

**Acoustics'08
Paris**
June 29-July 4, 2008
www.acoustics08-paris.org

Additivity of auditory masking using Gaussian-shaped tones

B. Laback^a, P. Balazs^a, G. Toupin^a, T. Necciari^b, S. Savel^b, S. Meunier^b, S. Ystad^b and R. Kronland-Martinet^b

^aAustrian Academy of Science / Acoustics Research Institute, Wohllebengasse 12-14, 1040 Vienna, Austria

^bCNRS-LMA, 31, chemin Joseph Aiguier, 13402 Marseille, France
bernhard.laback@oeaw.ac.at

Both temporal and spectral masking have been studied extensively in the literature. Mostly, they have been regarded as separate phenomena. Very little is known about the interaction between these two effects, i.e. masking in the time-frequency domain. Data on the time-frequency spread of masking evoked by a single Gaussian-shaped tone pulse are presented in an accompanying study at the same conference (Necciari et al.). The current study gathers data on the additivity of masking by up to four, approximately equally effective Gaussian maskers (ERB=600 Hz, ERD=1.7 ms), separated either along the time or the frequency axis. For temporal shift, the amount of masking increases with the number of maskers, with excess masking (exceeding linear additivity) of up to 25 dB. For frequency separation excess masking amounts up to 15 dB, and the higher-frequency masker (relative to the target) contributes more to the additivity than the lower-frequency maskers. The data suggest strong basilar membrane compressivity even for such short stimuli. Combined with the single masker time-frequency spread data, these data may serve as a basis for modeling time-frequency masking effects in complex signals.

1 Introduction

Auditory masking refers to the fact that the presence of a masker signal decreases the audibility of a target signal. Masking has been studied extensively in the frequency domain and in the time domain (see [1] for a review) but little is known about masking in the time-frequency domain. Frequency-domain studies most often used relatively long stimuli, which allows to create stimuli with very narrow spectra. On the other hand, time-domain studies most often used relatively broadband stimuli, allowing to create stimuli with fast temporal changes. Existing studies provide very little information on how the masking effect spreads across the time-frequency plane. This seems to be due to the difficulty to control both domains accurately at the same time.

A signal which has minimal spread in time and frequency at the same time is a Gaussian-shaped tone pulse. It minimizes Heisenberg's uncertainty principle [2] and has a Gaussian shape both in time and frequency. Due to these properties, Gaussians seem to be optimal to study time-frequency masking. The use of this type of signal for studying time-frequency masking was motivated by a study attempting to measure the shape of the auditory time-frequency window [3].

An accompanying study presented at the same conference [10] measured the spread of masking evoked by a single Gaussian using the same signal type as target. Here we present data on the additivity of the masking effect of multiple Gaussian maskers in the time and in the frequency domain. These data have both theoretical and practical implications. From the theoretical point of view, they will show if and under which conditions short Gaussians elicit nonlinear additivity of masking. In the literature, nonlinear additivity of masking has been found both for temporally non-overlapping maskers and for temporally overlapping maskers. Maskers that were used were broadband and short, broadband and long, or narrowband and long (see [4] for a summary). Here we use a stimulus that is compact in the time-frequency domain for measuring both temporally overlapping and non-overlapping conditions. From the practical point of view, the data may provide a basis to model and predict masking effects in arbitrarily complex signals.

2 General methods

2.1 Stimuli

All masker and target signals were Gaussian-shaped tone bursts. They are defined by the expression

$$s(t) = \sqrt{\Gamma} \cdot \sin\left(2\pi f_0 t + \frac{\pi}{4}\right) \cdot e^{-\pi(\Gamma t)^2}, \quad (1)$$

where f_0 is the tone frequency and $\Gamma = \alpha f_0$ defines the equivalent rectangular bandwidth of the signal. Γ was set constant at 600 Hz (-6 dB bandwidth: 563 Hz). The equivalent rectangular duration was 1.7 ms (-6 dB duration: 1.6 ms). The phase is set to $\pi/4$ to make the energy of the signal independent of f_0 . Raised-cosine onset and offset ramps (10 ms) were applied to the whole stimulus in an experimental interval (which could consist of multiple Gaussians). In some experimental conditions, a lowpass-filtered, continuous Gaussian noise was used to mask combination tones.

