Neural correlates of masked and unmasked tones: psychoacoustics and late auditory evoked potentials (LAEPs)

Hyojin Kim[1], Viktorija Ratkute[2], Bastian Epp[1]*

[1]Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark

[2]Biomedical Engineering, Technical University of Denmark, Kgs. Lyngby, DK-2800, Denmark

*bepp@dtu.dk

ABASTRACT

Hearing thresholds can be used to quantify one's hearing ability. In various masking conditions, hearing thresholds can vary depending on the auditory cues. With comodulated masking noise and interaural phase disparity (IPD), target detection can be facilitated, lowering detection thresholds. This perceptual phenomenon is quantified as masking release: comodulation masking release (CMR) and binaural masking level difference (BMLD). As these measures only reflect the low limit of hearing, the relevance of masking release at supra-threshold levels is still unclear. Here, we used both psychoacoustic and electro-physiological measures to investigate the effect of masking release at supra-threshold levels. We investigated whether the difference in the amount of masking release will affect listening at supra-threshold levels. We used intensity just-noticeable difference (JND) to quantify an increase in salience of the tone. As a physiological correlate of JND, we investigated late auditory evoked potentials (LAEPs) with electroencephalography (EEG). The results showed that the intensity JNDs were equal at the same intensity of the tone regardless of masking release conditions. For LAEP measures, the slope of the P2 amplitudes with a function of the level was inversely correlated with the intensity JND. In addition, the P2 amplitudes were higher in dichotic conditions compared to diotic conditions. Estimated the salience of the target tone from both experiments suggested that the salience of masked tone at supra-threshold levels may only be beneficial with BMLD.

Introduction

Acoustic scenes in everyday life consist of a mixture of sounds. Our auditory system can segregate this mixture into a target sound and background, and thereby enable communication in acoustically complex environments. It is assumed that our auditory system binds various acoustic features arising from the same source into a sound object, or an acoustic stream. As an example of such features, speech shows coherent amplitude modulation patterns across a wide frequency range. Previous studies have shown that coherent modulation, or comodulation, is beneficial for sound detection in noise [1]. This suggests that comodulation is used for grouping spectral components across different frequency bands as these components likely stem from the same source. Such grouping can facilitate the segregation of the target signal from the noise, and results in enhancement in detection performance. Similarly, spatial information can also facilitate sound detection. When an acoustic source is lateralized relative to the listeners' head direction, a phase difference between the ears can be induced. This interaural phase difference (IPD) can facilitate the target identification by grouping acoustic features from the same source. For instance, when the target tone in the noise is presented with an IPD between left and right ears, detection thresholds of the tone is lower compared to the case with no IPD. In psychoacoustics, such enhancement in detection performance is considered as "release from masking", commonly referred to as masking release. Masking release can be quantified with an amount of decrease in the detection threshold. For masking release induced by comodulation, a decrease in detection threshold is termed comodulation masking release (CMR). For masking release induced by binaural cue is termed as binaural masking level difference (BMLD).

In simple cases where the target tone is presented with both comodulation and IPD cue, the amount of masking release was found to be a superposition of CMR and BMLD [2]. They interpreted the psychoacoustical measures, CMR and BMLD, as the result of enhanced neural representations by bottom-up neural processing. For CMR, the earliest neural encoding of CMR was found at CN [3], [4], and its neuronal representation get sharper at the inferior colliculus (IC) . For BMLD, neural correlates of IPD were found at IC [5], [6], [7]. Based on these findings, they suggested that the superposition of CMR and BMLD is the combination of neural encoding for CMR and BMLD [2]. However, it is not yet known whether this applies in any other cases where the amount of CMR is affected by the temporal context of the spectral masker components. When a comodulated masker is preceded and followed by another masker (temporal fringe), CMR can be reduced or increased depending on the type of preceding and following maskers [8], [9] [10]. When the temporal fringe is comodulated, CMR can be enhanced compared to without temporal fringe as the grouping of maskers is preceded and followed, facilitating separation of target sound from noise. On the contrary, when the temporal fringe is uncorrelated, CMR can be reduced compared to without temporal fringe as the maskers cannot be separated from the target. The results were interpreted with the view of stream formation by frequency grouping in time. Assuming the stream formation to be a higher-order process, these results suggest the influence of higher-order processing on CMR [8], [9], [10]. As a physiological correlates, the neurons at A1 showed enhanced neural responses to the stimuli that preceded by comodulated masker compared to the one by uncorrelated masker [11]. Nevertheless, it remains unclear whether the improved neural response is merely the result of relayed encoding from the CN and IC to A1, or if A1 plays an additional role in encoding, or if there is a cortical feedback from A1 to CN [11]. Furthermore, little is known about the effect of the stream formation on BMLD, which raises the first question to be investigated in this study: how the stream formation affects both CMR and BMLD? Our hypothesis was that if the stream formation is the result of higher-order processing (e.g., temporal integration), both CMR and BMLD will be affected by the stream formation. Behavioral outcome of the effect of the stream formation on the combination of CMR and BMLD would reveal the neural encoding strategies underlying sound source separation in the auditory system.

The second part of this study investigates the question of the benefit of masking release at supra-threshold levels. Masking release has been widely used to investigate beneficial acoustic features for the target detection. The amounts of CMR and BMLD are significantly high up to 10 [2], suggesting the sensitivity of the auditory system to these cues. CMR and BMLD are well characterized at low intensities (near thresholds). However, one might argue that target detection in communication occurs at levels well above threshold, or supra-threshold levels. This leads to a question of its relevance for communication in complex acoustical environments. Several studies have investigated the perception at supra-threshold levels in masking release conditions. Common goal was to map from the physical properties (e.g., the increment in the intensity of a sound) to the psychophysical variables (e.g., the increment in loudness or salience). Related studies used categorical loudness scaling and continuous scaling of the perceived salience. In categorical loudness scaling, they used loudness matching method between the target tone in modulated noise and the one in unmodulated noise [12]. The level of unmodulated noise was reduced by the amount of threshold difference

between two noise types. Their results showed that the supra-threshold perception of the target tone in the modulated noise was similar to that in the unmodulated noise at reduced level, suggesting that the masking release is the result of reduced the masker level internally. The study with continuous scaling tested on both CMR and BMLD. The data obtained by continuous scaling showed individual variability in the ratings, presumably because some listeners confused the loudness of the overall sound with the salience of the target tone in noise [13]. The limitation of those methods is that those measures strongly depend on the subjective criterion for decision making by listeners.

Here, rather than asking listeners to quantify the salience, we used the method of limits or a just-noticeable difference (JND). With this method, the intensity decreases with steps, until the listener cannot judge if one with decreasing level is more louder than the others with fixed level. This minimum difference in the intensity of two sounds that induces a just-noticeable difference is an intensity JND. This has potential in excluding subjective criteria for judging the salience of the target tone. Previous study showed that the loudness JND (ΔL) of broadband noise? is proportional to $L^{1/p}$, where L is the loudness and p is a piece-wise intensity-independent constant ranging from 1 to 2 [14]. Previous study with the continuous scaling of the salience suggested that the salience would be the same at the same supra-threshold levels. Their results were not in line with their hypothesis, possibly due to subjective ratings. As an extension of their study, our hypothesis was that the intensity JND in masking release will follow the power-law, and would show the same p with the function of supra-threshold levels, regardless of masking release conditions. If the internal neural representation is enhanced proportional to the amount of masking release, the condition with higher amount of CMR and BMLD would show lower intensity JNDs at the same level.

