting in time-frequency distribution maps. In Step 2, the phase is extracted from these maps at each time-frequency point for each condition and trial. Step 3 involves calculating the phase difference at each time-frequency point. Step 4 computes the normalised effective gain, PLV and WPLI using the phase difference and values from the time-frequency distribution. Moving to Step 5, the metrics obtained in Step 4 are averaged over all time points to obtain the frequency-dependent effective gain, PLV and WPLI for each experimental condition. In Step 6, these frequency-dependent metrics are averaged over all trials for each condition. Finally, Step 7 identifies the dominant frequency bands with higher effective gain and extracts the peak values of PLV and WPLI at these bands for each experimental condition. Through these steps, the study enables the analysis and comparison of frequency-dependent metrics and the identification of dominant frequency bands in the EEG and sound signals across different experimental conditions.

Algorithm 1. The algorithm for choosing the *downsampling factor*

Input: desired frequency resolution, sampling frequency for EEG signal, sampling frequency for sound signal
Output: downsampling factor
1: downsampling factor = 1
2: **while** True **do**
3: sound time window length = $\frac{\text{sound sampling frequency}}{\text{modified frequency resolution}}$
4: EEG time window length = $\frac{\text{EEG sampling frequency}}{\text{modified frequency resolution}}$
5: **if** sound and EEG time window length are integers **then**
6: break
7: **else**
8: downsampling factor = downsampling factor + 1
9: **end if**
10: **end while**

Downsampling Factor. The frequency resolution of STFT is determined by the frequency resolution of the Fast Fourier Transform corresponding to each window size. To achieve a desired frequency resolution lower than the acquired resolution, the sample window length is increased to maintain an integer ratio between the desired and modified resolutions. This ratio, referred to as the *downsampling factor*, remains the same for both sound and EEG data. Consequently, the modified frequency resolution and sampling window length should also be integers. The *downsampling factor* is determined through Algorithm 1.

Effective Gain, PLV and WPLI. STFT generates a 2D time-frequency map where each time-frequency point corresponds to a complex value, representing the phase angle [4]. The STFT of a speech signal $x(t)$ and an EEG signal $y(t)$ can be expressed as:

$$X(\tau,\omega) = \sum_{t=0}^{L} x(t)W(t-\tau)e^{-i\frac{\omega t}{L}} \qquad Y(\tau,\omega) = \sum_{t=0}^{L} y(t)W(t-\tau)e^{-i\frac{\omega t}{L}} \tag{1}$$

where τ is the centre of the window, W is the window function, L is the length of the window and ω is the frequency.

The wrapped phase of the sound and EEG can be expressed as:

$$\Delta\phi_x = \tan^{-1}\left[\frac{\text{imag}(X(\tau,\omega))}{\text{real}(X(\tau,\omega))}\right] \qquad \Delta\phi_y = \tan^{-1}\left[\frac{\text{imag}(Y(\tau,\omega))}{\text{real}(Y(\tau,\omega))}\right] \tag{2}$$

Both of these phases are unwrapped to get the continuous phases:

$$\phi_x(\tau) = \sum_{t=0}^{\tau} \Delta\phi_x(t) \qquad \phi_y(\tau) = \sum_{t=0}^{\tau} \Delta\phi_y(t) \tag{3}$$

Now, the effective gain at time τ can be expressed as:

$$|H_{eff}(\tau,\omega)| = \frac{|Y(\tau,\omega)|}{|X(\tau,\omega)|} \tag{4}$$

A higher effective gain value is indicative of a stronger perception of the EEG signal in response to the sound stimulus. The time-average effective gain:

$$|H_{eff}(\omega)| = \frac{1}{N}\sum_{\tau=0}^{N-1}|H_{eff}(\tau,\omega)| \tag{5}$$

where N is the total number of samples.