2.2 Procedure

Thresholds were measured using an adaptive three-interval three-alternative forced choice (3I-3AFC) paradigm (odddity task). The target was presented randomly in one of the three intervals. In absolute threshold measurements, the other two intervals were silent. In masked threshold measurements, all three intervals contained the masker. The subjects were instructed to indicate the interval which sounded different from the other two by pressing one of three buttons. Each interval was marked visually on a computer screen. Response feedback was provided after each trial by highlighting the interval containing the target. In addition, the correctness of the response was indicated by inserting a text.

In the adaptive procedure the target level was varied using the three-down one-up rule to converge at the 79.4% point on the psychometric function. The target level started at a sufficiently high level to be easily audible. The initial stepsize was 5 dB and was halved after two reversals. A run was terminated after 12 reversals and the threshold was calculated by averaging the last eight reversals.

The different experimental conditions were presented in random order. One experimental block contained one repetition for each condition. The entire block was then

repeated several times while tracking learning effects for each condition. Testing was continued until at least three stable thresholds (standard deviation less than 3.4 dB) were obtained for each condition. Runs where the adaptive procedure did not converge were discarded.

For conditions with multiple maskers, experimental conditions were presented in blocks with equal numbers of maskers. We chose this design to avoid switching too often between conditions which may sound quite different and thus hoped to obtain more stable thresholds. The order of blocks was randomized.

The level of a Gaussian was specified by the equivalent dB SPL level of a long-lasting sinusoid of the same frequency and the amplitude corresponding to the peak amplitude of the Gaussian.

2.3 Subjects

Six normal hearing listeners participated in the experiments. Four of the listeners participated in both experiments, while NH24 participated in experiment 1 only and NH14 in experiment 2 only. None of the listeners had an indication of a hearing disorder and the audiometric thresholds were within 10 dB of the normal range. All listeners except NH24 and NH14 had previous experience with psychoacoustic experiments. All listeners received training with all masking conditions.

2.4 Apparatus

Experiment control and stimulus generation was performed using the custom-made software ExpSuite and Matlab. Stimuli were output at a sampling rate of 44.1 kHz and a 24 bit resolution via Pure Data, sent to an external D-A converter (AD/DA 2402, Digital Audio Denmark), a headphone amplifier (HB6, TDT) and an attenuator (PA4, TDT). A circumaural headphone (HDA200, Sennheiser) was used. The experiment was performed in a double-walled sound booth.

3 Experiment 1: Multiple maskers shifted in time

The aim of this experiment was to measure the additivity of masking from four Gaussian maskers, separated in time, on a Gaussian target. Each of the four maskers was adjusted to be equally effective in masking the target.

3.1 Experimental conditions

All signals, the target and the maskers, had a frequency of 4000 Hz. The four maskers were temporally shifted relative to the target (peak-to-peak distance) by -24 ms (M1), -16 ms (M2), -8 ms (M3), and +8 ms (M4). Thus, there were three forward maskers and one backward masker. In order to obtain equally effective maskers, each masker has to be adjusted in level. The temporal shifts were determined in pilot tests to fulfill the following requirements: (1) maximum temporal shift between the maskers, (2) equal shift between M1, M2, M3, the target, and M4, and (3) not

exceed a comfortable level for both the most distant forward masker (M1) and the backward masker (M4) which required the highest levels.

