DTU Library

Auditory stream formation affects comodulation masking release retroactively

Dau, Torsten; Ewert, Stephan; Oxenham, A. J.

Link back to DTU Orbit

Auditory stream formation affects comodulation masking release retroactively[a)]

Torsten Dau[b)]
Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark

Stephan Ewert
Medizinische Physik, Institut für Physik, Universität Oldenburg, D-26111 Oldenburg, Germany

Andrew J. Oxenham
Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

Many sounds in the environment have temporal envelope fluctuations that are correlated in different frequency regions. Comodulation masking release (CMR) illustrates how such coherent fluctuations can improve signal detection. This study assesses how perceptual grouping mechanisms affect CMR. Detection thresholds for a 1-kHz sinusoidal signal were measured in the presence of a narrowband (20-Hz-wide) on-frequency masker with or without four comodulated or independent flanking bands that were spaced apart by either 1/6 (narrow spacing) or 1 octave (wide spacing). As expected, CMR was observed for the narrow and wide comodulated flankers. However, in the wide (but not narrow) condition, this CMR was eliminated by adding a series of gated flanking bands after the signal. Control experiments showed that this effect was not due to long-term adaptation or general distraction. The results are interpreted in terms of the sequence of "postcursor" flanking bands forming a perceptual stream with the original flanking bands, resulting in perceptual segregation of the flanking bands from the masker. The results are consistent with the idea that modulation analysis occurs within, not across, auditory objects, and that across-frequency CMR only occurs if the on-frequency and flanking bands fall within the same auditory object or stream. © 2009 Acoustical Society of America. [DOI: 10.1121/1.3082121]

I. INTRODUCTION

The audibility of a target sound embedded in another masking sound can be improved by adding sound energy that is remote in frequency from both the masker and the target (Hall *et al.*, 1984). This effect is known as comodulation masking release (CMR) and is observed when the remote sound and the masker share coherent patterns of amplitude modulation. Most ecologically relevant sounds, such as speech and animal vocalizations, have coherent amplitude modulation patterns across different frequency regions, suggesting that the detection and recognition advantages conveyed by such coherent modulations may play an important role in our ability to deal with natural complex acoustic environments (e.g., Klump, 1996; Nelken *et al.*, 1999).

CMR has been measured in two ways. The first, often referred to as the "band-widening experiment," is to use a single band of noise, centered around the signal frequency, as a masker and to compare thresholds for modulated and unmodulated noise maskers as a function of the masker bandwidth (e.g., Hall *et al.*, 1984; Carlyon *et al.*, 1989). For the random noise, with irregular fluctuations in amplitude that are independent of different frequency regions, the signal threshold increases as the masker bandwidth increases up to about the critical bandwidth at that frequency and then remains constant, in broad agreement with the classical power spectrum model of masking (Fletcher, 1940; Patterson and Moore, 1986). For the modulated noise, a random noise that is amplitude modulated using a lowpass filtered noise as a modulator, the pattern of results is quite different. Here, signal thresholds typically decrease as the bandwidth increases beyond about 100 Hz (for a signal frequency of 2 kHz); thus, increasing the masker energy and bandwidth makes the signal easier to detect. These findings suggest that listeners may compare the outputs of different auditory filters to enhance signal detection. The fact that the decrease in threshold with increasing bandwidth only occurs with the modulated noise indicates that fluctuations in the masker are critical and that the fluctuations need to be correlated across frequency bands.

The second method is to use a masker consisting of several narrow masker bands of noise, typically with bandwidths between 20 and 50 Hz, which have relatively slow inherent amplitude fluctuations. One band is centered at the signal frequency (on-frequency band) and one or more other bands (flanking bands) are spectrally separated from the on-frequency band (e.g., Hall *et al.* 1984; Schooneveldt and Moore, 1987). When the flanking bands are uncorrelated with the on-frequency band, there is sometimes a slight el-

evation but typically no effect on signal threshold, so long as the flanking bands are not so close in frequency as to produce direct masking. However, when the amplitude fluctuations of the flanking bands are correlated with those of the on-frequency band, the addition of the flanking bands can produce a release from masking (Hall *et al.*, 1984; Schooneveldt and Moore, 1987; Cohen and Schubert, 1987). CMR has been found even if the signal and on-frequency band are presented to one ear and the flanking bands to the other ear (Schooneveldt and Moore, 1987; Cohen and Schubert, 1987; Buss and Hall, 2008).