PLV calculation requires the phase angle of the same frequency component from two different signals at different times:

$$\mathrm{PLV}_{xy}(\omega) = \frac{1}{N}\sum_{\tau=0}^{N-1}\exp(i\theta(\tau,\omega)) \tag{6}$$

where $\theta(\tau,\omega) = \phi_x(\tau,\omega) - \phi_y(\tau,\omega)$

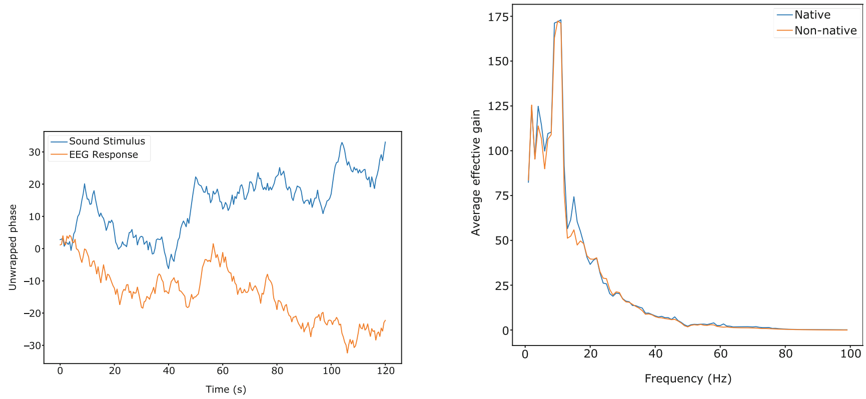
PLV serves as a valuable metric for assessing the inter-trial variability of the phase difference between two signals at a specific time and frequency [4,12]. By calculating the average over multiple trials, the PLV provides insights into the degree of phase synchronisation between the signals. The PLV ranges between 0 and 1. A PLV of 1 signifies a constant phase difference consistently present at $\frac{\pi}{2}$ or $\frac{3\pi}{2}$ radians. Conversely, a PLV of 0 indicates that the phase difference is uniformly distributed across the range of [0, 2π] radians, suggesting a lack of synchronisation between the signals.

WPLI shares a similar purpose to the PLV, as it evaluates the distribution patterns of phase differences between $x(t)$ and $y(t)$ at a specific time and frequency. Its primary focus lies in quantifying the asymmetry within the distribution of phase differences, providing valuable insights into the synchronisation characteristics of the signals. Assuming $S(\omega) = X(\tau,\omega)Y^*(\tau,\omega)$, where $*$ represent complex-conjugate, WPLI can be expressed as:

$$\mathrm{WPLI}(\omega) = \frac{|\mathbb{E}\{|S(\omega)|\,\mathrm{sgn}(S(\omega))\}|}{\mathbb{E}\{|S(\omega)|\}} \tag{7}$$

where $\mathbb{E}$ is the expectation operator and sgn is the sign operator. When the phase differences of all trials share the same sign, the WPLI value becomes 1. As a result, a WPLI value closer to 0 indicates a more symmetric distribution of phase differences.

The evoked potential of EEG signals is characterised by both time-locked and phase-locked responses to external stimuli. In general, higher values of WPLI and PLV indicate a more concentrated distribution of phase differences, reflecting increased phase synchronisation at specific time-frequency points. However, solely relying on elevated PLV and WPLI values at a particular time-frequency point is insufficient for identifying the presence of evoked potentials. It is also crucial to consider the corresponding effective gain value. Generally, evoked potentials are more likely to be located at time-frequency points with relatively higher values of all three metrics: effective gain, PLV and WPLI [12,18].



(a) (b)

Fig. 1. (a) Phase waveform of a typical sound and an EEG signal at 20 Hz and (b) Normalised effective gain for native and non-native cases using conventional in-ear phones (bone-conducting devices also result in a very similar trend).

3 Results and Discussion

3.1 Unwrapped Phase Waveform

The phase of both EEG and sound signals is obtained by extracting the phase values at each time-frequency point from the STFT results of respective signals. Figure 1a illustrates the phase values at a frequency of 20 Hz for both the EEG and sound signals over a time span of 0 to 120 s. The EEG signal consistently exhibits lower phase values compared to the sound signal. Notably, the phase difference between the two signals demonstrates an increasing trend, starting from approximately 0 at 0 s and reaching approximately 60 at 120 s.

3.2 Frequency-Dependent Normalised Effective Gain

The time-frequency distribution involves applying STFT in each time window to obtain the time-specified spectrum. The sampling rate of the EEG signal is 200 Hz. So, according to the Nyquist-Shannon sampling theorem, a maximum frequency of 100 Hz is acquired in the time-frequency distribution. In the case of the sound signal, a higher sampling frequency (>10,000 Hz) results from a wider frequency range in the extracted spectrum. However, to align with the frequency range of interest in the corresponding EEG signal, only components within 0 to 100 Hz are considered from the sound signal's time-frequency distribution.

