The effect of the preceding masking noise on monaural and binaural release from masking

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

*hykim@dtu.dk

ABSTRACT

When a target tone is preceded by a noise, the threshold for target detection can be increased or decreased depending on the type of a preceding masker. The effect of preceding masker to the following sound can be interpreted as either the result of adaptation at the periphery or at the system level. To disentangle these, we investigated the time constant of adaptation by varying the length of the preceding masker. For inducing various masking conditions, we designed stimuli that can induce masking release. Comodulated masking noise and binaural cues can facilitate detecting a target sound from noise. These cues induce a decrease in detection thresholds, quantified as comodulation masking release (CMR) and binaural masking level difference (BMLD), respectively. We hypothesized that if the adaptation results from the top-down processing, both CMR and BMLD will be affected with increased length of the preceding masker. We measured CMR and BMLD when the length of preceding maskers varied from 0 (no preceding masker) to 500 ms. Results showed that CMR was more affected with longer preceding masker from 100 ms to 500 ms while the preceding masker did not affect BMLD. In this study, we suggest that the adaptation to preceding masking sound may arise from low level (e.g. cochlear nucleus, CN) rather than the temporal integration by the higher-level processing.

Introduction

Cocktail party problem questions how one can listen to a particular talker while other people speak simultaneously. This problem was defined as "auditory scene analysis", suggesting that our auditory systems perceive the acoustic environment as a combination of multiple streams [1]. This enables us to segregate the target from the noise. This is analogous to the visual system that uses visual cues such as colors, textures, shapes to separate visual information into different objects. Similarly, a stream can be thought of as a basic unit of sounds resulting from a perceptual grouping of auditory cues.

Perceptually, when frequency components are closed in frequency or in time, the more they tend to be grouped into one stream. In more context-dependent cases, one of the key principle of all for the stream formation is that frequency components should come from the same sound source. Firstly, as our vocal tract vibrates, it modulates frequency components arising from its vibration. Such co-modulation can be a "grouping cue" for our auditory system. Experimentally, the benefit of co-modulation has been shown with a detection task. When the task is to detect a pure tone in noise, the detection performance increases when the noise is comodulated compared to the random noise. This enhancement in the detection performance can be quantified as a decrease in masked thresholds, comodulation masking release (CMR) [2]. This can be interpreted that co-modulation groups frequency components of noise into one stream, facilitating the segregation of tone from noise. Secondly, similar to comodulation, spatial information can induce spatial masking release. When sounds are from certain location, it induces binaural difference in phase, time and level, and this can be beneficial for target detection. For instance, when the tone is presented to left and right ears with an interaural phase difference (IPD), masked thresholds decrease and it can be quantified as binaural masking level difference (BMLD) [3], [4]. In the study by Epp et al. [5], they showed the superposition of masking release where the sum of CMR and BMLD is equal to the masking release with both comodulation and IPD present. To show the superposition of CMR and BMLD, they implemented conceptual model with an assumption of serially-aligned processing of CMR and BMLD. However, the exact order of CMR processing and IPD processing is unknown.

In case of CMR, studies in both psychoacoustics and physiology have suggested that CMR can be the result of combined effect of primitive segregation (bottom-up processing) and schema-driven segregation (top-down processing). While primitive processing means that the auditory object arises by innate and short process of incoming acoustic input, schema-driven processing assumes that the auditory system uses stored knowledge of sound input in long time scale to form the auditory object [1]. In physiological studies, neural correlates of CMR have been found at various stage along the auditory pathway. The earliest neuronal encoding exists in the cochlear nucleus (CN) as an increased neuronal response to tone in comodulated noise compared to ones to uncomodulated noise [6], [7]. This enhanced neuronal responses progressively sharpens along the auditory pathway; the inferior colliculus, medial geniculate body, and auditory cortex [8], [9]. This may suggest that CMR is induced by bottom-up processing, however, several studies have found the top-down processing affects CMR. When the tone in comodulated masker is preceded or/and followed by another masker (temporal fringe), CMR decreases or increases depending on the type of temporal fringe maskers [10]. This implies that CMR is affected by the stream formed before and after the target is presented [11], [12]. They found that both pre-cursors and post-cursors could eliminate CMR. In their later study in 2009, they suggested that for the effect of precursors, it is hard to exclude the effect of peripheral neural adaptation. Furthermore, the neural correlates of the influence of preceding maskers on CMR (e.g., reduced/increased CMR depending on temporal contexts) have been found in A1 [13]. This study showed that a preceding stream is formed as the auditory system adapts to the sound, and such system level adaptation could/may exist at the level of A1 (top-down). Nevertheless, it is still unknown whether there is cortical feedback on the sub-cortical processing of CMR or A1 plays a role blah blah. In case of BMLD induced by IPD, physiological evidence suggests that IPD information is processed at the inferior colliculus (IC), which is located after the CN [14]. Contrary to CMR, little has been found regarding the stream formation effect on BMLD.