Even though CMR has been investigated in many studies, the underlying mechanisms are still not clear. It was originally assumed that CMR results from across-channel comparisons of temporal envelopes (e.g., Buus, 1985). However, there is evidence that within-channel cues, i.e., information from only the one peripheral channel tuned to the signal frequency, can account for a considerable part of the effect in some conditions, suggesting that within-channel processing can lead to an overestimation of "true" across-channel CMR (Schooneveldt and Moore, 1987). This conclusion was supported by simulations of data from a band-widening experiment, using a modulation filterbank analysis of the stimuli at the output of the auditory filter tuned to the signal frequency (Verhey *et al.*, 1999; Piechowiak *et al.*, 2007). Additionally, for the CMR experiments using flanking bands, McFadden (1986) pointed out that it is imprecise to assume that one channel is receiving only the on-frequency band plus signal and another channel is receiving only the flanking band. Often, the two bands will be incompletely resolved. When this happens, the resulting waveform may contain envelope fluctuations resulting from beats between the carrier frequencies of the on-frequency and the flanker bands. These beats can facilitate signal detection without across-channel comparisons being involved (Schooneveldt and Moore, 1987). Thus, at least part of the CMR in many situations can be explained in terms of the use of within-channel rather than across-channel cues.

The authors of several studies have suggested that higher-level processes, such as object formation, may be involved in CMR, because certain stimulus manipulations designed to perceptually segregate the masker from the flanking bands have resulted in a reduction or elimination of CMR (McFadden and Wright, 1992; Grose and Hall, 1993). However, when manipulating perceptual grouping, it is often difficult to rule out mechanisms such as neural inhibition, forward suppression, or adaptation (Calford and Semple, 1995; Ulanovsky *et al.*, 2004; Wehr and Zador, 2005) that might at least partly be based on more peripheral processing. For example, neuronal adaptation, the decline over time of neural responses during sensory stimulation, might have affected the neural representation of the flanking masker bands in the experimental conditions of Grose and Hall (1993). In their study, CMR was reduced or eliminated either by gating the flanking bands on earlier and gating off later than the on-frequency masker band, or by presenting a series of precursor bands at the same frequencies as the flanking masker bands to perceptually segregate the on-frequency from the flanking masker bands. In both cases, it is at least conceiv-

able that the main effect of the precursors or leading onset asynchronies was to reduce the neural response to the flanking masker bands.

To exclude adaptation as a possible basis for the reduction or elimination of CMR, the present study focused on sounds that occurred *after* the target in time. Sounds occurring after a target and masker interval could in principle affect their perception by, for instance, binding with the flanking bands to form a separate perceptual stream (Dannenbring and Bregman, 1978; Bregman, 1990). Our hypothesis was that if across-frequency modulation analysis (and hence CMR) occurs primarily within auditory objects, then CMR could be eliminated by sounds that occur after the target, so long as these sounds are successfully segregating the on-frequency masker and the flanking masker bands into different auditory objects, thereby disrupting the across-frequency (but within-object) modulation processing. On the other hand, if across-frequency modulation processing is a lower-level or "automatic" process that is not governed by auditory grouping mechanisms, then sounds occurring after the target in time should not affect CMR.

II. RETROACTIVE STIMULUS EFFECTS ON CMR

A. Method

1. Listeners

Six normal-hearing listeners ranging in age from 25 to 39 years participated in the experiments. Two of the listeners were the first and second authors. All listeners received several hours of listening experience prior to the final data collection.

2. Apparatus and stimuli

Listeners were seated in a double-walled sound attenuating booth in front of a computer keyboard and monitor. The stimuli were presented diotically via Sennheiser HD580 headphones. Signal generation and presentation during the experiments were controlled by computer using the AFC software package for MATLAB, developed at the University of Oldenburg and the Technical University of Denmark. The stimuli were digitally generated at a sampling rate of 32 kHz and converted to analog signals by a high-quality 24-bit sound card (RME DIGI96/8 PAD).

