

Masking a Frequency Band in a Musical Fragment Played by a Single Instrument

P. KLECZKOWSKI*, M. PLEWA AND M. PLUTA

Department of Mechanics and Vibroacoustics, AGH University of Science and Technology

al. A. Mickiewicza 30, 30-059 Kraków, Poland

The phenomenon of masking in the environment of musical sounds has not yet been deeply investigated. An experiment was conducted to find out whether masking of a band of frequencies in a musical instrument sound by the remaining spectral content of that instrument could be explained by “energetic” masking, i.e. masking resulting from overlapping patterns of excitation on the basilar membrane. An isolated, about one octave wide, band of frequencies of the saxophone sound playing a couple of notes legato was masked alternatively by two maskers: CBN — the band of noise occupying the same frequency range plus the remaining spectral content of the saxophone sound, and BN — the same band of noise acting alone. For both maskers, the level of noise was increased gradually to detect the respective thresholds of masking. Masker CBN was more efficient in masking by about 15 dB, indicating that the harmonics below and above the analysed band contributed substantially more to masking than noise inside the band. As energetic masking of those harmonics was weak, non-energetic or “informational” masking prevailed.

PACS: 43.66.Dc, 43.66.Lj, 43.66.Ba, 43.60.Hj

1. Introduction

When a musical instrument is played, complex masking phenomena between its partials take place. This issue is of particular interest in view of the finding that when listening to a mixture of musical instrument sounds it is possible to remove large parts of these sounds from their time-frequency representation with marginal audible effect [1].

A psychoacoustic phenomenon that should be considered in the first place is informational masking. In 1975, Pollack first used the two terms: “informational masking” and “energetic masking” [2]. Both terms have not yet found a widely accepted definition. Energetic masking usually means peripheral masking, resulting from overlapping patterns of excitation on the basilar membrane. Most of the research in masking falls within this category. The simplest and most often used explanation of the term “informational masking” is: “masking that occurs beyond that which can be attributed to energetic masking” [3]. A meaningful contribution to the research on informational masking was done by Watson et al. and has been reviewed in [4].

In the process of listening to an instrument playing, the listener is faced with uncertainty, therefore informational masking may be expected. The objective of this work was to investigate whether masking of a frequency

band in an instrument by frequency components outside of that band was energetic or non-energetic, i.e. informational. An isolated band of frequencies of the saxophone sound playing a couple of notes legato (the signal in masking experiment) was masked alternatively by two maskers: CBN (for complex plus band of noise) — the band of noise occupying the same frequency range plus all frequency components of the saxophone sound outside of that band, and BN (for band of noise) — the same band of noise acting alone. For both maskers, the level of noise was increased gradually in order to detect the respective thresholds of masking. The threshold was found to be substantially higher for masker BN, indicating that the contribution of noise in masker CBN was low. The band of noise was a purely energetic masker in this experiment, therefore the main component of masker CBN was non-energetic, i.e. informational, according to the present terminology.

2. Experiment

2.1. Stimuli

The basis for the construction of the stimuli was a recording of a couple of notes played legato on a saxophone, lasting 3.7 s. The time-frequency representation of this sound is shown in Fig. 1, and its long-term frequency spectrum is shown in Fig. 2. Out of the full spectrum two frequency bands were chosen, marked grey in Fig. 2.

In part 1 of the experiment masker CBN was evaluated and two sound samples (A and B) were compared:

* corresponding author; e-mail: kleczkow@agh.edu.pl

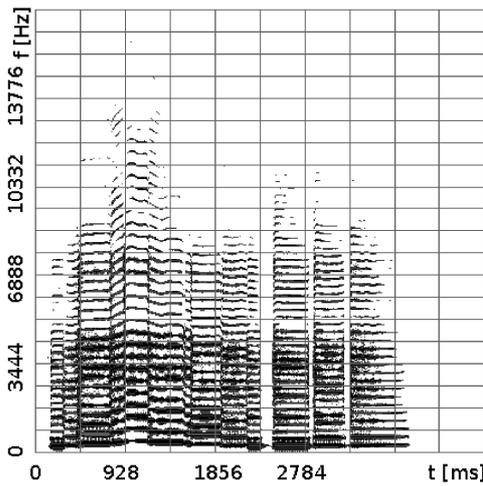


Fig. 1. The time-frequency representation of the saxophone sound used as a basis for the stimuli.

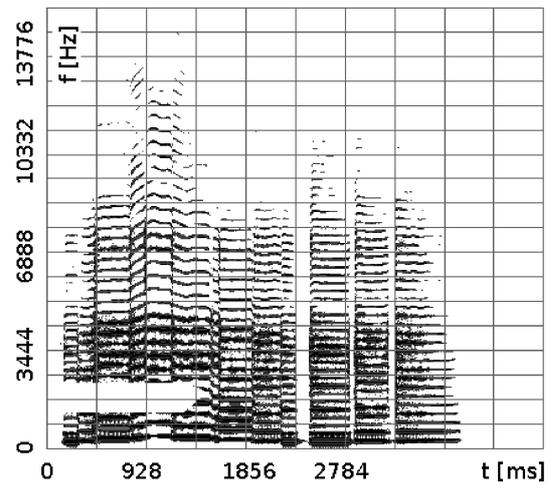


Fig. 4. The saxophone sound with a spectral “gap” — a band of frequencies (1227–2261 Hz) removed from it up to some point in time.