From both psychoacoustics and physiology sides, the neural basis of the effect of preceding masker on CMR and BMLD is in question. In this study, we varied the duration of preceding maskers to disentangle the adaptation effect on CMR at periphery and at cortical level (bottom-up and top-down). In addition, we measured BMLD with the same condition as CMR. We investigated i) how much extent the adaptation affects on CMR with time constant, and ii) whether an effect of duration is at system level. Our hypothesis was that higher-order processing at system level would affect both CMR and BMLD. In the first experiment, we measured masked thresholds with varying the length of the preceding maskers to test if preceding masker' effect on CMR is the result of higher-processing or not. In the second experiment, we measured masked thresholds with binaural cue (IPD of π) in the same conditions as the first experiment. We calculated both

CMR and BMLD in dichotic conditions to investigate the effect of preceding masker on CMR and BMLD.

Methods

Stimuli We measured CMR and BMLD with six different lengths of preceding maskers varied from 0 ms (no preceding masker) to 500 ms. For conditions with no preceding maskers (0 ms), two maskers R, C were used. With the preceding masker (20 ms, 100 ms, 200 ms, 300 ms, 500 ms), we used four different maskers: the reference condition with uncorrelated masker (RR), and three maskers consist of a comodulated masker preceded by three different maskers: uncorrelated masker (RC), comodulated masker (CC), and the masker with comodulated flanking-band (FC). For BMLD, we used interaural phase difference (IPD) of pi. Each masker consist of five narrow noise bands with 20 Hz bandwidth and the level of the masker was set to 60 dB SPL. Tone was centered at 700 Hz with signal-centered band (SCB). Flanking bands were centered at 460, 580, 820, and 940 Hz. We chose spectral distance between narrow bands to maximize CMR based on Grose et al. [10].

Protocol After the training session, each listener performed three threshold measurements for all conditions. For each measurement, conditions were in randomized order. The thresholds were estimated by averaging three trials. The additional measurement was done if the thresholds from the last three measurements had high variance (SD > 3 dB). During the threshold measurement, the listeners were seated in a double-walled, soundproof booth with ER-2 headphones. 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 [15], [16]. Three sound intervals were presented with a pause of 500 ms in between. Two were with only maskers, and the one was the target tone with maskers. The listeners' task was to choose the interval with a target tone by pressing the corresponding number key (1, 2, 3) on the keyboard. Whenever the listener pressed the keyboard, visual feedback was provided, indicating whether the answer was "WRONG" or "CORRECT". The target tone's start level was set to 75 dB and depending on the answer, and 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 those measurements was estimated as the threshold.