The last part of this study was to investigate a electro-physiological measure of the salience. EPP et al (2013) evaluated auditory evoked potentials (AEPs) with the hypothesis that the neural representation of the target tone in masker at the cortical level would be correlated with the perceptual outcome, the salience [15]. They assumed that the signal-to-noise ratio of the neural representation is enhanced by comodulation and IPD cues, and would be reflected on AEPs. They measured AEPs at various intensities of the tonal component in fixed level of the masker. They found that the amplitude of P2 component of the late auditory evoked potentials (LAEPs) correlated with the amount of masking release, CMR and BMLD. The growth function of the P2 amplitude across conditions showed similar slope. Based on this finding, in the follow-up study by EGGER, they hypothesized that LAEPs measured at the same level above masked threshold (e.g., threshold + 5 dB, + 10 dB etc) will evoke the same amplitude of the P2 components regardless of masking release conditions. They measured LAEPs at the six different supra-threshold levels, together with the salience of the tonal component in maskers with scaling method. The results of LAEPs showed that the tone at the same supra-threshold levels evoked similar P2 amplitudes, however, ratings of the salience were not correlated with P2. In this study, instead of comparing the amplitudes of P2, we estimated the slope of changes in P2 amplitudes with increased level. Our hypothesis was that the increment in P2 with increasing tone level would be inversely proportional to the intensity JND.

To summarize, we investigated: i) the effect of stream formation induced by preceding masker on CMR and BMLD; ii) the intensity JND as a measure of salience of the tonal component in masking release conditions at supra-threshold levels; iii) the correlation between the intensity JND measures and LAEP measures. With both the intensity JND measures and LAEP measures, we estimated the salience of the target tone in masking release conditions.

Materials and methods

Stimuli

Our study consists of three experiments. Two psychoacoustical experiments: i) threshold measurements to quantify CMR and BMLD; ii) the intenisty JND measurements. One EEG experiment to measure iii) LAEPs. For all three experiments, we used the same eight masking release conditions as shown in Fig 1. In the frequency domain (Fig 1a), the stimulus consisted of a masker constructed of five noise bands and a pure tone as a target signal: one noise band was centred at the frequency of the target tone. The other bands were equally placed apart with 120 Hz distance above and below the center band (flanking bands, FBs). Each masker band had a bandwidth of 20 Hz and a level of 60 dB SPL. The tone was centered at 700 Hz. We chose this frequency setting to maximize the effect of the stream formation on CMR based on the previous work by [8].

In the time domain (Fig 1b), each stimulus consisted of a 500ms interval of “preceding masker”, and a 200ms interval of the masked target tone, “target interval”. We used four masker conditions: the reference condition with maskers with random intensity fluctuations across frequency for both the “preceding masker” and the

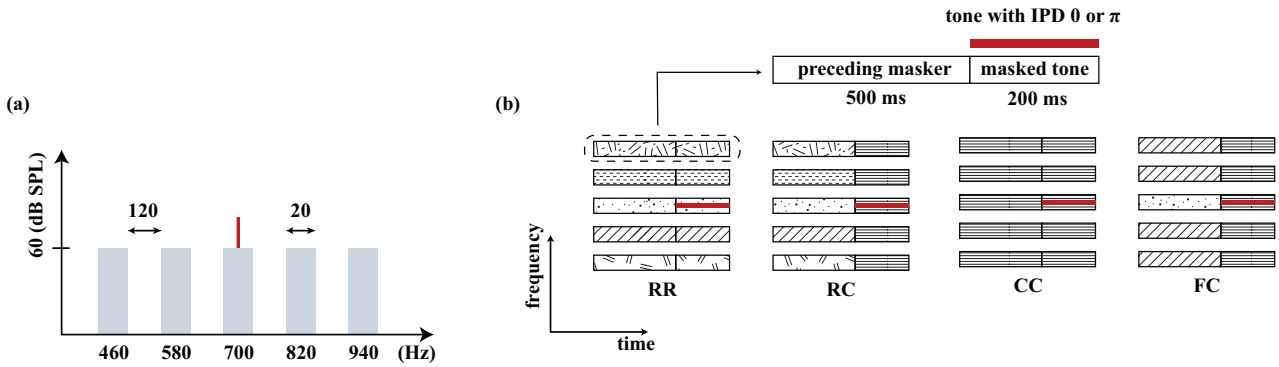

Figure 1: (a) Spectra of the stimulus. A target tone (700 Hz) was presented with a masking noise consisting of five narrow-band maskers: One signal centered band (SCB) and four flanking bands (FBs). The bandwidth of each masker band was 20 Hz and the frequency spacing was 120 Hz. The overall level of the noise was set to 60 dB SPL. (b) Schematic spectrograms of the stimulus conditions. Each stimulus consist of a preceding masker (500 ms) and masked tone (200 ms). Four types of maskers were used: RR, RC, CC, and FC. The RR was used as the reference condition with uncorrelated masker bands. In the other three conditions, the maskers consisted of a comodulated masker preceded by three different maskers: uncorrelated masker (RC), comodulated masker (CC), and the masker with comodulated flanking-bands (FC). The thick black line represents a tone which was presented with an IPD of 0 or π.

"target interval" (RR). In the three other conditions, the target tone was in comodulated noise, and this target interval was preceded by one of the following masker type: a masker with random intensity fluctuations across frequency (RC), a comodulated masker (CC), and a masker where only the FBs were comodulated (FC). To induce BMLD, with the same four masker types, the target tone was presented with either an IPD of 0 or π.

All maskers were presented diotically and had 20 ms raised-cosine on- and offset ramps. For RC and FC conditions where there is a transition between masker types, same on- and offset ramps were added. The noise bands were generated in the frequency domain, and transformed into the time domain. The noise bands were assigned numbers from a uniformly distributed random process to the real and imaginary parts of the respective frequency components with a frequency resolution of 2 Hz. For R masker, different numbers were assigned for each noise band, and for C masker, the same numbers were used for all five noise bands. The stimuli were generated for each interval and each trial.

Apparatus

During all three experiments, the listeners were seated in a double-walled, soundproof booth with ER-2 headphones (Etymotic). All stimuli were generated by MATLAB with a sampling rate of 44100 Hz and a 16-bit resolution, converted from digital to analogue (RME Frieface UCX). The headphones were calibrated and equalized at the signal frequency of the tone. For the recording of AEPs, we used a g.Tec HIamp system with a sampling rate of 1024 Hz. The 64 channels of active electrodes were set up with highly conductive electrode gel to reduce the impedance between the scalp and electrodes. The reference electrodes were placed close to the mastoid of both ears and other electrodes were placed based on g.GAMMAcap 64 channel setup from g.tec.

Listeners

From hearing screening, we recruited fifteen normal hearing listeners. None of them reported any history of hearing impairment and had pure-tone hearing thresholds within 15 dB HL for the standard audiometric frequencies from 125 to 4000 Hz. One listener was tested with 20 dB at 125 Hz. All participants provided informed consent and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391). All of them participated in the first experiment, eleven of them participated in the second experiment, and ten of them participated in the third experiment.

