ARCHIVES OF ACOUSTICS
32, 4 (Supplement), 331–336 (2007)

MODULATION MASKING PHENOMENON FOR MASKING SIGNALS OF DIFFERENT SPECTRAL AND STATISTICAL PROPERTIES

Dariusz J. KUTZNER

Adam Mickiewicz University
Institute of Acoustics
Division of Room Acoustics and Psychoacoustics
Umultowska 85, 64-614 Poznań, Poland
e-mail: konsbol@wp.pl

(*received July 15, 2007; accepted October 2, 2007*)

Modulation masking effect for modulation maskers of different spectral and statistical properties was investigated in this study. Three noisebands centred at 16 Hz were used as modulation maskers, namely: 1) 32-Hz wide Gaussian noise (GN 32 Hz); 2) 32-Hz wide low-noise noise (LNN 32 Hz) and 3) 4-Hz wide low-noise noise (LNN 4 Hz). The GN 32 Hz and LNN 32 Hz were characterized by the same power spectrum density and different probability density functions. Conversely, the LNN 32 Hz and LNN 4 Hz had the same probability density functions, but different power spectrum densities. The root mean square of the modulation maskers was normalized to 0.2. During the measurements audibility of a sinusoidal probe modulation presented simultaneously with the modulation maskers was investigated by means of the 3AFC adaptive procedure with a 3-down/1-up decision rule. The results of the experiment indicated that modulation masking was mainly determined by power density spectrum of the modulation masker, however, due to peripheral compression, temporal (statistical) properties of the masker might have played also some role in the modulation masking.

Keywords: amplitude modulation, envelope perception, modulation masking, modulation filterbank concept, Gaussian noise, low-noise noise, crest factor.

1. Introduction

A vast majority of natural sounds (for example speech, environmental signals) are characterized by temporal changes of their parameters, i.e. amplitude and/or frequency. These temporal changes, known as amplitude (AM) or frequency (FM) modulation, are widely recognized to convey an important part of information “coded” in acoustic stimuli (for example: speech). Accordingly, many experiments concerning various aspects of modulated sounds analysis and processing at different stages of the auditory system have been carried out over the last decades.

Results of many investigations eloquently suggest that auditory system performs some kind of spectral analysis with respect to signal envelope and reveals some (limited) ability to decompose its spectral content. The frequency selectivity of the auditory system in the modulation frequency domain is usually modelled and interpreted in terms of the so-called modulation filterbank (MFB) concept, which assumes an activity of an array of bandpass, linear, overlapping filters tuned into different modulation frequencies, existing at the higher stages of the auditory system [1]. In other words, the MFB concept postulates two stages of acoustic stimuli analysis in the auditory system, namely: cochlear filtering in the audible frequency domain (the auditory filters), which is followed by the modulation filters functioning in the envelope rate domain. Although, the hypothetical modulation filters are thought to work analogously to the auditory filters, it is assumed that their selectivity is much poorer ($Q \approx 1$ [2]) than that of the auditory filters ($Q \approx 8$ [3]). It should be stressed, that the most recent model of modulation filters argues that the sound envelope is processed at the higher stages of the auditory system in a set of filters characterized by non-negative impulse response [4, 5].

The selective properties of the auditory system in the modulation rate domain have been described by results of many psychoacoustical experiments in which similar effects, as in the audible frequency domain, were observed: "tuning" [1, 2, 6, 7], frequency discrimination [8], perception of modulation asynchrony [9], some phase effects [10] and independence of masking from masker temporal repetition [11]. Thought, the mentioned above experiments concerning different aspects of modulation perception, the existence of "critical band" in the modulation rate domain was proved in the measurements concerning the so-called modulation masking phenomenon, i.e. the effect in which modulation perception is disturbed in a presence of another modulation signal, called as a "modulation masker". It has turned out that the modulation masking effectiveness increased when spectral separation (in the modulation frequency domain) between the masked (probe) and the masker was decreased (like in the audible frequency domain), which can be explained in terms of decreasing signal-to-noise ratio, SNR, in a hypothetical modulation channel tuned to the masked modulation rate [1, 2, 6, 7]. However, it should be stressed that modulation masking is also interpreted in terms of perceptual grouping of the probe and the masking modulation [12].