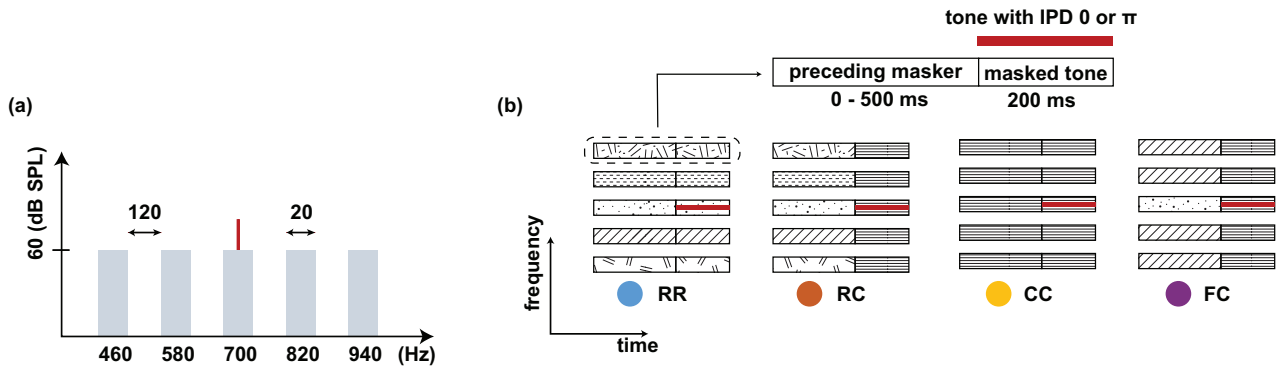

Figure 1: (a) Spectra of the stimulus. A target tone (700 Hz) was presented with a masking noise consisting of five narrow-band maskers: One signal centered band (SCB) and four flanking bands (FBs). The bandwidth of each masker band was 20 Hz and the frequency spacing was 120 Hz. The overall level of the noise was set to 60 dB SPL. (b) Schematic spectrograms of the stimulus conditions. Each stimulus consist of a preceding masker (0 - 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.

Listeners From hearing screening, we recruited eleven 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).

Results

The threshold measurements In the first experiment, we measured masked thresholds in diotic conditions. Figure 2 shows the mean thresholds across all listeners. The results are plotted with solid lines. Each plot shows the the mean threshold values from 20 ms to 500 ms of preceding maskers for four masker types: RR, RC, CC, and FC. As a reference, we had two masker types without preceding maskers: R and C. This was plotted with black color to compare to conditions with preceding maskers. In the second experiment, we measured masked thresholds in dichotic conditions with IPD of π. The results are plotted with dotted lines in Figure 2. As same as diotic conditions, each plot shows the the mean threshold values from 20 ms to 500 ms of preceding maskers for four masker types: RR, RC, CC, and FC.

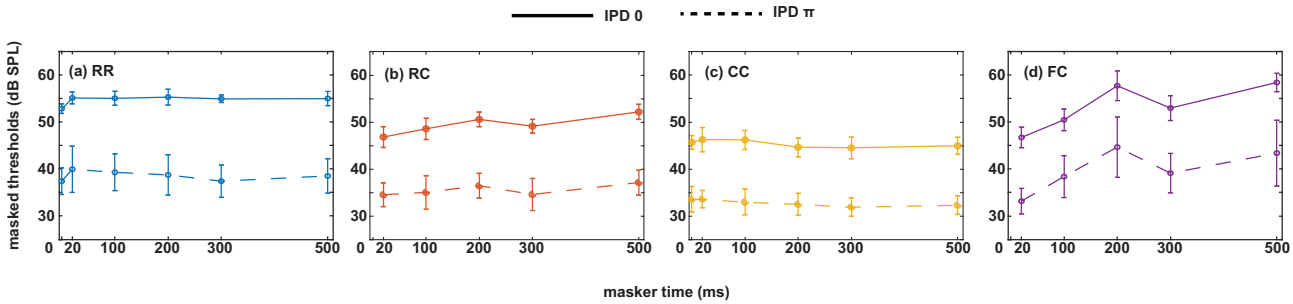

Figure 2: Mean masked thresholds. The threshold measurements from all listeners were averaged. Thresholds are plotted for each masker type with colors (RR - blue, RC - orange, CC - yellow, FC - purple). For diotic conditions with IPD of 0, masked thresholds are plotted with solid lines. For dichotic conditions with IPD of π, masked thresholds are plotted with dotted lines.