Procedure

In the first experiment, we measured the masked detection thresholds individually for eight conditions. We used an adaptive, three interval, three-alternative forced-choice procedure (3-AFC) with a one-up, two-down rule to estimate the 70.7% of the psychometric function [16], [17]. Two intervals contained the masking noise only, and the one interval contained the target tone in the same masking noise as other two. Three sound intervals were presented with a pause of 500 ms in between. The listeners' task was to select the interval with the target tone by pressing the corresponding number key (1, 2, 3) on the keyboard. Visual feedback was provided, indicating whether the answer was "WRONG" or "CORRECT". The initial level of the target tone was set to 75 dB and was adjusted with an initial step size of 8 dB. The step size was halved after each lower reversal until it reached the minimum step size of 1 dB. The signal level at a minimum step size of 1dB was measured six times, and the mean of the last six reversals were used as the estimated threshold. Each listener performed three threshold measurements for all conditions, and for each measurement, conditions were in randomized order. The average of three measurements were used as individual masked thresholds for the next two experiments. The additional measurement was done if the thresholds from the last three measurements had high variance (SD > 3 dB). This was to reduce the variability in the second and third experiment data where the level of the target tone was set based on masked threshold measurements.

In the second experiment, we measured the intensity JNDs individually at six supra-threshold levels for all conditions. The intensity of the tone was individually adjusted for each listener to match levels of +0 dB (threshold), +5dB, +10dB, +15 dB, +20 dB, and +25 dB above threshold. The individual mean of three threshold measurements from the first experiment was used to set supra-threshold levels. We used the same setup and 3-AFC method as for the first experiment. Two intervals contained the masked target tone with a fixed level at one of the supra-threshold levels (reference interval), and the interval contained the masked target tone with higher level than the others (target interval). Three sound intervals were presented with a pause of 500 ms in between. Listeners were asked to select the interval with the loudest tone by pressing the corresponding number key (1, 2, 3) on the keyboard. Visual feedback was provided, indicating whether the answer was "WRONG" or "CORRECT". The order of conditions and supra-threshold levels were randomized. The start level of the tone in the target interval was set to 75 dB, and The level of the target decreased until listeners could not discriminate from reference intervals. The tone level was adjusted with the initial step size of 8 dB. The step size was halved after each lower reversal until it reached the minimum step size of 1 dB. The signal level at a minimum step size of 1dB was measured six times, and the mean of the last six reversals were used as the minimum level of disciminable tone. Listeners were familiarized with the task by a test run. Each listener performed three trials for all conditions. If the supra-threshold level exceeded 80 dB, the intensity JND measure was skipped. We calculated the intensity JND by subtracting the level of reference intervals from the the minimum level of discriminable tone in the target interval.

In the third experiment, we measured the late auditory evoked potentials (LAEPs) at three supra-threshold levels for all conditions. The intensity of the tone was individually adjusted for each listener to match levels of +15 dB, +20 dB, and +25 dB above threshold. The individual mean of three threshold measurements from the first experiment was used to set supra-threshold levels. The stimuli for each condition and supra-threshold level were presented 400 times in random order, and we added noise only stimuli (10%). The presentations were separated by a random inter-stimulus interval of 500ms with jitter. This was added to eliminate AEPs by previous stimulus. During the experiment, a silent movie with subtitles was presented on a low-radiation screen. The listeners were asked to sit comfortably and avoid movement as much as possible. The experiment was divided in six blocks of approximately 38 minutes each. These were divided into two sessions on different days.

Data analysis

The threshold measurements We used CMR and BMLD to quantify the amount of masking release in eight conditions. We used several acronyms for masking release measures for each condition as follows. For comodulation masking release (CMR),

$$\mathrm{CMR_{m/ipd}} = \mathrm{threshold[RR_{ipd}]} - \mathrm{threshold[m_{ipd}]}, \tag{1}$$

Here, m stands for one of three masker types (RC, CC, FC) and ipd stands for the IPD of the tone between two ears (0 or π). As an example, CMR_{CC_π} is the amount of a decrease in thresholds between RR_π condition and CC_π. A positive value indicates a decreased detection threshold, and a negative value indicates a increased

detection threshold. For binaural masking level difference (BMLD),

$$\mathrm{BMLD_m} = \mathrm{threshold[m_0]} - \mathrm{threshold[m_\pi]}, \tag{2}$$

Here, m stands for one of three masker types (RC, CC, FC). As an example, $BMLD_{CC}$ is the difference in the thresholds between CC without IPD and with IPD of π.

The intensity JNDs We calculated the intensity JND by subtracting the level of reference intervals from the the minimum level of discriminable tone in the target interval. We first fitted JND measures for each condition with power law. In addition, we estimated the Weber fraction k by dividing the intensity JND (ΔL) with the intensity of the target tone (L). To use coherent measures used in other studies, we also calculated $10log(\Delta L/L)$. From the fitted intensity JNDs, we estimated the salience from 1 to 5. For each condition, we assumed that the salience was one at the masked thresholds. We increased the salience by one when the level was increased by the intensity JND at the current level. We repeated this estimation until the salience reached five.

The late auditory evoked potentials (LAEPs) Collected data was analyzed by The data analysis was performed using FieldTrip. The EEG data were partitioned into epochs from -300 to 850 ms relative to the onset of the masker. The region of interest was the central position (Cz), and the reference signals were the average of two electrodes near the mastoids. Each epoch was low pass filtered with a cut-off frequency of 20 Hz. Detrending, base line correction and weighted averaging were applied to increase the signal-to-noise ratio. Trials containing signals exceeding 100 μV were rejected as artifacts. For auditory evoked potentials (AEPs), from averaged sweeps, we extracted the signals from 100 ms before the onset of the target tone, and 100 ms after the offset of the target tone. Baseline correction was applied considering a 300 ms pre-stimulus period. The grand mean of AEPs was computed with an arithmetic mean over all individual AEPs (Fig 7). For late auditory evoked potentials (LAEPs), we chose the electrode placed at the vertex (Cz) where the largest response is elicited [13]. We defined the peak of first negative component between 100 ms and 200 ms (with respect to the target onset) as N1, and the peak of second positive component between 200 ms and 300 ms as P2. This was estimated for each individual AEPs rather than as from grand mean of AEPs as each individual AEP shows the difference in latency. Peak amplitudes were extracted by MATLAB function *findpeaks* by locating minima and maxima within the time frame defined for N1 and P2, respectively. Extracted LAEPs were visually verified, and when multiple components were found, the ones with the largest amplitude were selected. When there was no component was found, we excluded it from analysis.

Results

Experiment 1. The masked thresholds

Figure 2a shows the threshold measurements in eight masking release conditions. The observed mean threshold across all the participants for the RR_0 condition was 55.4 dB. The $RC0$ condition had a mean threshold of 52.23 dB. In the CC_0 condition, the threshold was 45.70 dB. In the FC_0 condition, the mean threshold was found to be 58.68 dB. With the IPD of π, the RR_π had a mean threshold of 39.51 dB, and the RC_π condition had a mean threshold of 38.27 dB. In the CC_π, the mean threshold was 33.13 dB, and that of the FC_π condition was 45.68 dB.