The frequency-dependent normalised effective gain for native and non-native cases using conventional in-ear phones is depicted in Fig. 1b. The trends are similar for both native and non-native cases. It is evident that the frequency range from 0 Hz to 40 Hz has relatively higher values compared to other frequencies.

Table 1. Dominant frequency bands for auditory evoked potential characterised by phase locking value (PLV). Here, $H(j\omega)$ is the normalised effective gain.

Device	Native			Non-native		
	Freq. (Hz)	$H(j\omega)$	Peak PLV	Freq. (Hz)	$H(j\omega)$	Peak PLV
In-ear	8–10	**171.608**	0.066	6–8	109.982	0.071
	17–19	44.540	0.063	8–10	**171.608**	0.068
	21–23	36.004	**0.070**	10–12	132.346	0.064
	28–30	18.892	0.070	15–17	57.704	**0.080**
	30–32	15.941	0.067	17–19	44.540	0.068
				19–21	37.749	0.070
				30–32	15.941	0.067
				33–35	13.225	0.072
				38–40	8.281	0.071
Bone-conducting	1–3	109.952	0.068	2–4	110.126	0.078
	7–9	140.798	0.076	6–8	109.982	0.066
	9–11	**172.493**	**0.076**	9–11	**172.493**	0.061
	11–13	74.092	0.066	12–14	59.001	0.072
	15–17	57.704	0.068	16–18	51.726	0.072
	25–27	19.576	0.063	19–21	37.749	**0.080**
	27–29	20.429	0.063	22–24	29.038	0.066
	32–34	13.632	0.064	24–26	23.050	0.068
				33–35	13.225	0.060
				36–38	9.413	0.064

As our focus is on the dominant frequency bands that normally exhibit higher levels of evoked potential, the frequency range of 0 Hz to 40 Hz is selected for subsequent analysis.

3.3 Dominant Frequency Characterised by Phase Locking Value

Table 1 summarises dominant frequency bands and corresponding normalised effective gain and phase locking value (PLV) peaks for both native and non-native experiments.

Native Listener. There are five PLV peaks observed in the in-ear experiment, while the bone-conducting experiment exhibits eight peaks in the frequency range of 0 Hz to 40 Hz (Table 1). The maximum value of average normalised effective gain (171.608) for the in-ear experiment is obtained in the frequency band of 8 Hz to 10 Hz, and the corresponding peak PLV is 0.066. The highest peak PLV value (0.070) is observed in the frequency band of 21 Hz to 23 Hz, with a corresponding average normalised effective gain of 36.004. For the bone-conducting experiment, the maximum value of average normalised effective gain (172.493) is observed in the frequency band of 9 Hz to 11 Hz. The corresponding peak PLV, which is also the highest value, is 0.076.

A higher value of PLV implies a lower level of phase difference between the EEG and sound signals. Additionally, a higher value of effective gain represents a higher ratio of EEG amplitude to sound amplitude, indicating a stronger

EEG response at frequencies where evoked potentials occur. Compared to those who used conventional in-ear phones, the bone-conducting group shows a higher number of peaking PLV values, larger peak PLV values and larger effective gain corresponding to the highest peak PLV. This suggests that native subjects using bone-conducting devices experience a relatively higher level of evoked potential compared to those using conventional in-ear phones.

Non-native Listener. In the case of non-native experiments with conventional in-ear phones, the average normalised effective gain is highest (171.608) in the frequency band of 8 Hz to 10 Hz, with a corresponding peak PLV of 0.068 (Table 1). The highest peak PLV value (0.080) occurs at 15 Hz to 17 Hz, with an average normalised effective gain of 57.704.

In the case of bone-conducting devices, the maximum average normalised effective gain is obtained in the frequency band of 9 Hz to 11 Hz, with a corresponding peak PLV of 0.061, which is slightly lower than that observed in conventional in-ear phones. The highest peak PLV value (0.080) is achieved at 19 Hz to 21 Hz, with an average normalised effective gain of 37.749. Comparing the data with in-ear phones, we observe an additional peaking PLV. However, the largest peaking PLV is very similar to that observed in conventional in-ear phones.

Comparison Between Native and Non-native Listeners. Compared to native subjects using conventional in-ear phones, the non-native subjects exhibit higher peaking PLV, higher maximum peaking PLV and corresponding effective gain values. Additionally, the peaking value corresponding to the highest normalised effective gain for non-native subjects using conventional in-ear phones is higher than that observed in native subjects.