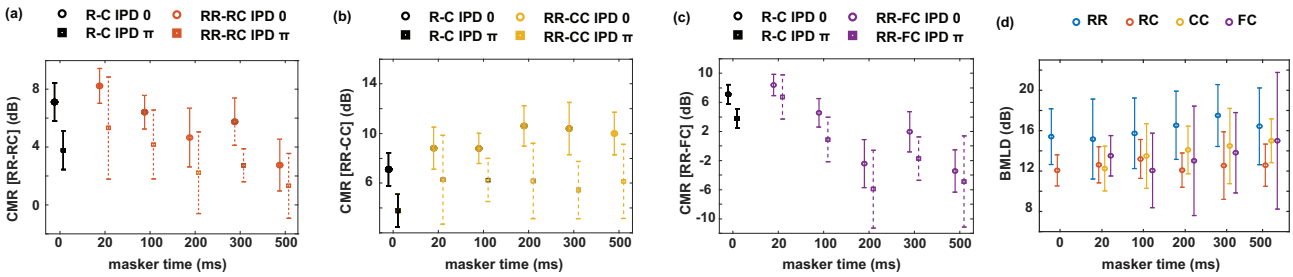

Figure 3: The plots for CMR and BMLD. Each plot shows changes in CMR and BMLD with varying the duration of preceding maskers. As a reference, we calculated CMR for the condition with no preceding masker by subtracting the masked threshold of C from the one of R. Reference condition is plotted with black color. The data with solid lines are of diotic conditions, and the data with dotted lines are of dichotic conditions. (a) The plot of CMR changes in RC condition. (b) The plot of CMR changes in CC condition. (c) The plot of CMR changes in FC condition. (d) The plot of BMLD changes in all conditions.

CMR We calculated CMR for diotic conditions for three conditions (RC, CC, and FC) by subtracting their thresholds from the thresholds of RR condition. Resulting CMR is shown in Figure 3. Each plot shows changes in CMR with varying the duration of preceding maskers. As a reference, we calculated CMR for the condition with no preceding masker by subtracting the masked threshold of C from the one of R. Reference condition is plotted with black color. From Figure 3(a) to (c), each plot shows the CMR changes in RC, CC conditions, and FC condition, respectively.

We tested the significant changes in CMR when the duration of the preceding masker was increased from 0 ms to 500 ms (Table 1). In the RC condition with IPD of 0, CMR was significantly decreased from 8 to 3 dB (one-way ANOVA: F(4,45) = 16.17, $p < 0.001$). In the CC condition with IPD of 0, CMR was significantly increased from 7 to 10 dB (one-way ANOVA: F(5,54) = 6.44, $p < 0.001$). In the FC condition with IPD of 0, CMR was significantly decreased from 9 to -4 dB (one-way ANOVA: F(4,45) = 36.47, $p < 0.001$).

For dichotic conditions with IPD of π, the CC condition showed no significant changes in CMR. In the RC condition, CMR was decreased 5 dB to 2 dB (one-way ANOVA: F(4,45) = 3.9, p = 0.008). In the FC condition, CMR significantly decreased from 7 dB to -5 dB (F(4,45) = 13.57, $p < 0.001$).

BMLD We calculated BMLD for four conditions (RR, RC, CC, and FC) by subtracting masked threshold of dichotic conditions from the that of diotic conditions. In Figure , each plot shows the BMLD for each condition (RR, RC, CC, and FC). Unlike CMR, BMLD did not show significance change (one-way ANOVA) with varying the duration of preceding maskers (RR condition (F(5,54)=0.64, p=0.67), CC condition (F(5,54)=0.37, p=0.86), RC condition (F(4,45)=1.43, p=0.24, and FC condition (F(4,45)=0.54, p=0.70)). For the FC condition, we observed that the variance of BMLD measures increased from 2 dB to 6.7 dB while the variances of other conditions were around 2 to 3.8 dB.