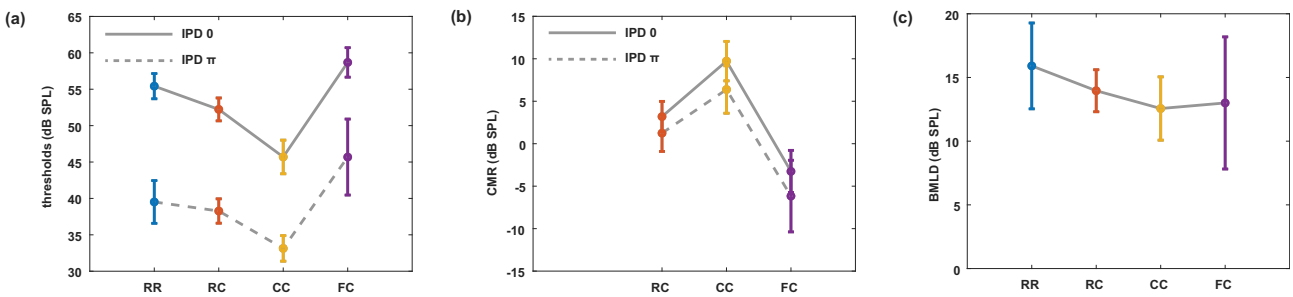

Figure 2: The plots of threshold measurements and CMR, BMLD (a) The plot of the threshold measurements from eight masking release conditions. Error bars indicate plus-minus one standard deviation. (b) The plot of CMR. (c) The plot of BMLD.

Figure 2b shows the CMR calculated for each condition by subtracting the threshold of RR condition from its

threshold (eq. (1)). In the diotic conditions, CMR_{RC0} was 3.19 dB, CMR_{CC0} was 9.73 dB and CMR_{FC0} was -3.26 dB. In dichotic conditions, $CMR_{RC\pi}$ was 1.24 dB, $CMR_{CC\pi}$ was 6.38 dB and $CMR_{FC\pi}$ was -6.17 dB. To test the significant difference in CMR measures between conditions, one-way ANOVA followed by Tukey's multiple comparison tests were used. Test showed that CMR measures were different between different masker types. Between diotic and dichotic conditions with the same masker type, only CMR_{CC0} and $CMR_{CC\pi}$ showed significant difference. This could be due to the flooring effect, where the maximum amount of masking release is reached and masked threshold cannot further decreases.

Figure 2c shows the BMLD calculated for each condition by subtracting the threshold of dichotic condition (eq. (2)). $BMLD_{RR}$ was 15.91 dB, $BMLD_{RC}$ was 13.86 dB, $BMLD_{CC}$ was 12.56 dB, and $BMLD_{FC}$ was 13 dB. One-way ANOVA showed $p < 0.0477$, and to further check the significant difference between conditions, we used Tukey's multiple comparison tests. Multiple comparison tests showed that the $BMLD_{CC}$ and $BMLD_{RR}$ were significantly different.

Experiment 2. The intensity JNDs

In Figure 3, the intensity JND measure for each condition was plotted separately, starting from the threshold (+ 0 dB) to + 25 dB. For each condition, the intensity JND measures were fitted with power function. The goodness of fitness for each condition with IPD of 0 was: RR(R^2=0.4748), RC(R^2=0.6614), CC(R^2=0.4732), FC(R^2=0.5604). The goodness of fitness for each condition with IPD of π was: RR(R^2=0.2743), RC(R^2=0.1613), CC(R^2=0.2946), FC(R^2=0.6170). The intensity JND measures were re-scaled with $10log(\Delta L/L)$, and fitted with power function as shown in Figure 4. The goodness of fitness for each condition with IPD of 0 was: RR(R^2=0.6386), RC(R^2=0.7914), CC(R^2=0.7472), FC(R^2=0.6491). The goodness of fitness for each condition with IPD of π was: RR(R^2=0.6338), RC(R^2=0.6316), CC(R^2=0.4993), FC(R^2=0.7559). Overall, the condition with lower detection thresholds (e.g., CC IPD π) showed higher JNDs compared to the one with higher detection thresholds (e.g., RR IPD 0). This indicates that the degree of enhancement in the salience depends on the level of the target tone. Figure 5 shows the averaged intensity JND (left) and re-scaled JND (right) at six supra-threshold levels: + 0 dB, + 5 dB, + 10 dB, + 15 dB, + 20 dB and + 25 dB. The intensity JND measures of all conditions and listeners are plotted with scatter, and fitted with the power function and plotted with the line. The intensity JNDs decreased with increasing level of the target tone in all masking release conditions.

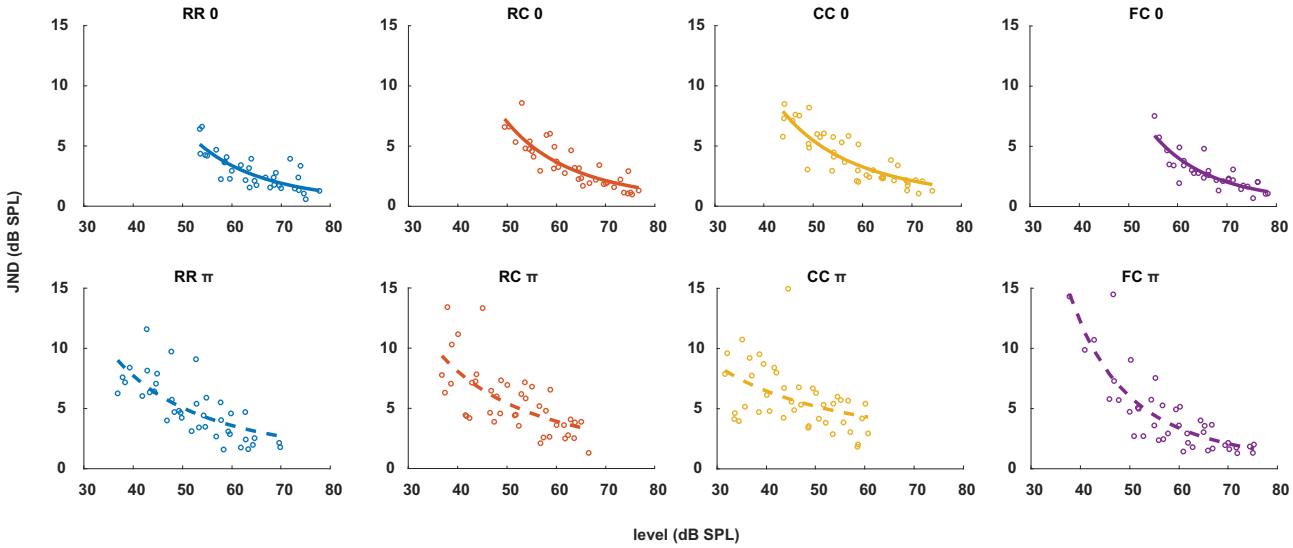

Figure 3: The plots of the intensity JNDs. The plot of the intensity JNDs as a function of the level. Individual data is plotted with scatter. The data for each condition is fitted with the power function and plotted with the line. Blue corresponds to RR condition, orange to RC, yellow to CC, and purple to FC condition. Solid line represents the data of IPD 0 and dotted line represents the data of IPD π.

Estimated salience The salience refers to the perceptual quantify that describes how clear the tone is perceived in noise. Assuming that the salience is the same at the threshold level, the salience would increase as the level of the target tone is increased. We hypothesized that the salience rating would increase when the target tone level is increased by the intensity JND. The estimated salience is shown in Fig 6. In dichotic conditions, the salience was higher compared to the diotic conditions with same masker type.