The non-native subjects using either conventional in-ear phones or bone-conducting devices have higher peaking PLV values across most sub-frequency bands compared to native subjects, especially where the effective gain reaches its maximum. This suggests that non-native subjects experience relatively higher evoked potentials (characterised by PLV) within the dominant frequency range. Since the bone-conducting devices exhibit a very similar level of evoked potential to conventional devices for both native and non-native experiments, it can be said that the choice of different playing devices has minimal impact on the evoked potential. With the higher peaking PLV value, it can be inferred that subjects tend to focus on a broader range of frequencies when listening to non-native songs compared to native subjects, regardless of the playing device used.

3.4 Dominant Frequency Characterised by Weighted Phase Lag Index

Table 2 summarises the dominant frequency bands characterised by weighted phase lag index (WPLI), with corresponding normalised effective gain and WPLI peaks for both native and non-native experiments.

Table 2. Dominant frequency bands for auditory evoked potential characterised by weighted phase lag index (WPLI). Here, $H(j\omega)$ is the normalised effective gain.

Device	Native			Non-native		
	Freq. (Hz)	$H(j\omega)$	Peak WPLI	Freq. (Hz)	$H(j\omega)$	Peak WPLI
In-ear	2–4	110.126	0.121	2–4	110.126	0.106
	4–6	107.132	0.113	5–7	104.708	0.107
	7–9	**140.798**	0.093	7–9	140.798	0.109
	10–12	132.346	0.124	9–11	**172.493**	0.109
	14–16	67.360	0.116	12–14	59.001	0.096
	16–18	51.726	0.103	15–17	57.704	0.151
	19–21	37.749	0.095	19–21	37.749	0.112
	22–24	29.038	0.125	21–23	36.004	0.109
	24–26	23.050	0.135	23–25	25.956	0.121
	28–30	18.892	0.135	25–27	19.576	0.111
	30–32	15.941	0.097	31–33	14.715	0.130
	33–35	13.225	0.133	33–35	13.225	**0.158**
	36–38	9.413	**0.140**	37–39	9.055	0.134
Bone-conducting	1–3	109.952	0.130	5–7	104.708	0.092
	7–9	**140.798**	**0.159**	7–9	140.798	0.105
	12–14	59.001	0.088	9–11	**172.493**	**0.143**
	15–17	57.704	0.102	13–15	67.903	0.106
	18–20	38.607	0.102	15–17	57.704	0.086
	20–22	39.490	0.100	19–21	37.749	0.098
	22–24	29.038	0.132	21–23	36.004	0.101
	25–27	19.576	0.104	27–29	20.429	0.114
	28–30	18.892	0.094	31–33	14.715	**0.143**
	33–35	13.225	0.135	34–36	12.558	0.101
	36–38	9.413	0.124	37–39	9.055	0.119

Native Listener. In the conventional in-ear experiment, a total of 13 WPLI peaks are observed, and the largest peaking value (0.140) occurs in the sub-frequency band of 36 Hz to 38 Hz, with a corresponding average normalised effective gain of 9.413 (Table 2). Conversely, the lowest peaking value of the WPLI (0.093) is observed in the sub-frequency band of 7 Hz to 9 Hz, with a corresponding average normalised effective gain of 140.798 (the highest value observed). In the case of bone-conducting devices, a total of 11 WPLI peaks are observed, and the largest peaking value (0.159) occurs in the sub-frequency band of 7 Hz to 9 Hz, with a corresponding average normalised effective gain of 140.798 (the highest value observed).

Compared with conventional in-ear phones, the bone-conducting devices exhibit a higher value for the largest peaking WPLI. The WPLI at the frequency bands where the largest normalised effective gain is obtained is also higher than the value observed in conventional in-ear phones. Additionally, there are two fewer peaking WPLIs observed in the dominant frequency range. The higher value of the peaking WPLI and the corresponding average normalised effective gain indicate that the evoked potential (characterised by WPLI) is at a higher level for bone-conducting devices compared to conventional in-ear phones.