First, the absolute threshold of a single Gaussian was measured. In the second stage, the equally effective masker levels were determined. Using an iterative approach, the masker level was adjusted from run to run to obtain a masked threshold of the target approximately 8 dB above its threshold in quiet. On average, about four to six iterations (masker level changes) were required. After the final masker level had been determined, further measurements at this level were performed, which represent the masked thresholds for the single masker conditions. For each listener and masker, the masked threshold had to be significantly above the absolute threshold of the targets ($p < 0.05$). Otherwise, the iterative procedure was continued with an adjusted masker level. Note that a more direct approach to determine equally effective masker levels would be to vary the masker level instead of the target level. However, when using this method in pilot tests we had often obtained non-converging adaptive procedures. Therefore, we discarded this method.

In the third stage, the multiple masker test, masked thresholds were measured for selected combinations of maskers. These combinations were 2-3, 3-4, 1-2-3, 2-3-4, and 1-2-3-4. After completion of the multiple masker tests, the single masker conditions were retested to check for learning effects. There were no systematic learning effects observed.

Subject	Absolute threshold (dB SPL)	Equally effective masker level			
		M1 (-24 ms)	M2 (-16 ms)	M3 (-8 ms)	M4 (+8 ms)
NH2	19.3	77.3	76.3	65.3	78.3
NH19	11.3	71.3	62.3	52.3	72.3
NH23	21.3	79.3	75.3	63.3	72.3
NH24	15.3	71.3	62.3	55.3	69.3
NH25	16.3	79.3	65.3	55.3	71.3

Tab. 1: Exp. 1: Absolute target thresholds and equally effective masker levels, resulting in masked thresholds of the target approximately 8 dB above the absolute threshold.

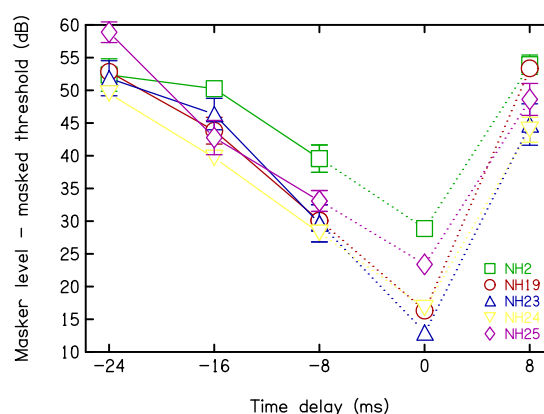


Fig. 1: Experiment 1: For each masker (at -24, -16, -8 and +8 ms) the difference between the masker level and the corresponding masked threshold of the target is shown. Error bars indicate 95% confidence intervals.

3.2 Results and discussion

Table 1 shows the absolute thresholds of the target and the equally effective masker levels leading to masked thresholds about eight dB above the absolute threshold. Fig. 1 shows, for each single masker, the difference between the masker level and the corresponding masked threshold of the target. The values at a shift of 0 ms are arbitrarily determined by linear extrapolation from the values at -16 and -8 ms to facilitate reading the plot. According to the expectation, with increasing shift a higher masker level is required to mask the target. Furthermore, backward masking is much weaker than forward masking: the equally effective masker level at +8 ms (backward masking) corresponds to that between -16 and -24 ms (forward masking).