Figure 1 shows schematic spectra of the stimuli: The target (a 1-kHz tone) was masked by a narrow (20-Hz wide) band of noise centered at 1 kHz. Four flanking bands of noise (each 20-Hz wide) were presented with temporal envelopes that were either random (condition R) and thus uncorrelated with that of the on-frequency masker, or coherent (condition C) with the masker envelope. The envelope fluctuations, or modulations, of the bands are indicated by the different shades of gray. The coherent across-frequency modulation of condition C was expected to enhance the audibility of the target, and hence reduce its detection threshold, relative to its threshold in condition R. The novel conditions investigated the retroactive influence of stimulus presentation on CMR. Several additional bursts of noise, termed postcursors, were presented at the frequencies of the original flanking bands (conditions PR and PC). The enve-
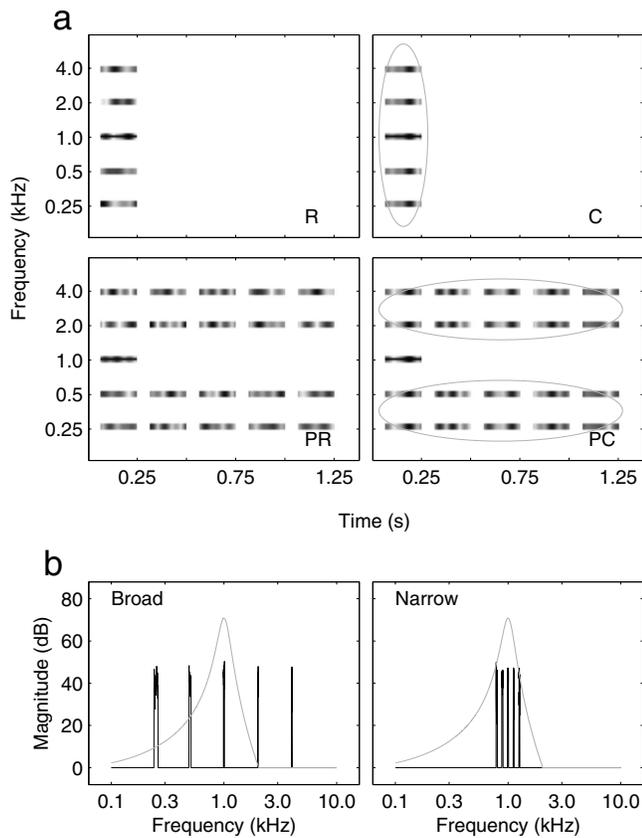
FIG. 1. Schematic representation of the experimental conditions. (a) Four conditions with a 1-kHz target tone (black horizontal line) masked by a noise band centered at 1-kHz with random flankers (R), comodulated flankers (C), random flankers followed by four postcursors (PR), and comodulated flankers followed by four comodulated postcursors (PC). The shades of gray indicate the distribution of envelope fluctuations in the masker and flanker bands. (b) Power spectra for the broadband configuration (with one-octave spacing between the noise bands) and narrowband configuration (with one-sixth octave spacing). The gray curves represent the magnitude response of the auditory filter centered at the target frequency.

lope coherence of the postcursors was the same as that for the original flanking masker bands. The postcursors were designed to "capture," and to form a single auditory stream with, the original flanking bands and thus to perceptually segregate them from the masker (Bregman and Pinker, 1978), as shown schematically in Fig. 1(a) (gray ellipses).

The experiment was performed using two spectral configurations [Fig. 1(b)]. In the broadband configuration, the noise bands were centered at 250, 500, 1000, 2000, and 4000 Hz, i.e., with one-octave spacing between the bands such that they primarily stimulated separate auditory filters along the tonotopic axis. The gray curve in Fig. 1(b) indicates the magnitude transfer function of a gammatone bandpass filter (e.g., Patterson *et al.*, 1995) tuned to the signal frequency (1 kHz). In the narrowband configuration, the noise bands were centered at 794, 891, 1000, 1123, and 1260 Hz, representing a sixth-octave spacing centered around the target frequency. In this case, within-channel processes were likely to play a strong role (Schooneveldt and Moore, 1987; Verhey *et al.*, 1999), because all the components fell within the same frequency region. To the extent that the effect of the postcursors is limited to across-frequency processing, they should not affect CMR for the narrowband configuration. Hence, the

narrowband configuration acted as a control condition for any potential non-specific distraction or interference effects produced by the postcursors.

In both cases, the noise bands were generated in the time domain as independent Gaussian noise tokens for each of the presentation intervals. The noise tokens were restricted to the appropriate bandwidth in the spectral domain via a Fourier transform. Comodulated noises were frequency-shifted versions of the masker band at 1000 Hz. The level of each of the noise bands was 60 dB sound pressure level (SPL). The four postcursors at the flanking-band frequencies all had the same duration and level as the masker bands (187.5 ms, including 20-ms raised-cosine ramps) and were separated by gaps of 62.5 ms, giving an overall repetition period of 250 ms.

3. Procedure

An adaptive, three-interval, three-alternative forced-choice procedure was used in conjunction with a 1-up, 2-down tracking rule to estimate the 70.7% correct point on the psychometric function (Levitt, 1971). The intervals were marked on a computer monitor and feedback was provided after each trial. Listeners responded via the computer keyboard or mouse. The initial step size of the target level was 8 dB, which was reduced to 4 and 2 dB after the second and fourth reversals, respectively. The adaptive run then continued for a further six reversals at the final step size, and threshold was defined as the mean of the levels at those last six reversals. Four threshold estimates were obtained and averaged from each listener in each condition. The intra-individual standard deviations were typically around 1–2 dB and rarely exceeded 4 dB. Final thresholds reported here are the mean across listeners, who all showed comparable patterns of results across conditions. Typical individual differences were around 2–3 dB and maximally reached 5–6 dB.