Non-native Listener. In conventional in-ear phones, a total of 13 peaking values in the WPLI are observed (Table 2). The highest peaking WPLI (0.158) is obtained in the sub-frequency band of 33 Hz to 35 Hz, and the second-highest peaking WPLI (0.151) is observed in the sub-frequency band of 15 Hz to 17 Hz, with a corresponding average normalised effective gain of 57.704. In the case of bone-conducting devices, a total of 11 peaking WPLI values are observed. The maximum peaking WPLI (0.143) is obtained in the sub-frequency band of 7 Hz to 9 Hz, with a corresponding average normalised effective gain of 172.493 (also the largest value of average normalised effective gain).

Comparison Between Native and Non-native Listeners. For subjects using conventional in-ear phones, both the top two peaking values of the WPLI for native cases are higher than non-native cases. However, at other sub-frequency bands within the dominant frequency range, higher values of peaking WPLI are observed for native cases. These findings indicate that the number of sub-frequency bands where non-native subjects experience a higher level of evoked potential, characterised by the WPLI, compared to native subjects is reduced compared to the case of evoked potential characterised by the PLV. This discrepancy arises due to the different mathematical definitions between them.

According to the mathematical definition, a higher PLV implies a lower phase difference between two signals averaged over the trials. However, such a lower phase difference does not necessarily result in a higher value of the WPLI. The WPLI incorporates the absolute value and sign of the imaginary part of the cross-spectrum between two signals. Therefore, a larger WPLI indicates that the phase difference tends to be densely distributed either in the range from 0 to π or from π to 2π, or the phase difference across all trials tends to have the same sign. Consequently, the phase difference for non-native subjects tends to be distributed around 0 for all trials, whereas the phase difference for native subjects tends to have the same sign but not as close to 0.

These findings indicate that native subjects using in-ear phones experience a lower level of evoked potential with a higher latency value, while non-native subjects using in-ear phones experience a higher level of evoked potential with a lower latency value. A larger latency value of the evoked potential often reflects a more complex cognitive brain process following external stimuli [13]. Hence, the acquired data is consistent with the hypothesis that native subjects experience a more complex cognitive brain process than non-native subjects.

The maximum peak WPLI for non-native subjects using bone-conducting devices is lower than that of native subjects (Table 2). Only at the sub-frequency bands from approximately 8 to 15 Hz and from 28 to 33 Hz, the peaking WPLI is higher for non-native cases. This comparison is similar to the case of conventional in-ear phones. Thus, the choice of the playing device has minimal impact on the evoked potential level experienced by native and non-native subjects.

The native subjects experience a higher evoked potential characterised by the WPLI at more frequencies from 0 Hz to 40 Hz. However, as discussed earlier, non-native subjects experience a higher evoked potential characterised by the PLV than native subjects. This supports the hypothesis that native subjects undergo more complex cognitive brain processes, as reflected by the evoked potential with a larger latency. In contrast, non-native subjects experience more intense evoked stimuli, indicated by the presence of the evoked potential with a lower latency.

4 Conclusion

This study analyses EEG and sound signals from participants using in-ear phones or bone-conducting devices while listening to both native and non-native songs. The data was subjected to time-frequency analysis to extract phase information and calculate normalised effective gain, phase locking value (PLV) and weighted phase lag index (WPLI). The results indicate that the frequency range of 0 Hz to 40 Hz exhibits a dominant band for higher normalised effective gain. A comparison of PLV and WPLI values reveals that non-native subjects generally exhibit higher PLV but lower WPLI across most sub-frequencies. This implies that non-native subjects experience a higher level of evoked potential with a lower phase difference, whereas native subjects display a lower evoked potential level with a higher phase difference. The outcome is useful for detecting music genres [15] or speech recognition [8] considering native and non-native listeners. Our findings also support to the hypothesis that native subjects engage in more complex cognitive processes. Furthermore, the choice of playing device was found to have a limited impact on the evoked potential level. Future research endeavours could expand the sample size by including a larger number of participants and investigating induced potentials to identify the dominant frequency bands associated with high induced potential levels across different experimental conditions.