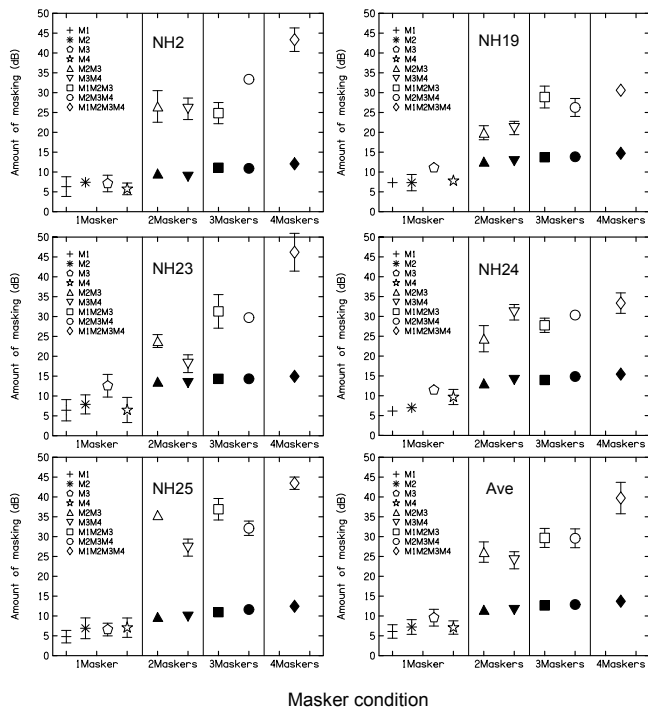


Fig. 2: Experiment 1: Different masker combinations are indicated with symbols shown in the legend. Error bars indicate 95% confidence intervals.

Figure 2 shows the masked thresholds for the single maskers and for all combinations of maskers. Both individual results and the average result across listeners are shown. The ordinate depicts the amount of masking, i.e. the difference between the masked threshold and the absolute threshold. Despite considerable inter-subject variability, we focus on the description of the average data. All these effects are substantiated by a Repeated Measures ANOVA performed on the group data.

Combining two maskers increased the amount of masking relative to the single maskers ($p < 0.05$; ave: 18 dB). Adding a third masker increased the amount of masking further ($p < 0.05$; 5 dB). Adding a fourth masker increased masking even further ($p < 0.05$; 11 dB). On average across the group, there was no systematic difference in the additivity of masking across subjects when combining either forward maskers only or forward maskers with a backward masker ($p > 0.05$). However, the individual subjects showed exceptions from this. For example, NH2

showed more masking in condition 2-3-4 compared to 1-2-3.

The filled symbols in Fig. 2 show the amount of masking as it corresponds to a linear additivity model. According to this model, combining two equally effective maskers increases the amount of masking relative to the single maskers by 3 dB. It is obvious that the data show a large amount of additional masking relative to the linear prediction. The difference between the linear prediction and the actual data is referred to as “excess” masking. On average across the subjects, excess masking amounts to 14 dB for the two-masker conditions, 17 dB for the three-masker conditions and 26 dB for the four-masker-condition. These results are qualitatively consistent with previous studies showing nonlinear additivity of temporal masking for maskers which have either no or minimum temporal overlap [5-8].

It is interesting to consider that nonlinear additivity of masking occurs for signals with an equivalent rectangular duration of only 1.7 ms. According to widespread models of masking additivity [4], the nonlinearity is a result of individual maskers being independently compressed on the basilar membrane (BM) and subsequently being summed. Our finding of highly nonlinear masking additivity indicates that the Gaussians are subject to strong BM compression. This is consistent with the physiological finding that the BM starts to be highly compressive within 500 to 700 μ s after the onset of a click (measured in the 8-10 kHz region in the chinchilla) [9].

4 Experiment 2: Multiple maskers separated in frequency

This experiment is analogous to experiment 1, with the difference that the maskers were separated in frequency rather than in time.

Subject	Absolute threshold (dB SPL)	Equally effective masker level			
		M1 (-7 ERBs)	M2 (-5 ERBs)	M3 (-3 ERBs)	M4 (+3 ERBs)
NH2	32.2	63.9	59.6	57.3	69.2
NH19	26.2	53.9	51.6	48.3	58.2
NH23	24.2	57.9	54.6	53.3	63.2
NH25	25.2	57.9	54.6	53.3	63.2
NH14	29.2	59.9	57.6	55.3	66.2

Tab. 2: Experiment 2: as for Tab. 1

4.1 Experimental conditions

All signals, the target and the maskers, were presented simultaneously. The target had a frequency of 5611 Hz. The four maskers were separated in frequency from the target by -7 ERBs (M1), -5 ERBs (M2), -3 ERBs (M3), and +3 ERBs (M4). Thus, three maskers had lower frequencies and one masker had a higher frequency relative to the target. The choice of frequencies was based on pilot experiments, following the same requirements as in experiment 1, but in the frequency domain. However, here we were not able to fulfill all requirements. The compromise was to have smaller frequency separations between maskers M1, M2, and M3 (two ERBs) than between M3, the target, and M4 (three ERBs).