B. Results and discussion

The experimental data are shown in Fig. 2. In the broadband configuration (circles and filled bars), the results for conditions R and C were as expected from previous studies (Schooneveldt and Moore, 1987): The target threshold in the presence of the masker was significantly lower for the coherently modulated flanking bands than for the randomly modulated flanking bands [paired t-test; $t(5)=7.21$, $p<0.001$]. This difference in threshold, reflecting the amount of CMR, was 6.1 dB. However, this CMR was eliminated when the postcursors were added: Thresholds in condition PC were not significantly different from those in condition R [paired t-test; $t(5)=0.73$, $p=0.50$]. Similarly, there was no significant difference between thresholds in conditions PR and PC [paired t-test; $t(5)=0.59$, $p=0.58$], confirming the lack of effect of coherent amplitude modulations when the postcursors were present. The elimination of the CMR by sounds occurring after the target suggests that the postcursors led to perceptual segregation of the flanking bands from the masker, so that the coherent modulations in the flanking bands were no longer processed with those of the masker.

The results from the narrowband configuration (Fig. 2, squares and open bars) show that the effect of the postcursors
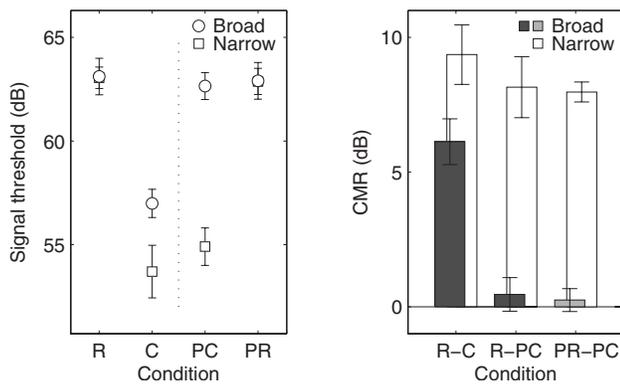
FIG. 2. Mean masked thresholds for the target tone (left) and amount of CMR (right) for the broadband (circles and filled bars) and the narrowband configurations (squares and open bars). Error bars denote one standard error across subjects. Conditions are indicated on the abscissa (R=random modulations of the flanking bands, C=comodulated flanking bands, PC =postcursors with comodulated flanking bands, and PR=postcursors with randomly modulated flanking bands). The amount of CMR, defined as the difference between thresholds in the random and the comodulated conditions, is indicated for the standard condition without postcursors, R-C, and for the conditions with postcursors (R-PC and PR–PC).

is unlikely to be a non-specific distraction effect: Here, the average target threshold in the presence of the masker (R) was 9.4 dB lower for the coherently modulated than for the randomly modulated flanking bands [paired t-test; $t(5)$ =8.46, $p<0.001$], as expected (Schooneveldt and Moore, 1987). However, in contrast to the broadband configuration, no significant reduction in CMR was produced when the postcursors were added [C-PC, $t(5)=-1.06$, $p=0.34$]. When compared to the random condition with postcursors, PR-PC (right open bar), the amount of CMR was 8.0 dB, in contrast to a non-significant 0.25 dB in the broadband configuration (right light-gray bar). Thus, postcursors eliminated CMR in the broadband configuration where CMR is likely to be based on across-frequency processing, but did not significantly affect target detection in the narrowband configuration, where CMR is more likely to be based on within-channel cues. In other words, the postcursors were successful at eliminating CMR only when CMR was likely to be based on a true across-frequency analysis of coherent modulations occurring in remote frequency bands.

Overall, the results are difficult to account for in terms of traditional neuronal adaptation or inhibition mechanisms because the critical sound components (the postcursors) occurred after the presentation of both the masker and target. However, it is possible in principle that the postcursors in a given trial affected the representation of the flankers in the following trial, via some form of long-term adaptation. In other words, despite the temporal gaps between successive trials, the postcursors may have influenced the response to the flanking bands in the next trial, which in turn may have reduced CMR. Long-term adaptation effects have been observed in the auditory pathways, particularly at higher levels, such as cortex (e.g., Ulanovsky et al., 2004; Altmann et al., 2007). Another possibility is that the postcursors induced some distraction effect that selectively impaired performance in the wideband, but not the narrowband conditions. For instance, if attention were exogenously diverted toward the

frequency regions of the flanking bands by the postcursors, and if this shift in attention affected signal detection, then this would be expected to selectively affect results for the wideband condition. Such an attentional effect would be less likely to affect results for the narrowband condition, because the frequency region of the flankers was close to that of the target. Both these possibilities were addressed in the following control conditions.