Discussion

In this study, we investigated i) the time constant of the effect of preceding masker on CMR and BMLD, and ii) whether the effect of preceding masker is the result of high-order processing at system level (e.g., temporal integration). We measured masked thresholds with varying the length of the preceding maskers. In diotic conditions, all types of preceding maskers affected CMR. In dichotic conditions, however, the CC condition did not show any changes in CMR with increased duration of the preceding masker. This may be due to the low limit of the auditory system in which a maximum amount of masking release is reached. In addition, in dichotic conditions, CMR measures were lower compared to the diotic conditions. This is in line with previous studies [17], [18], [19]. Possible neural basis has been proposed as a contralateral projection of the wide-band inhibitor [20], [19], [21].

Furthermore, we measured masked thresholds with binaural cue (IPD of π) to investigate the effect of preceding masker on CMR and BMLD. Results showed that CMR was more affected with longer preceding masker while the preceding masker did not affect BMLD. This indicates that the grouping of frequencies in the preceding masker only influences following frequency grouping by comodulation but not by IPD. This may suggest that the possible mechanism of the adaptation to preceding maskers could exist at CN level, rather than the system level. In visual system, the adaptation to the light contrast occurs at retina level (equivalent to cochlea) ranging from 0.1 to 17 seconds [22]. These can be done by different retinal cells such as bipolar cells, amacrine cells, and ganglion cells. Slow adaptation was suggested to be encoded as a gradual hyperpolarization. Similarly, in CN, various types of neurons exist and these neurons are connected to each other [23]. Various functional circuit types exist such as feed-forward excitation/inhibition, feed-back inhibition, and mutual excitation. For instance, L-cells that project feed-back inhibition as delayed input to T-stellate cells that encode CMR [24]. This may contribute the slow adaptation in auditory system as delayed wide-inhibitor, and further physiological evidence is needed to provide insights on possible mechanisms of the adaptation to preceding masker.

Conclusion

The present study investigated the effect of the preceding masker on CMR and BMLD by varying the length of the preceding masker. The effect of the preceding masker on the following sound has been suggested to be either the result of the adaptation at peripheral level or system level. Our hypothesis was that if CMR and BMLD change with increased duration of the preceding masker, the effect of preceding masker could be due to higher-level processing. From the data acquired in this study, we showed that the time constant of the effect of preceding masker can range up to 500 ms on CMR. However, the preceding masker did not affect BMLD. This may indicate that the effect of preceding masker may be the result of neural processing at CN rather than that of high-order processing at system level (e.g., temporal integration).

Table 1: The significant differences in CMR between conditions $p - values$ with varying the length of the preceding masker.

Condition	IPD	Duration (ms)	0	20	100	200	300	500
RR-CC	IPD 0	**0**	-	0.205	0.217	$<$ **0.001**	$<$ **0.001**	**0.003**
		20	-0.205	-	0.999	0.159	0.274	0.594
		100	0.217	0.999	-	0.149	0.260	0.574
		200	$<$ **0.001**	0.159	0.149	-	0.999	0.960
		300	$<$ **0.001**	0.274	0.260	0.999	-	0.993
		500	**0.003**	0.594	0.574	0.960	0.993	-
	IPD π	**0**	-	0.285	0.292	0.329	0.713	0.344
		20	0.285	-	0.999	0.999	0.979	0.999
		100	0.292	0.999	-	0.999	0.981	0.999
		200	0.329	0.999	0.999	-	0.988	0.999
		300	0.713	0.979	0.981	0.988	-	0.990
		500	0.344	0.999	0.999	0.999	0.990	-
RR-RC	IPD 0	**0**	-	-	-	-	-	-
		20	-	-	0.101	$<$ **0.001**	**0.010**	$<$ **0.001**
		100	-	0.101	-	0.119	0.887	$<$ **0.001**
		200	-	$<$ **0.001**	0.119	-	0.546	0.076
		300	-	**0.010**	0.887	0.546	-	**0.001**
		500	-	$<$ **0.001**	$<$ **0.001**	0.076	**0.001**	-
	IPD π	**0**	-	-	-	-	-	-
		20	-	-	0.855	0.068	0.177	**0.008**
		100	-	0.855	-	0.436	0.716	0.106
		200	-	0.068	0.436	-	0.991	0.929
		300	-	0.177	0.716	0.991	-	0.721
		500	-	**0.008**	0.106	0.929	0.721	-
RR-FC	IPD 0	**0**	-	-	-	-	-	-
		20	-	-	**0.014**	$<$ **0.001**	$<$ **0.001**	$<$ **0.001**
		100	-	**0.014**	-	$<$ **0.001**	0.176	$<$ **0.001**
		200	-	$<$ **0.001**	$<$ **0.001**	-	**0.003**	0.901
		300	-	$<$ **0.001**	0.176	**0.003**	-	$<$ **0.001**
		500	-	$<$ **0.001**	$<$ **0.001**	0.901	$<$ **0.001**	-
	IPD π	**0**	-	-	-	-	-	-
		20	-	-	**0.033**	$<$ **0.001**	$<$ **0.001**	$<$ **0.001**
		100	-	**0.033**	-	**0.009**	0.669	**0.038**
		200	-	$<$ **0.001**	**0.009**	-	0.214	0.982
		300	-	$<$ **0.001**	0.669	0.214	-	0.495
		500	-	$<$ **0.001**	**0.038**	0.982	0.495	-