Simultaneous maskers may involve the perception of combination tones. To test their contribution, we tested in pilot tests with two listeners (NH2 and NH23) a critical case, namely M3+T, with and without lowpass noise. Since even repeated measurements showed no significant difference in the masked thresholds, we decided to omit the lowpass noise at all.

4.2 Results and discussion

Table 2 shows the absolute thresholds of the target and the equally effective masker levels leading to masked thresholds about eight dB above the absolute threshold. Fig. 3 shows, for each single masker, the difference between the masker level and the corresponding masked threshold of the target. The values at zero ERBs frequency separation are arbitrarily determined by linear extrapolation from the values at -5 and -3 ERBs to facilitate reading the plot. According to the expectation, with increasing frequency separation a higher masker level is required to mask the target. Furthermore, in correspondence with the masking literature [1], masking towards higher frequencies is much stronger than towards lower frequencies.

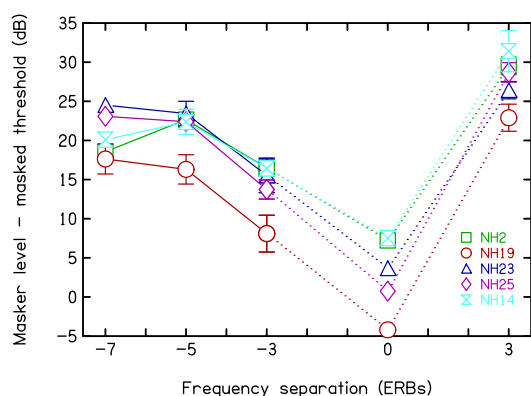


Fig. 3: Experiment 2: For each masker (at -7, -5, -3 and +3 ERBs) the difference between the masker level and the corresponding masked threshold of the target is shown. Error bars indicate 95% confidence intervals.

Fig 4 shows the masked thresholds for the single maskers and for all combinations of maskers, analogous to Fig. 2. Even though there is some inter-subject variability, all subjects show the same effects and thus only the average data are discussed. In contrast to experiment 1, the effect of combining two maskers differs strongly between the masker combinations. For masker pair 2-3 there is very little increase ($p < 0.05$; 3 dB) whereas for masker pair 3-4 there is a strong increase ($p < 0.05$; 15 dB). Adding masker M1 to the pair 2-3 increases masking ($p < 0.05$; 5 dB) while adding M2 to the pair 3-4 has no effect ($p > 0.05$; ave: 0 dB). Adding a fourth masker has a strong effect relative to the combination 1-2-3 ($p < 0.05$; 9 dB) but not relative to the combination 2-3-4 ($p > 0.05$; 0 dB).

There is no excess masking for the pair 2-3 but strong excess masking for the pair 3-4 (of 15 dB). Excess masking for the triples ranges from 5 dB (1-2-3) to 13 dB (2-3-4) and increases only marginally to 14 dB for the quadruple. Overall, the results are consistent with the concept that

excess masking occurs always when adding a masker that has no energy overlap with at least one of the maskers it is combined with. This seems to hold particularly if the maskers to which the masker is added are themselves overlapping.