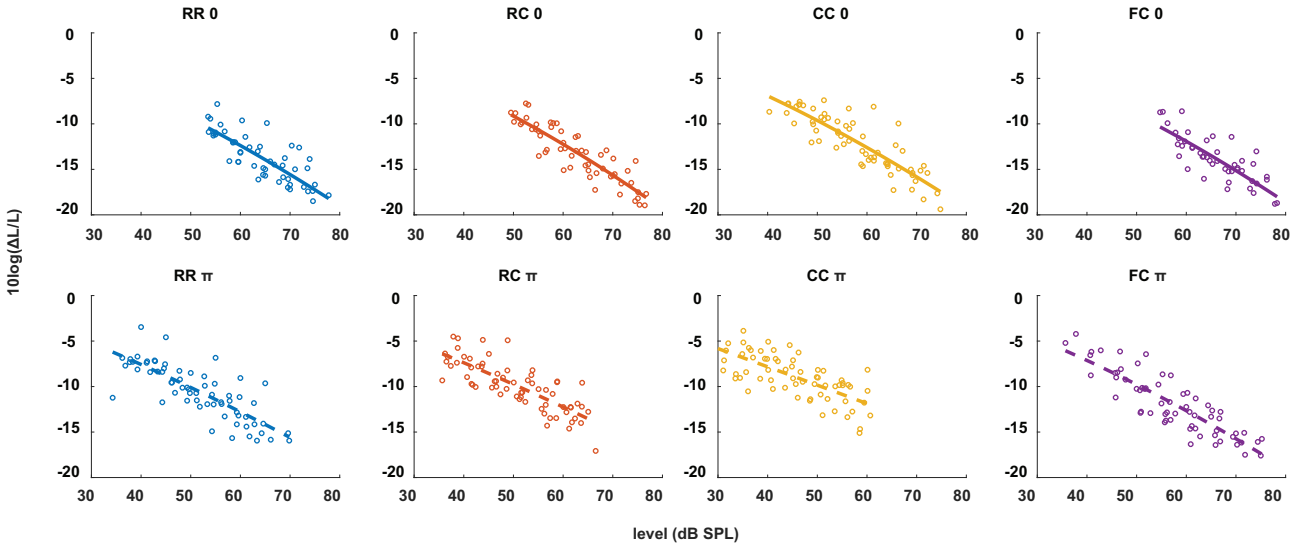

Figure 4: The plots of the intensity JNDs. The plot of the intensity JNDs as a function of the level. Individual data is plotted with scatter. The data for each condition is fitted with the power function and plotted with the line. Blue corresponds to RR condition, orange to RC, yellow to CC, and purple to FC condition. Solid line represents the data of IPD 0 and dotted line represents the data of IPD π.

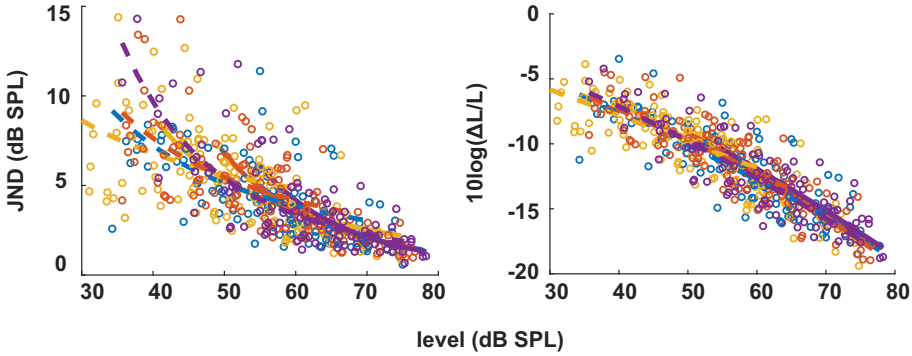

Figure 5: The intensity JNDs. (a) The plot of the averaged intensity JND at six supra-threshold levels: + 0 dB, + 5 dB, + 10 dB, + 15 dB, + 20 dB and + 25 dB. (b) The plots of the intensity JNDs over level of the target tone, $\Delta L/L$ at six supra-threshold levels: + 0 dB, + 5 dB, + 10 dB, + 15 dB, + 20 dB and + 25 dB. Individual data is plotted with scatter. The data for each condition is fitted with the power function and plotted with the line. Blue corresponds to RR condition, orange to RC, yellow to CC, and purple to FC condition. Solid line represents the data of IPD 0 and dotted line represents the data of IPD π.

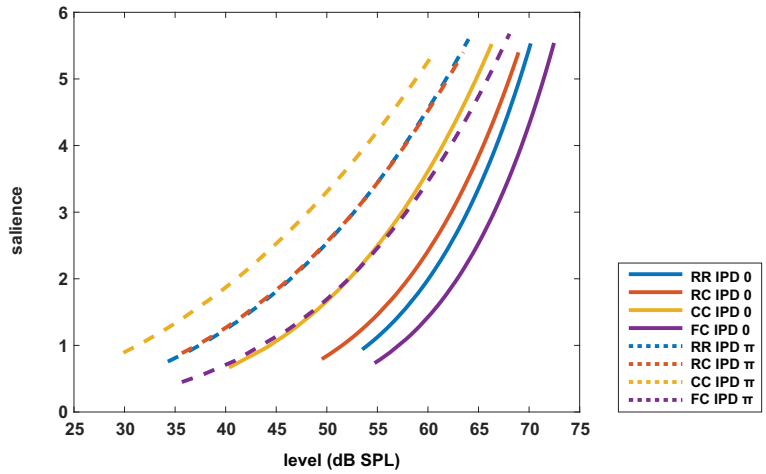

Figure 6: Estimated salience ratings. Each color represents four masker types. Solid lines represent diotic conditions and dotted lines represent dichotic conditions.

Experiment 3. Late auditory evoked potentials

Figure 7 shows the grand mean AEPs across all listeners for each condition. The plot shows the AEPs of diotic signals (solid lines) and dichotic signals (dashed lines) in four maker types RR, RC, CC, and FC, respectively. After the onset of the stimuli, onset response was elicited and go back to the baseline after 500 ms. Figure 8 shows the mean of LAEPs across all listeners for each condition. We selected first negative component (N1), and the second positive component (P2). We fitted P2 amplitudes with power function of the target tone level as shown in Figure 9 and 10. Each panel shows the LAEPs of each masker type with both diotic (solid lines) and dichotic (dashed lines) target tones. For each condition, the goodness of fitness with IPD of 0 was: RR(R^2=0.1820), RCR^2=(0.3253), CC(R^2=0.0970), FC(R^2=0.1646). The goodness of fitness with IPD of π was: RR(R^2=0.1847), RC(R^2=0.3224), CC(R^2=0.0601), FC(R^2=0.3161). For N1 amplitudes, the goodness of fitness (R^2) with IPD of 0 was: RR(R^2=0.2776), RC(R^2=0.1657), CC(R^2=0.2179), FC(R^2=0.2893). The goodness of fitness with IPD of π was: RR(R^2=0.2075), RC(R^2=0.1390), CC(R^2=0.0469), FC(R^2=0.0288).