References

[1] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

[2] J. W. Hall, M. P. Haggard, and M. A. Fernandes, "Detection in noise by spectro-temporal pattern analysis," *Journal of the Acoustical Society of America*, vol. 76, no. 1, pp. 50–56, 1984.

[3] L. A. Jeffress, H. C. Blodgett, T. T. Sandel, and C. L. Wood III, "Masking of tonal signals," *The Journal of the Acoustical Society of America*, vol. 28, no. 3, pp. 416–426, 1956.

[4] S. van de Par and A. Kohlrausch, "Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1940–1947, 1999.

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

[6] 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.

[7] 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.

[8] 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.

[9] L. Las, E. A. Stern, and I. Nelken, “Representation of tone in fluctuating maskers in the ascending auditory system,” *Journal of Neuroscience*, vol. 25, no. 6, pp. 1503–1513, 2005.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] G. P. Schooneveldt and B. C. Moore, “Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking bands,” *The Journal of the Acoustical Society of America*, vol. 85, no. 1, pp. 262–272, 1989.

[18] J. W. Hall III, J. H. Grose, and M. P. Haggard, “Effects of flanking band proximity, number, and modulation pattern on comodulation masking release,” *The Journal of the Acoustical Society of America*, vol. 87, no. 1, pp. 269–283, 1990.

[19] S. M. Ernst and J. L. Verhey, “Role of suppression and retro-cochlear processes in comodulation masking release,” *The Journal of the Acoustical Society of America*, vol. 120, no. 6, pp. 3843–3852, 2006.

[20] J. L. Verhey, D. Pressnitzer, and I. M. Winter, “The psychophysics and physiology of comodulation masking release,” *Experimental Brain Research*, vol. 153, no. 4, pp. 405–417, 2003.

[21] N. J. Ingham, S. Bleeck, and I. M. Winter, “Contralateral inhibitory and excitatory frequency response maps in the mammalian cochlear nucleus,” *European Journal of Neuroscience*, vol. 24, no. 9, pp. 2515–2529, 2006.

[22] S. A. Baccus and M. Meister, “Fast and slow contrast adaptation in retinal circuitry,” *Neuron*, vol. 36, no. 5, pp. 909–919, 2002.

[23] P. B. Manis and L. Campagnola, “A biophysical modelling platform of the cochlear nucleus and other auditory circuits: from channels to networks,” *Hearing research*, vol. 360, pp. 76–91, 2018.

[24] T. Ngodup, G. E. Romero, and L. O. Trussell, “Identification of an inhibitory neuron subtype, the l-stellate cell of the cochlear nucleus,” *ELife*, vol. 9, p. e54350, 2020.