III. CONTROL CONDITIONS: MISSING AND OFF-FREQUENCY POSTCURSORS

A. Rationale

To address the possibility that the effect of the postcursors was due to longer-term effects on flankers in following trials, the first burst of postcursors following the flankers was replaced with a silent gap. The gap was expected to reduce or eliminate the perceptual grouping of the flankers with the postcursors, but would not be expected to eliminate any longer-term adaptation effects. To address the possibility that the effect of the postcursors was due to an attentional shift away from the target frequency, a second control condition was run in which the postcursors were shifted by a half-octave away from the frequencies of the flankers. In this case, attention would still be expected to shift from the target frequency, but the postcursors would no longer be expected to form a perceptual stream with the flanking masker bands. Thus, if the effect of the postcursors in experiment 1 was primarily due to an exogenous attentional shift, then CMR should be reduced even when the center frequencies of the postcursors are shifted; however, if the effect was due to perceptual grouping of the flanking maskers and postcursors into a single stream, then CMR should remain when the center frequencies are shifted, because the postcursors should no longer form a stream with the flanking masker bands.

B. Methods

The target and the flankers were the same as for the broadband configuration in the main experiment. The same listeners took part, and the apparatus and procedures for estimating thresholds were also the same. The left panel of Fig. 3(a) shows schematic spectrograms of the stimuli for the first control condition, where the first of the four postcursors was removed (gap-postcursor condition, GP); the right panel of Fig. 3(a) shows schematic spectrograms of the stimuli for the second control condition, where the center frequencies of the postcursors were shifted relative to the flanker center frequencies. The off-frequency postcursors were positioned with half-octave separation from the respective flanking bands, i.e., at 354, 707, 1414, and 2828 Hz.

C. Results and discussion

The data in Fig. 3(b) show that CMR is not affected by the postcursors, if the first postcursor is omitted. The threshold obtained in the comodulated condition with gap-postcursor (GPC) was not significantly different from the threshold obtained in the original condition C [from Fig. 2; $t(5)=-0.76$, $p=0.483$]. CMR was 5.5 dB when defined as
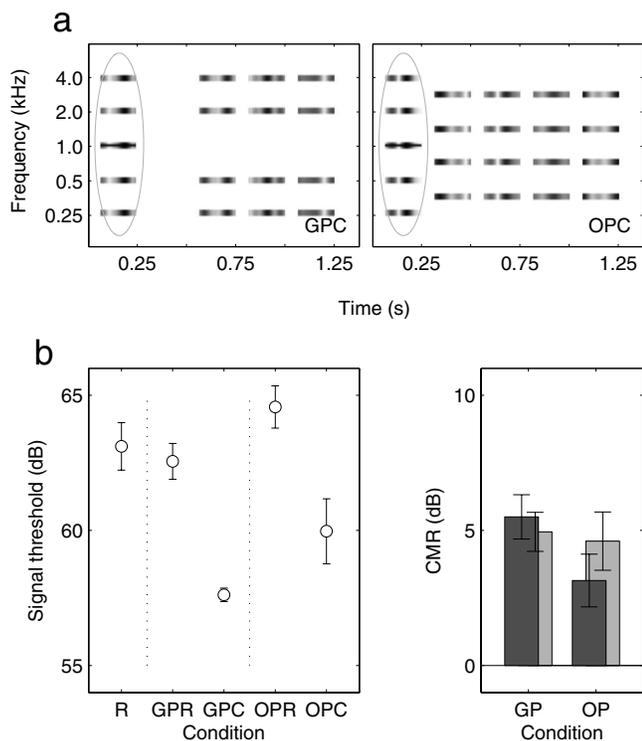
FIG. 3. Control stimulus conditions to further test the hypothesis that CMR is associated with auditory grouping. (a) Conditions with gap and postcursors (GPC, left) where the first postcursor was eliminated, and with off-frequency postcursors (OPC, right) where the postcursors were presented at intermediate frequencies. The last letter C in the abbreviations indicates that the comodulated condition is shown. (b) Mean masked thresholds and standard errors (left panel) are shown for the different stimulus conditions. The corresponding amounts of CMR for the conditions GP and OP are shown in the right panel. Dark-gray bars indicate the conditions R-GPC and R-OPC and light-gray bars indicate the conditions GPR-GPC and OPR-OPC.

R-GPC (dark-gray bar), and 4.9 dB when defined as GPR-GPC (light-gray bar). Both CMR effects were highly significant [$t(5) = 6.73$, $p = 0.001$ and $t(5) = 6.85$, $p = 0.001$, respectively] and were not significantly different from one another [$t(5) = 0.81$, $p = 0.456$]. The results from the second control experiment, where the postcursors were presented at intermediate frequencies (off-frequency postcursors-OP) show slightly elevated thresholds both for the random (OPR) and the comodulated (OPC) conditions compared to the standard thresholds, R and C. However, the amount of CMR, as measured by the difference between OPR and OPC conditions [right-hand light-gray bar; 4.6 dB, $t(5) = 4.26$, $p < 0.01$], was highly significant and was close (within 1.5 dB) to that found in the standard condition without postcursors, R-C.