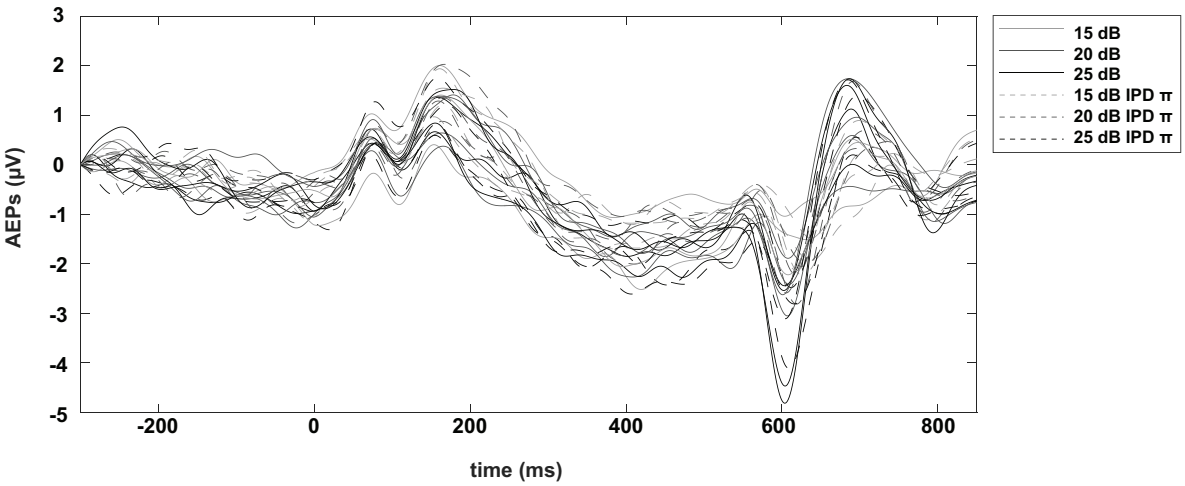

Figure 7: The averaged AEPs. The plots of the averaged auditory evoked potentials. Solid lines represent four masker types with diotic target signals, and dotted lines represent four masker types with dichotic target signals.

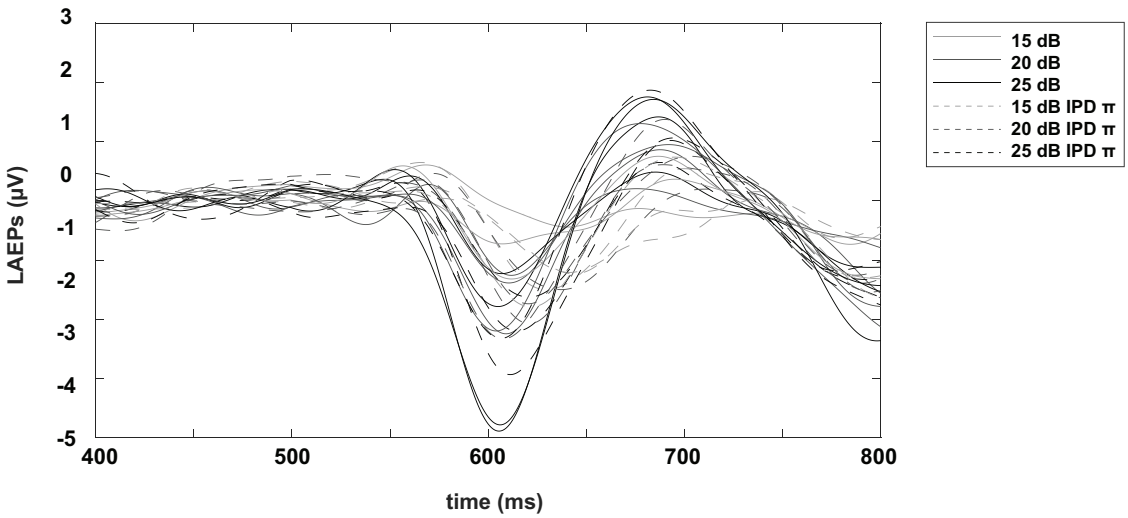

Figure 8: The averaged LAEPs. The plots of the averaged late auditory evoked potentials (LAEPs). Solid lines represent four masker types with diotic target signals, and dotted lines represent four masker types with dichotic target signals.

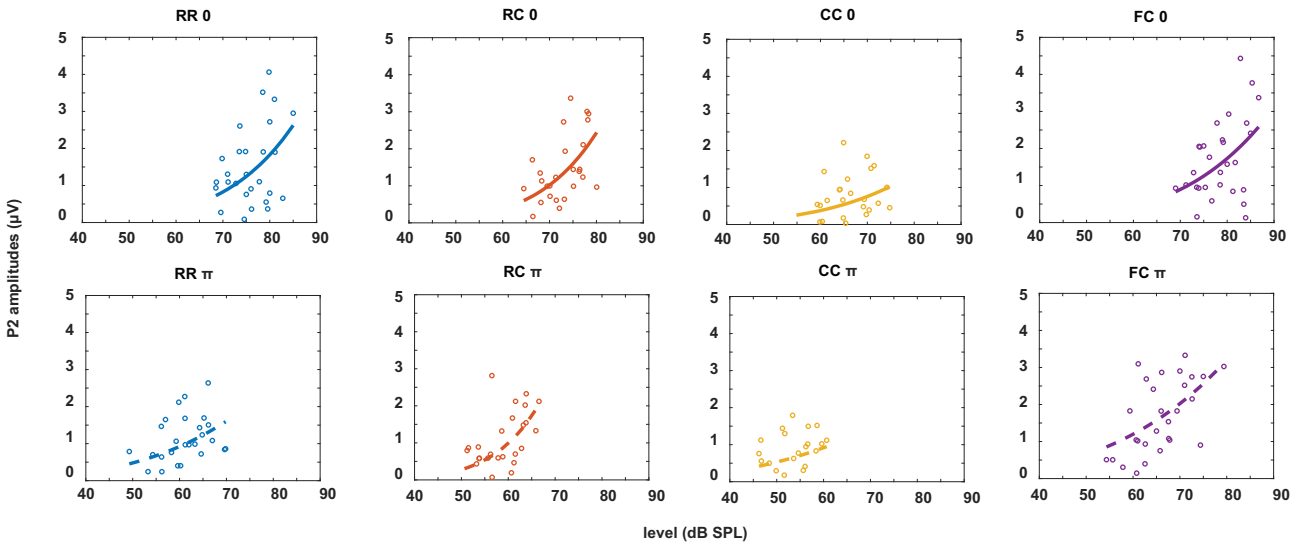

Figure 9: The plots of the P2 amplitudes with a function of target tone level at supra-threshold levels: + 15 dB, + 20 dB and + 25 dB. Individual data is plotted with scatter. The data for each condition is fitted with the power function and plotted with the line. Blue corresponds to RR condition, orange to RC, yellow to CC, and purple to FC condition. Solid line represents the data of IPD 0 and dotted line represents the data of IPD π.