The main purpose of the present investigation was to gain further insight into mechanisms underlying the modulation masking, namely to examine this phenomenon with respect to spectral as well as statistical properties of the masker.

It had been assumed that if the masking changed with a change in spectral properties of the modulation masker only, it would be another strong argument supporting a concept of multi-channel envelope processing at the higher stages of the auditory pathway. In contrast, if the masking patterns changed when statistical properties of the masker were modified, such an effect might suggest that a temporal representation of the masker has some influence on this phenomenon.

2. Psychoacoustical measurements

2.1. Stimuli

General formula describing the stimuli used in this investigation was as follows:

$$y(t) = (1 + x_{\mathrm{mod}}(t)) \sin(2\pi f_c t), \tag{1}$$

where f_c is a carrier frequency; $x_{\mathrm{mod}}(t)$ is a modulation signal:

$$x_{\mathrm{mod}}(t) = m_p \cos(2\pi f_p t + \phi) + m_m x_{\mathrm{mask}}(t), \tag{2}$$

where m_p, f_p and ϕ are the probe (masked) modulation depth, frequency and initial phase, respectively; $x_{\mathrm{mask}}(t)$ is the masking modulation, m_m its modulation depth. Three different masking signals $x_{\mathrm{mask}}(t)$ were used: 1) a 32-Hz wide Gaussian noise (GN 32 Hz); 2) a 32-Hz wide low-noise noise (LNN 32 Hz) and 3) a 4-Hz wide low-noise noise (LNN 4 Hz). All the maskers were centered at 16 Hz and their root-mean-square value, m_{mrms}, was normalized to 0.2. Therefore, 2 kinds of noise waves and 2 spectral bandwidths were used. Figure 1 presents power spectrum densities (upper panels) and distributions (lower panels) of the respective modulation maskers: GN 32 Hz (left panels), LNN 32 Hz (middle panels) and LNN 4 Hz (right panels).

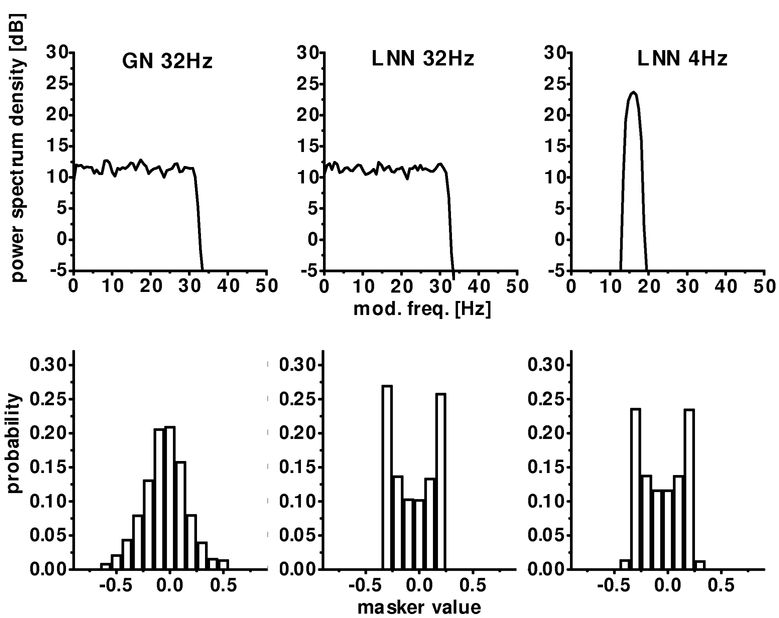

Fig. 1. Power spectra (upper panels) and histograms (bottom panels) of masking modulators used in the measurements: GN 32 Hz (left panels), LNN 32 Hz (middle panels) and LNN 4 Hz (right panels).