In summary, the results from the two additional control conditions indicate that the effects of the postcursors cannot be ascribed to either long-term, across-trial adaptation or to a general distraction or attentional effect produced by the presence of the postcursors. The findings support the idea that retroactive effects of perceptual segregation can lead to a deterioration in target detection.

IV. DISCUSSION

The results show that across-frequency modulation processing may interact with the processes that give rise to au-ditory object and stream formation. The effects of poststimulus manipulations make an interpretation based solely on neural inhibition or forward suppression (Las *et al.*, 2005) unlikely. Instead, the current results suggest the influence of higher-level processes, whereby modulations can be processed efficiently across different frequency regions only if they form part of the same auditory object. Stated another way, the modulation analysis observed in the tasks of the present study seems to be performed on objects, rather than frequency channels.

Retroactive effects in hearing, although rare, have been reported before. For instance, in speech perception, segments of a sound occurring after the offset of a vowel can affect the perceived identity of the vowel (Darwin, 1984; Darwin *et al.*, 1989; Roberts and Moore, 1990); noise bursts can be perceived differently depending on the following vowel (Liberman *et al.*, 1952); and certain features of sounds are perceived or remembered less well when followed by a masking sound (Massaro, 1975). Warren (1970) found strong retroactive effects of context on the recognition of "missing" speech sounds, when parts of a speech sound in recorded sentences were replaced with an extraneous sound (such as a tone or a gap). Retroactive effects on the simple detection of an auditory target are less commonly observed. One example is backward masking, which occurs when a brief target, e.g., a tone pulse, is presented just before a masker (e.g., Elliott, 1962; Oxenham and Moore, 1995). Another example is related to the detection of a brief target that is gated on synchronously with a masker, similar to the well-known "overshoot" effect (e.g., Zwicker, 1965). When the masker is gated on and off with the target, thresholds are often higher than when the masker continues beyond the offset of the target (Kidd and Wright, 1994). In this case, the additional masker energy *improves* performance, possibly by eliminating the potential masking produced by the masker offset transient. Such effects are typically observed only for very short target durations and for maskers that follow immediately (within 10 ms) of the end of the masker. The targets used in the present study were much longer (187.5 ms), as were the gaps between the flankers and the postcursors (62.5 ms), suggesting that the effects observed here are probably not related to those of backward masking or overshoot.

The results place strong constraints on the search for neural correlates of across-frequency modulation processing. The current findings seem incompatible with recent physiological studies suggesting that neural correlates of CMR may be found at the brainstem level, i.e., at an early stage of auditory processing (Pressnitzer *et al.*, 2001; Verhey *et al.*, 2003; Neuert *et al.*, 2004). In Pressnitzer *et al.* (2001), some of the recorded units in the cochlear nucleus (CN) of guinea pigs showed responses consistent with perceptual CMR. The addition of a comodulated flanking band in a CMR paradigm produced a strong reduction in the response to the masker band modulation, making the signal more salient in the corresponding poststimulus time histograms. A decision statistic based on d' showed that threshold was reached at lower signal levels for the comodulated condition than for the reference condition. Using a computational model, Pressnitzer *et al.* (2001) and Meddis *et al.* (2002) demonstrated that a

simple neural circuit consisting of the inhibition of a narrow-band unit by a wideband inhibitor is able to replicate many of the physiological findings. These results thus provided evidence for an enhanced representation of the signal in the brainstem when presented in comodulated backgrounds. It is also possible that certain neurons in the brainstem that respond after a signal has ended show an altered representation when a subsequent masker is presented. However, the data shown in the present study lead us to question to what extent such across-channel processing actually reflects a direct physiological correlate of the perceptual across-channel CMR.

Several studies have investigated neural correlates of CMR at higher stages in the auditory pathways. In the primary auditory cortex of the cat, Nelken et al. (1999) considered the disruption of a neuron's envelope-following response as a correlate for CMR. Using a flanking-band experiment conceptually similar to the one considered in the present study, Nelken et al. (1999) showed that single units in the auditory cortex can demonstrate a response consistent with CMR. These units tended to lock to the envelope of the slowly fluctuating noise, whereas the addition of a low-level tone suppressed the envelope locking, a phenomenon that was referred to as "locking suppression." Such disruption of the envelope-following response to the masker at a cortical level was considered as a complementary strategy of achieving an enhanced signal representation in a modulated noise background (Nelken et al., 1999; Langemann and Klump, 2001; Verhey et al., 2003).