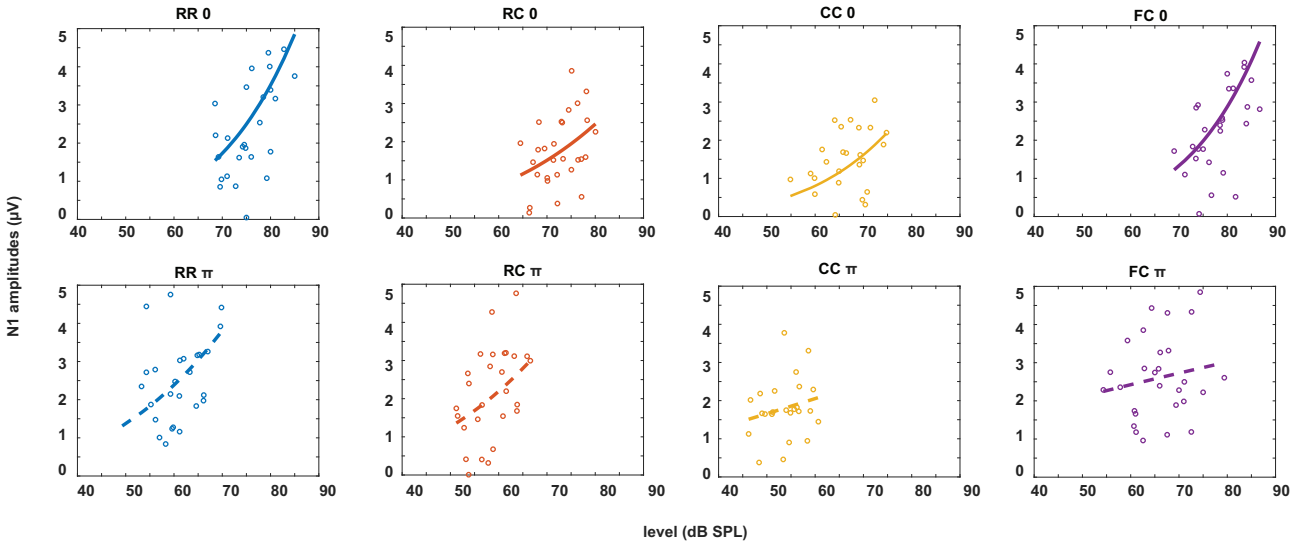

Figure 10: The plots of the N1 amplitudes with a function of target tone level at supra-threshold levels: + 15 dB, + 20 dB and + 25 dB. Individual data is plotted with scatter. The data for each condition is fitted with the power function and plotted with the line. Blue corresponds to RR condition, orange to RC, yellow to CC, and purple to FC condition. Solid line represents the data of IPD 0 and dotted line represents the data of IPD π.

Discussion

The effect of preceding stream formation on CMR and BMLD Results showed that the combination of comodulation, IPD, and the preceding masker can induce different amount of masking release. We designed our stimuli based on a previous study with both preceding and following maskers [8]. Although our stimuli had no following masker after the offset of the target tone, the thresholds measures were in line with their results. This may suggest that the preceding masker plays a role in inducing the streaming effect more than following maskers. Previous studies with preceding maskers interpreted the effect of the stream formation is by top-down processing [9], [10]. The auditory system uses accumulated information on incoming sound, or the adaptation at system level. We hypothesized that if this is the case, and BMLD processing occurs at IC, such top-down processing would affect to BMLD as it does to CMR. Interestingly, results showed that the adaptation to preceding maskers did not affect BMLD. This may suggest that the possible mechanism of the

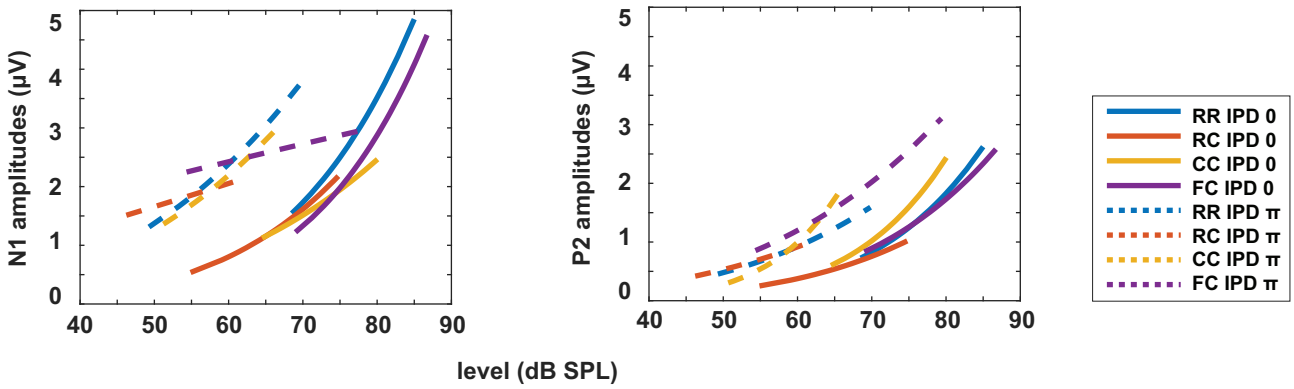

Figure 11: The plots of the LAEPs with a function of target tone level. The data of N1 (left) and P2 (right) is fitted with the power function and plotted with the line. Blue corresponds to RR condition, orange to RC, yellow to CC, and purple to FC condition. Solid line represents the data of IPD 0 and dotted line represents the data of IPD π.

adaptation may not the result of top-down processing (e.g. temporal integration) at the system level. Further psychoacoustic and physiological evidence is needed to support this idea.

The benefit of CMR and BMLD at supra-threshold levels In the second experiment, results showed that the intensity JNDs depend on the physical intensity of the target tone rather than the level above masked threshold (supra-threshold levels). Regardless of masking release conditions, the discrimination performance depends on the physical intensity of the target tone. The encoding of sound intensity (e.g., loudness) is often assumed to be in spike rates. At the level of periphery, the auditory nerve fibers (ANFs) can cover only the range of 30 - 40 dB. As such, if the intensity coding is with rate-based code, there should be some mechanisms along the auditory pathway. Several studies have suggested that the auditory cortex plays a role in intensity discrimination [18], [19]. However, the process before the cortex level is unclear. Here, we suggest that the such encoding may occur at cochlea nucleus (CN). From physiological studies, neural correlates of CMR were enhanced neural activities at CN. Intuitively, one might think that when level is increased, neural activity would increase more in CMR conditions compared to random condition. Rather, our results indicate that the physical intensity of the tone is encoded in addition to the enhanced neural representation at thresholds as an internal signal-to-noise ratio (iSNR). At the level of the cochlea nucleus (CN), neural response to the tonal component in comodulated noise was enhanced compared to in random noise [3], [4]. Based on this finding, we assumed that the iSNR is identical at threshold in all conditions. If the intensity-encoding is encoded by the spike rate, the iSNR will increase as the spike rate increases. In the same manner, we also assumed that the intensity JND can be encoded with spike rate differences. Behaviorally, the intensity JNDs decrease as the level increase. This means that the changes in spike rates should get steeper as level increases. At CN, small cells showed such responses to tone in noise [20], and this may play a role in the intensity coding in masking release conditions.