As can be seen from Fig. 1, GN 32 Hz and LNN 32 Hz have the same power spectra and different histograms: GN 32 Hz reveals the normal distribution, while the histogram of LNN 32 Hz is “U-shaped”. On the other hand, LNN 32 Hz and LNN 4 Hz have different power spectra, but are characterized almost identical histograms. Therefore, if modulation perception were determined by spectral properties of the masking signals, the identical masking patters for GN 32 Hz and LNN 32 Hz would be obtained and different pattern for LNN 4 Hz might be observed. Conversely, if the modulation masking depended on statistical properties of the masker, similar masking patterns could be observed for signals of comparable histograms (LNN 32 Hz and LNN 4 Hz), while the masking pattern for GN 32 Hz would be different. GN 32 Hz was generated via inverse fast Fourier transformation (*ifft*). Low-noise noise signals were generated by means of an iterative procedure described by KOHLRAUSCH *et al*. [13]. The carrier frequency f_c was 4 kHz. A probe modulation initial phase ϕ was randomised (the uniform distribution from $-\pi$ to π).

2.2. Method and apparatus

The modulation detection thresholds were determined for the following probe modulation frequencies (rates) f_p: 4, 6, 8, 12, 16, 24, 32, 48 and 64 Hz. The thresholds were determined in an absence of the modulation maskers (unmasked thresholds) as well as in a presence of the masker (masked thresholds). The standard *3AFC* staircase adaptive procedure with a *1-up/3-down* decision rule (converging to the 79.4%-correct response point on a psychometric function) [14] was used to determine the modulation thresholds. During measurements, three stimuli (“intervals”) were presented to a subject, whereas only one contained a probe modulation (“signal interval”). The position of the “signal interval” was randomised. In the unmasked and masked thresholds measurements $m_{mrms} = 0$ and $m_{mrms} = 0.2$, respectively, for all the intervals. A subject task was to indicate the signal interval (i.e. the one that differs perceptually from the others). The root mean square of the probe modulation, m_{prms}, was increased by some value (step) after one (*1-up*) incorrect response and was decreased after three succeeding correct responses (*3-down*). The listeners were informed whether their response was correct or incorrect (“feedback response”). The measurement lasted until 12 turnpoints were determined. The initial step was 4 dB (in terms of $20 \log m_{prms}$) and was reduced to 1 dB after first 4 turnpoints. A final threshold value was computed as a geometric mean of the last 8 turnpoints. The stimuli were generated by means of the RP-2 real time 24-bit processor (Tucker-Davis Technologies, TDT, System 3), passed through headphone amplifier HB7 (TDT System 3) and presented to the subjects monaurally via the Sennheiser HD 580 headphones. Finally, the modulation masking patterns were computed as difference between masked and unmasked modulation threshold values.

3. Results and discussion

The determined modulation masking patters were subjected to the three-way analysis of variance (ANOVA) with respect to the following factors: probe modulation rate

("rate"), modulation masker type ("type") and also inter-individual differences were taken into consideration ("subject"). The ANOVA revealed that the parameters of modulation turned out to be highly statistically significant ("rate" $\{F(8,242) = 82.72, p \ll 0.001\}$; "type" $\{F(2,242) = 461.73, p \ll 0.001\}$), while the "subject" factor was statistically insignificant $\{F(2,242) = 0.36, p = 0.69\}$. Moreover, a statistically significant interaction between "rate" and "type" $\{F(16,242) = 22.11, p \ll 0.001\}$ indicated not only quantitative, but also qualitative differences between the masking patters obtained for respective maskers.

Figure 2 presents modulation masking patterns determined for the respective maskers: GN 32 Hz (squares), LNN 32 Hz (circles) and LNN 4Hz (triangles). The respective points depict data averaged across subjects and repetitions.

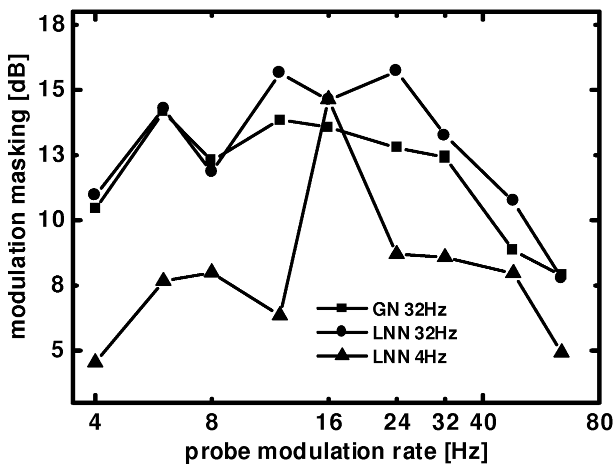

Fig. 2. Modulation masking patterns obtained for GN 32 Hz (squares), LNN 32 Hz (circles) and LNN 4 Hz (triangles).