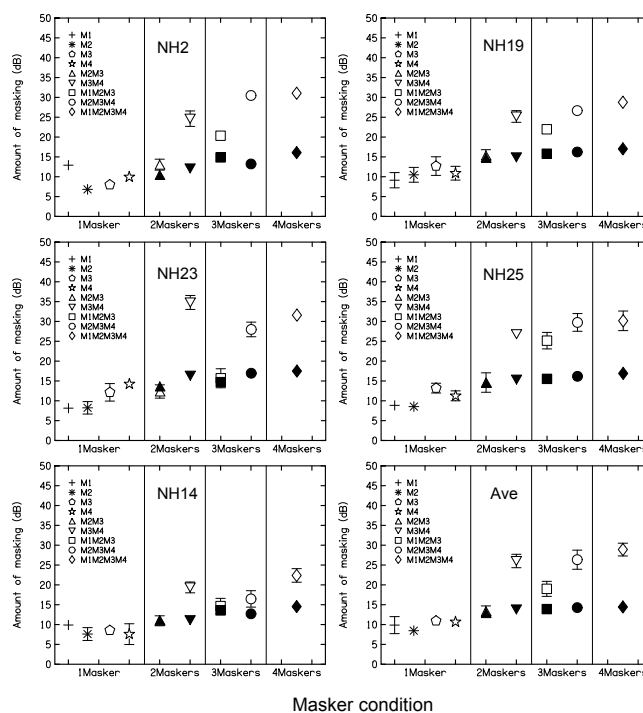


Fig. 4: Experiment 2: Different masker combinations are indicated with symbols shown in the legend. Error bars indicate 95% confidence intervals.

5 General Discussion and Conclusions

The most important finding of this study is that short Gaussian tone bursts elicit highly nonlinear additivity of masking if they are physically non-overlapping. In most cases and for most subjects, the amount of excess masking increased monotonically with the number of maskers, provided that the added masker resulted in a masker combination with no energy overlap. This could be observed mainly for the temporally shifted maskers, where there was no energy overlap between the individual maskers. This appears to indicate strong compressivity of the BM for short Gaussian tones.

The data may serve as a basis for modeling time-frequency masking effects in complex signals. Every complex signal can be decomposed into a time-frequency matrix of Gaussians with appropriate amplitudes and phases. Having a model for time-frequency masking effects between individual Gaussians would then allow to predict which Gaussian components of the time-frequency matrix corresponding to the complex signal are masked and can be removed without audible effects. This may be beneficial for signal codecs or for sound synthesis applications.

Acknowledgments

We would like to thank Piotr Majdak for providing support in the development of the software for the experiments. The work partly supported by WTZ (project AMADEUS).

the chinchilla cochlea,” J. Acoust. Soc. Am. 103, 1972-1989 (1998)

[10] Necciari, T., Savel, S., Meunier, S., Ystad, S., Kronland-Martinet, R., Laback, B., and Balazs, P. Auditory masking using Gaussian-windowed stimuli,” presented at Acoustics 08 conference (2008)

References

- [1] Moore, B. J. C. An introduction to the psychology of hearing, Academic Press Limited, London (1992)
- [2] K. Gröchenig, *Foundations of time-frequency analysis*, Birkhäuser, Boston (2001)
- [3] van Schijndel, N. H., Houtgast, T., Festen, J. M. “Intensity discrimination of Gaussian-windowed tones: indications for the shape of the auditory frequency-time window,” J Acoust Soc Am. 105, 3425-35 (1999)
- [4] Humes, L. E., Jeseadt, W. “Models of the additivity of masking,” J. Acoust Soc Am. 85, 1285-1294 (1989)
- [5] Penner M. J. “The coding of intensity and the interaction of forward and backward masking”, J. Acoust. Soc. Am. 67, 608-616 (1980)
- [6] Wilson R. H., Carhart R. “Forward and Backward Masking: Interactions and Additivity”, J. Acoust. Soc. Am. 49, 1254-1263 (1971)
- [7] Widin G. P. , Viemeister N. F. “Masker interaction in pure-tone forward masking”, J. Acoust. Soc. Am. 68, 475-479 (1980)
- [8] Oxenham, A. J. and Moore, B.C.J. “Modeling the additivity of nonsimultaneous masking,” Hearing Research, 80, 105-118 (1994)
- [9] Recio A, Rich NC, Narayan SS, Ruggero MA. “Basilar-membrane responses to clicks at the base of