The LAEPs and the intensity JNDs

In previous studies, P2 component was suggested to be correlated with the supra-threshold levels. They showed that P2 amplitudes were correlated with the amount of masking release, CMR and BMLD [15]. In other words, when the target tone was presented with the same level, P2 amplitudes were high proportional to the amount of CMR and BMLD. In later study by Eager et al. [13], they measured P2 amplitudes at the same supra-threshold levels, or sensation level, P2 amplitudes were same regardless of the amount of CMR and BMLD. They also measured the salience of the target tone with continuous scaling method. However, they could not find correlation between the P2 amplitudes and the salience ratings. In this study, we estimated how P2 amplitudes grow as a function of the intensity JNDs, instead of comparing the absolute values of P2 amplitudes at supra-threshold levels. With fitted function for both P2 and the re-scaled intensity JND ($\Delta L/L$), we estimated those values at supra-threshold levels of each individual. We used masked threshold measures from fifteen listeners. As shown in Figure 12, the slope of P2 amplitudes get steeper as the intensity JNDs decreases, and for the same intensity JNDs, P2 amplitudes of dichotic conditions were higher. This was coherent with the estimated salience in Fig 6

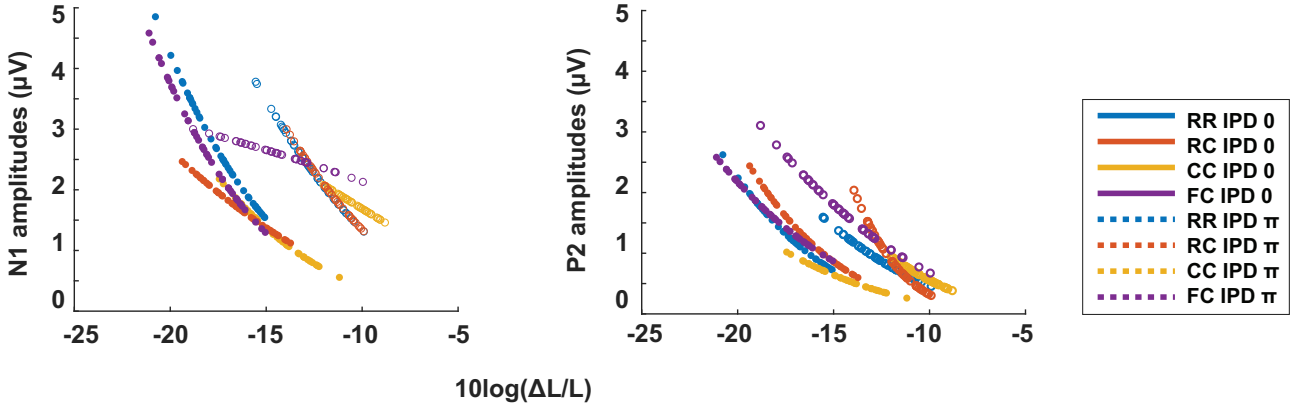

Figure 12: The plots of the LAEPs with a function of re-scaled intensity JNDs ($\Delta L/L$). The data of N1 (left) and P2 (right) is fitted with the power function and plotted with the line. Blue corresponds to RR condition, orange to RC, yellow to CC, and purple to FC condition. Solid line represents the data of IPD 0 and dotted line represents the data of IPD π.

Conclusion

In this study, we investigated the detection and discrimination of masked tones in masking release conditions. Auditory cues such as comodualtion and IPD, and the preceding masker, could enhance the detection performance. On the other hand, at supra-threshold levels, the discrimination performance was highly dependent on the level of the tone. Regardless of the masking release conditions, the intensity JND measures were correlated with the intensity of the tone. Furthermore, estimated salience of target tone suggested that binaural cues may be more beneficial than comodulation cue at supra-threshold levels. Lastly, the behavioural measure of the salience was correlated with the P2 amplitudes.

References

[1] I. Nelken, Y. Rotman, and O. B. Yosef, “Responses of auditory-cortex neurons to structural features of natural sounds,” *Nature*, vol. 397, no. 6715, pp. 154–157, 1999.

[2] B. Epp and J. L. Verhey, “Superposition of masking releases,” *Journal of computational neuroscience*, vol. 26, no. 3, pp. 393–407, 2009.

[3] D. Pressnitzer, R. Meddis, R. Delahaye, and I. M. Winter, “Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus,” *Journal of Neuroscience*, vol. 21, no. 16, pp. 6377–6386, 2001.

[4] V. Neuert, J. L. Verhey, and I. M. Winter, “Responses of dorsal cochlear nucleus neurons to signals in the presence of modulated maskers,” *Journal of Neuroscience*, vol. 24, no. 25, pp. 5789–5797, 2004.

[5] T. M. Shackleton, B. C. Skottun, R. H. Arnott, and A. R. Palmer, “Interaural time difference discrimination thresholds for single neurons in the inferior colliculus of guinea pigs,” *Journal of Neuroscience*, vol. 23, no. 2, pp. 716–724, 2003.

[6] T. M. Shackleton, R. H. Arnott, and A. R. Palmer, “Sensitivity to interaural correlation of single neurons in the inferior colliculus of guinea pigs,” *JARO - Journal of the Association for Research in Otolaryngology*, vol. 6, no. 3, pp. 244–259, 2005.

[7] O. Zohar, T. M. Shackleton, I. Nelken, A. R. Palmer, and M. Shamir, “First spike latency code for interaural phase difference discrimination in the guinea pig inferior colliculus,” *Journal of Neuroscience*, vol. 31, no. 25, pp. 9192–9204, 2011.

[8] J. H. Grose, E. Buss, and J. W. Hall, “Within- and across-channel factors in the multiband comodulation masking release paradigm,” *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 282–293, 2009.

[9] T. Dau, S. D. Ewert, and A. J. Oxenham, “Effects of concurrent and sequential streaming in comodulation masking release,” in *Auditory signal processing*, pp. 334–342, Springer, 2005.

[10] T. Dau, S. Ewert, and A. J. Oxenham, “Auditory stream formation affects comodulation masking release retroactively,” *The Journal of the Acoustical Society of America*, vol. 125, no. 4, pp. 2182–2188, 2009.

[11] J. Sollini and P. Chadderton, “Comodulation enhances signal detection via priming of auditory cortical circuits,” *Journal of Neuroscience*, vol. 36, no. 49, pp. 12299–12311, 2016.

[12] J. L. Verhey and W. Heeren, “Categorical scaling of partial loudness in a condition of masking release,” *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 904–915, 2015.

[13] K. Egger, T. Dau, and B. Epp, “Supra-threshold perception and neural representation of tones presented in noise in conditions of masking release,” *PLoS ONE*, vol. 14, no. 10, pp. 1–18, 2019.

[14] J. B. Allen and S. T. Neely, “Modeling the relation between the intensity just-noticeable difference and loudness for pure tones and wideband noise,” *The Journal of the Acoustical Society of America*, vol. 102, no. 6, pp. 3628–3646, 1997.

[15] B. Epp, I. Yasin, and J. L. Verhey, “Objective measures of binaural masking level differences and comodulation masking release based on late auditory evoked potentials,” *Hearing research*, vol. 306, pp. 21–28, 2013.

[16] S. D. Ewert, “Afc—a modular framework for running psychoacoustic experiments and computational perception models,” in *Proceedings of the international conference on acoustics AIA-DAGA*, pp. 1326–1329, 2013.

[17] H. Levitt, “Transformed up-down methods in psychoacoustics,” *The Journal of the Acoustical society of America*, vol. 49, no. 2B, pp. 467–477, 1971.

[18] A. R. Dykstra, C. K. Koh, L. D. Braida, and M. J. Tramo, “Dissociation of detection and discrimination of pure tones following bilateral lesions of auditory cortex,” 2012.

[19] C. Micheyl, H. Kreft, S. Shamma, and A. J. Oxenham, “Temporal coherence versus harmonicity in auditory stream formation,” *The Journal of the Acoustical Society of America*, vol. 133, no. 3, pp. EL188–EL194, 2013.

[20] A. Hockley, C. Wu, and S. E. Shore, “Cochlear nucleus small cells use olivocochlear collaterals to encode sounds in noise,” *bioRxiv*, 2021.