In a later study, the evolution of locking suppression along the auditory pathway was investigated (Las et al., 2005). Recordings were made in the primary auditory cortex (A1) and in two preceding subcortical nuclei, the inferior colliculus (IC) and the medial geniculate body (MGB). Las et al. (2005) showed that, whereas responses in IC resembled in many aspects those in CN, a new response pattern appeared in MGB and became dominant in A1, whereby the representation of the tone was more explicit in the higher stages than in the brainstem. Specifically, Las et al. (2005) proposed that the enhancement of the representation of the low-level tone in slowly fluctuating noise-by suppression of envelope locking in the ascending auditory system-could be a correlate of the formation of an auditory object (the tone) as a separate entity from the background noise.

Nelken (2004) and Las et al. (2005) proposed that, while most of the physical attributes of the sound and many interesting auditory features might already be extracted in the brainstem (e.g., the IC), the *organization* of these features into auditory objects takes place in the auditory cortex using temporal and spectral contexts at several time scales. The nonlinear interaction between stimulus components in the primary auditory cortex thus results in a more abstract representation of sounds in terms of auditory objects. The perceptual data from the present study seem to be consistent with this interpretation. However, it remains unclear to what extent such a process may account for the grouping effects observed here. The decision mechanisms that can relate neural activity and percepts in scene analysis experiments need to be specified. Such mechanisms will need to operate over more than the duration of the stimulus, to account for the retroactive effects shown here. The stimulus configurations tested in the present study might provide a basis for explicitly testing the existing hypotheses regarding the neural representation of across-channel CMR.

Finally, the results of this study also provide challenges for future models of modulation processing and perception. While the recent model of Piechowiak et al. (2007), which reflects an across-channel extension of the modulation filter-bank model of Dau et al. (1997), can account for a variety of detection and masking data, including across-channel CMR, it is not able to simulate the elimination of CMR as a result of perceptual segregation of the masker band from the flanker bands. Likewise, recent models of spectro-temporal processing in the auditory system proposed by Chi et al. (2005) cannot account for this finding. In both modeling approaches, the signal energy across time and frequency is essentially integrated linearly. This might be successful in relatively simple sound conditions but fails in more complex sound situations where perception depends on the acoustical context. However, the output of the processing models might provide some of the important auditory features as input to the "central processor." The models have been shown to be valuable as pre-processors in, for example, automatic speech recognition and objective assessment of speech quality (Hansen and Kollmeier, 1999; Tchorz and Kollmeier, 1999; Chi et al., 2005). It remains to be seen how best such representations can be manipulated to predict the effects described in this study.

ACKNOWLEDGMENTS

Altmann, C. F., Nakata, H., Noguchi, Y., Inui, K., Hoshiyama, M., Kaneoke, Y., and Kakigi, R. (**2007**). "Temporal dynamics of adaptation to natural sounds in the human auditory cortex," Cereb. Cortex **18**, 1350–1360.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organisation of Sound* (MIT, Cambridge, MA).

Bregman, A. S., and Pinker, S. (**1978**). "Auditory streaming and the building of timbre," Can. J. Psychol. **32**, 19–31.

Buss, E., and Hall, J. W. (**2008**). "Factors contributing to comodulation masking release with dichotic maskers," J. Acoust. Soc. Am. **124**, 1905–1908.

Buus, S. (**1985**). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958–1965.

Calford, M. B., and Semple, M. N. (**1995**). "Monaural inhibition in cat auditory cortex," J. Neurophysiol. **73**, 1876–1891.

Carlyon, R. P., Buus, S., and Florentine, M. (**1989**). "Comodulation masking release for three types of modulator as a function of modulation rate," Hear. Res. **42**, 37–46.

Chi, T., Ru, P., and Shamma, S. A. (**2005**). "Multiresolution spectrotemporal analysis of complex sounds," J. Acoust. Soc. Am. **118**, 887–906.

Cohen, M. F., and Schubert, E. D. (**1987**). "The effect of cross-spectrum correlation on the detectability of a noise band," J. Acoust. Soc. Am. **81**, 721–723.

Dannenbring, G. L., and Bregman, A. S. (**1978**). "Streaming vs. fusion of sinusoidal components of complex waves," Percept. Psychophys. **24**, 369–376.

Darwin, C. J. (**1984**). "Perceiving vowels in the presence of another sound: Constraints on formant perception," J. Acoust. Soc. Am. **76**, 1636–1647.

Darwin, C. J., Pattison, H., and Gardner, R. B. (**1989**). "Vowel quality changes produced by surrounding tone sequences," Percept. Psychophys. **45**, 333–342.