As can be seen, the masking patterns reflect power spectrum density of the masker: these obtained for GN 32 Hz and LNN 32 Hz are approximately flat for probe rates up to 32 Hz; above this frequency a decrease in masking values is observed. The pattern for LNN 4 Hz reflects a clearly visible local maximum in power spectrum density of this masker: the masking effectiveness decreases as spectral separation between the probe and the masker increases. This remark is in agreement with the results of the previous measurements concerning narrowband modulation maskers [1, 2, 6, 7]. Although, the patterns obtained for GN 32 Hz and LNN 32 Hz are quite similar, an additional *post hoc* analysis (*Scheffe* test, $p = 0.007$) surprisingly revealed statistically significant differences between these results. Since the both maskers have the same power density spectra, the differences in their phase spectra and, consequently, parameters describing temporal representation of the maskers might have influenced the modulation masking. As for GN 32 Hz crest factor $C_f \approx 3.5$ ($C_f = \max|x_{\text{mask}}|/m_{mrms}$) and for LNN 32 Hz $C_f \approx 1.6$, it might have been that in the case of GN 32 Hz relatively high instantaneous values of masking modulation were much more effectively attenuated by cochlear compression [15], than in the case of LNN 32 Hz and slightly more effective modulation masking was observed for LNN 32 Hz.

References

[1] DAU T., *Modelling auditory processing of amplitude modulation*, Oldenburg BiS, 1996.

[2] SĘK A. P., SKRODZKA E. B., *Tuning in the amplitude modulation rate domain*, Archives of Acoustics, **23**, 4, 491–498 (1998).

[3] MOORE B. C. J., *An Introduction to the Psychology of Hearing*, 5th Ed., Academic Press, London 2003.

[4] KUTZNER D., *Modelling the sound envelope auditory processing using the non-negative-impulse-response modulation filters concept. I. Initial simulations*, Archives of Acoustics, **31**, 1, 29–45 (2006).

[5] KUTZNER D., *Sound envelope processing in the auditory system*, Ph.D. Dissertation, Institute of Acoustics, Adam Mickiewicz University, Poznań 2007.

[6] BACON S. P., GRANTHAM D. W., *Modulation masking: effects of modulation frequency, depth and phase*, J. Acoust. Soc. Am., **85**, 2575–2580 (1989).

[7] LEMAŃSKA J., SĘK A. P., RYBICKA W., *Masking in the amplitude modulation rate domain*, Archives of Acoustics, **28**, 3, 151–159 (2003).

[8] LEMAŃSKA J., SĘK A. P., SKRODZKA E. B., *Discrimination of the amplitude modulation rate*, Archives of Acoustics, **27**, 1, 3–22 (2002).

[9] KUTZNER D., LEMAŃSKA J., SĘK A. P., *Detection of asynchronicity in the modulation rate domain*, Archives of Acoustics, **30**, 3, 291–307 (2005).

[10] SĘK A. P., MOORE B. C. J., *Testing the concept of a modulation filter bank: The audibility of component modulation and detection of phase change in three-component modulators*, J. Acoust. Soc. Am, **113**, 5, 2803–2811 (2003).

[11] KUTZNER D., *Modulation masking for recurrent low-noise noise masker*, Archives of Acoustics, **31**, 4, 543–549 (2006).

[12] OXENHAM A. J., DAU T., *Modulation detection interference: effects of concurrent and sequential streaming*, J. Acoust. Soc. Am., **110**, 1, 402–408 (2001).

[13] KOHLRAUSCH A., FASSEL R., VAN DER HEIJDEN M., KORTEKAAS S., VAN DE PAR S., OXENHAM A. J., *Detection of tones in low-noise noise: Further eveidence for the role of envelope fluctuations*, Acustica, Acta-Acustica, **83**, 659–669 (1997).

[14] H. LEVIT, *Transformed up-down methods in psychoacoustics*, J. Acoust. Soc. Am., **49**, 467–477 (1971).

[15] BACON S.P., FAY R.R., POPPER A.N., *Compression. From Cochlea to Chochlear Implants*, Spinger Handbook of Auditory Research, Vol. 17 (2004).