References

1. Alain, C., Roye, A., Arnott, S.R.: Chapter 9 - middle- and long-latency auditory evoked potentials: what are they telling us on central auditory disorders? In: Celesia, G.G. (ed.) Disorders of Peripheral and Central Auditory Processing, Handbook of Clinical Neurophysiology, vol. 10, pp. 177–199. Elsevier (2013). https://doi.org/10.1016/B978-0-7020-5310-8.00009-0
2. Alzahab, N.A., et al.: Auditory evoked potential EEG-biometric dataset (2021). https://doi.org/10.13026/ps31-fc50. version 1.0.0
3. Aydore, S., Pantazis, D., Leahy, R.M.: A note on the phase locking value and its properties. Neuroimage **74**, 231–244 (2013). https://doi.org/10.1016/j.neuroimage.2013.02.008
4. Cohen, M.X.: Analyzing Neural Time Series Data: Theory and Practice. MIT Press, Cambridge (2014)
5. Dehaene-Lambertz, G.: Electrophysiological correlates of categorical phoneme perception in adults. NeuroReport **8**(4), 919–924 (1997). https://doi.org/10.1097/00001756-199703030-00021

6. Galambos, R.: A comparison of certain gamma band (40-hz) brain rhythms in cat and man. In: Başar, E., Bullock, T.H. (eds.) Induced Rhythms in the Brain, pp. 201–216. Birkhäuser Boston, Boston (1992). https://doi.org/10.1007/978-1-4757-1281-0_11
7. Goldberger, A.L., et al.: PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation **101**(23), e215–e220 (2000). https://doi.org/10.1161/01.CIR.101.23.e215
8. Hasan, M.R., Hasan, M.M., Hossain, M.Z.: Effect of vocal tract dynamics on neural network-based speech recognition: a Bengali language-based study. Expert. Syst. **39**(9), e13045 (2022). https://doi.org/10.1111/exsy.13045
9. Ibrahim, I.A., Ting, H.N., Moghavvemi, M.: The effects of audio stimuli on auditory-evoked potential in normal hearing Malay adults. Int. J. Health Sci. **12**(5), 25 (2018)
10. Imperatori, L.S., et al.: EEG functional connectivity metrics wPLI and wSMI account for distinct types of brain functional interactions. Sci. Rep. **9**(1), 8894 (2019). https://doi.org/10.1038/s41598-019-45289-7
11. Jagiello, R., Pomper, U., Yoneya, M., Zhao, S., Chait, M.: Rapid brain responses to familiar vs. unfamiliar music-an EEG and pupillometry study. Sci. Rep. 9(1), 15570 (2019). https://doi.org/10.1038/s41598-019-51759-9
12. Lachaux, J., Rodriguez, E., Martinerie, J., Varela, F.J.: Measuring phase synchrony in brain signals. Hum. Brain Mapp. **8**(4), 194–208 (1999). https://doi.org/10.1002/(SICI)1097-0193(1999)8:4$⟨$194::AID-HBM4$⟩$3.0.CO;2-C
13. Michalopoulos, K., Iordanidou, V., Giannakakis, G.A., Nikita, K.S., Zervakis, M.: Characterization of evoked and induced activity in EEG and assessment of intertrial variability. In: 2011 10th International Workshop on Biomedical Engineering, pp. 1–4. IEEE (2011). https://doi.org/10.1109/IWBE.2011.6079037
14. Morales, S., Bowers, M.E.: Time-frequency analysis methods and their application in developmental EEG data. Dev. Cogn. Neurosci. **54**, 101067 (2022). https://doi.org/10.1016/j.dcn.2022.101067
15. Rahman, J.S., Gedeon, T., Caldwell, S., Jones, R., Hossain, M.Z., Zhu, X.: Melodious micro-frissons: detecting music genres from skin response. In: 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2019). https://doi.org/10.1109/IJCNN.2019.8852318
16. Tallon-Baudry, C., Bertrand, O.: Oscillatory gamma activity in humans and its role in object representation. Trends Cogn. Sci. **3**(4), 151–162 (1999). https://doi.org/10.1016/S1364-6613(99)01299-1
17. Vialatte, F.B., Dauwels, J., Musha, T., Cichocki, A.: Audio representations of multi-channel EEG: a new tool for diagnosis of brain disorders. Am. J. Neurodegener. Dis. **1**(3), 292–304 (2012)
18. Vinck, M., Oostenveld, R., Van Wingerden, M., Battaglia, F., Pennartz, C.M.: An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. Neuroimage **55**(4), 1548–1565 (2011). https://doi.org/10.1016/j.neuroimage.2011.01.055
19. Wagner, M., Ortiz-Mantilla, S., Rusiniak, M., Benasich, A.A., Shafer, V.L., Steinschneider, M.: Acoustic-level and language-specific processing of native and non-native phonological sequence onsets in the low gamma and theta-frequency bands. Sci. Rep. **12**(1), 314 (2022). https://doi.org/10.1038/s41598-021-03611-2