Dau, T., Ewert, S. D., and Oxenham, A. J. (**2005**). "Effects of concurrent and sequential streaming in comodulation masking release," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer, New York), pp. 335–343.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2893–2905.

Elliott, L. L. (**1962**). "Backward masking: Monotic and dichotic conditions," J. Acoust. Soc. Am. **34**, 1108–1115.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Grose, J. H., and Hall, J. W. (**1993**). "Comodulation masking release: Is comodulation sufficient?," J. Acoust. Soc. Am. **93**, 2896–2902.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Hansen, M., and Kollmeier, B. (**1999**). "Continuous assessment of time varying speech quality," J. Acoust. Soc. Am. **106**, 2888–2899.

Kidd, G., and Wright, B. A. (**1994**). "Improving the detectability of a brief tone in noise using forward and backward masker fringes: Monotic and dichotic presentations," J. Acoust. Soc. Am. **95**, 962–967.

Klump, G. M. (**1996**). "Bird communication in the noisy world," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Comstock, Ithaca, NY), pp. 321–338.

Langemann, U., and Klump, G. M. (**2001**). "Signal detection in amplitude modulated maskers. I. Behavioural auditory thresholds in a songbird," Eur. J. Neurosci. **13**, 1025–1032.

Las, L., Stern, E. A., and Nelken, I. (**2005**). "Representation of tone in fluctuating maskers in the ascending auditory system," J. Neurosci. **25**, 1503–1513.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Liberman, A. M., Delattre, P., and Cooper, F. S. (**1952**). "The role of selected stimulus-variables in the perception of the unvoiced stop consonants," Am. J. Psychol. **65**, 497–516.

Massaro, D. W. (**1975**). "Backward recognition masking," J. Acoust. Soc. Am. **58**, 1059–1065.

McFadden, D. (**1986**) "Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band," J. Acoust. Soc. Am. **80**, 1658–1667.

McFadden, D., and Wright, B. A. (**1992**). "Temporal decline of masking and comodulation masking release," J. Acoust. Soc. Am. **92**, 144–156.

Meddis, R., Delahaye, R., O'Mard, L., Summer, C., Fantini, D. A., Winter, I., and Pressnitzer, D. (**2002**). "A model of signal processing in the cochlear nucleus: Comodulation masking release," Acta. Acust. Acust. **88**, 387–398.

Nelken, I. (**2004**). "Processing of complex stimuli and natural scenes in the auditory cortex," Curr. Opin. Neurobiol. **14**, 474–480.

Nelken, I., Rotman, Y., and Yosef, O. B. (**1999**). "Responses of auditory cortex neurons to structural features of natural sounds," Nature (London) **397**, 154–157.

Neuert, V., Verhey, J. L., and Winter, I. M. (**2004**). "Responses of dorsal cochlear nucleus neurons to signals in the presence of modulated maskers," J. Neurosci. **24**, 5789–5797.

Oxenham, A. J., and Moore, B. C. J. (**1995**). "Additivity of masking in normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **98**, 1921–1934.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.

Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, New York), pp. 123–178.

Piechowiak, T., Ewert, S. D., and Dau, T. (**2007**). "Modeling comodulation masking release using an equalization-cancellation mechanism," J. Acoust. Soc. Am. **121**, 2111–2126.

Pressnitzer, D., Meddis, R., Delahaye, R., and Winter, I. M. (**2001**). "Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus," J. Neurosci. **21**, 6377–6386.

Roberts, B., and Moore, B. C. J. (**1990**). "The influence of extraneous sounds on the perceptual estimation of the first format frequency in vowels," J. Acoust. Soc. Am. **88**, 2571–2583.

Schooneveldt, G. P., and Moore, B. C. J. (**1987**). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. **82**, 1944–1956.

Tchorz, J., and Kollmeier, B. (**1999**). "A model of auditory perception as front end for automatic speech recognition," J. Acoust. Soc. Am. **106**, 2040–2050.

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (**2004**). "Multiple time scales of adaptation in auditory cortex neurons," J. Neurosci. **24**, 10440–10453.

Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," J. Acoust. Soc. Am. **106**, 2733–2745.

Verhey, J. L., Pressnitzer, D., and Winter, I. M. (**2003**). "The psychophysics and physiology of comodulation masking release," Exp. Brain Res. **153**, 405–417.

Warren, R. M. (**1970**). "Restoration of missing speech sounds," Science **167**, 392–393.

Wehr, M., and Zador, A. M. (**2005**). "Synaptic mechanisms of forward suppression in rat auditory cortex," Neuron **47**, 437–445.

Zwicker, E. (**1965**). "Temporal effects in simultaneous masking by white-noise bursts," J. Acoust. Soc. Am. **37**, 653–663.