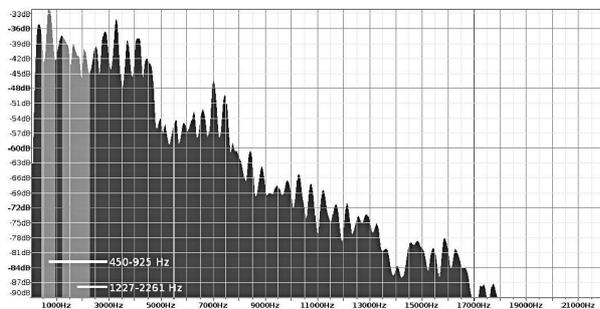


Fig. 2. The long-term frequency spectrum of the saxophone sound.

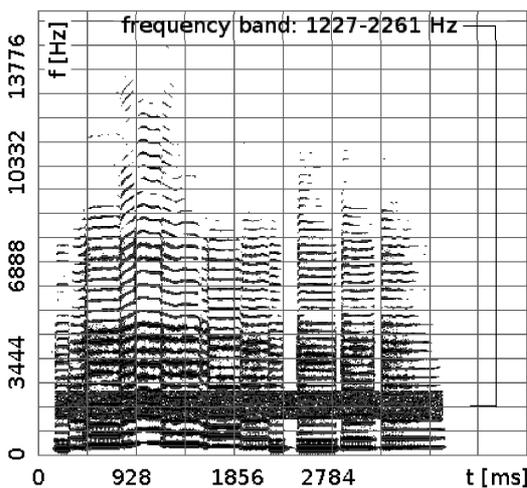


Fig. 3. Example of stimulus A in part 1 of the experiment: the saxophone sound with noise added in a band of frequencies (signal + masker CBN, noise at -11 dB).

A was the sound of Fig. 1, with noise in a band of frequencies added, as shown in Fig. 3 for the case of the band from 1227 Hz to 2261 Hz, and the noise at -11 dB. Sample B was based on the sound of Fig. 1 but the same band of frequencies was removed from it up to some point in time, creating a spectral “gap” as is shown in Fig. 4. Noise was also added to B. The noise signals added to A and B were identical copies. In order to eliminate possible biases resulting from cutting out the part of the sound only at its beginning, actually two comparisons were performed, with two kinds of the sample B, one as in Fig. 4, and the other as in Fig. 5. Part 2 of the experiment was similar, but in both A and B samples all spectral components below and above the selected band were removed, so that only energetic masking introduced by the band of noise (masker BN) contributed to the threshold of masking found. Figure 6 presents the band of frequencies from 1227 Hz to 2261 Hz and Fig. 7 presents the same band with the spectral “gap” on the left (corresponding to Fig. 4) and noise at -11 dB added.

Figure 8 shows an example of stimulus B in the frequency band from 450 Hz to 925 Hz, with spectral gap on the left and noise at -6 dB added.

In both part 1 and part 2 spectral gaps on the left and on the right were investigated and the masking in two bands of frequencies, shown in Fig. 2, was examined. Therefore eight different types of signals were generated as bases for sample B (Table I). Actual samples B were obtained by adding noise in a band. For each A/B comparison the only difference between samples A and B was that there was no spectral gap in A.

The level of noise added was calculated in reference to the level of the original saxophone sound in the respective band. RMS values were used. The level of noise was increased in steps. The details are presented in Table I.

All stimuli were generated before the test, saved as audio files, and during the test they were played back by the test operator.

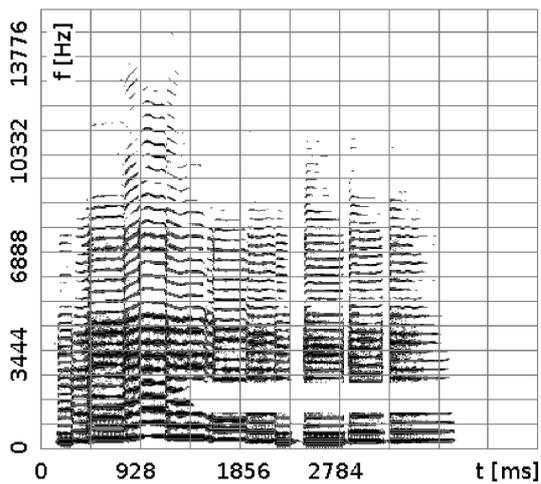


Fig. 5. The saxophone sound with a spectral “gap” — a band of frequencies (1227–2261 Hz) removed from it starting from some point in time.

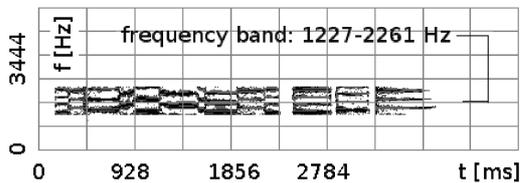


Fig. 6. The saxophone sound with all spectral components below and above the selected band removed (the signal).

2.2. Methodology

To find the threshold where the difference between samples A and B is not heard, i.e., to find the threshold of masking, a modified up-down procedure was used [5]. The level of the band of noise was changed depending on the previous answers. If the subject answered correctly, then the noise level was increased and if the answer was wrong then it was decreased in steps shown in Table I. The test was carried out by the operator who wrote down the responses of the subjects.

The actual comparison of A to B was performed according to the ABX procedure. Every ABX probe consisted of three stimuli: the standard, the first stimulus

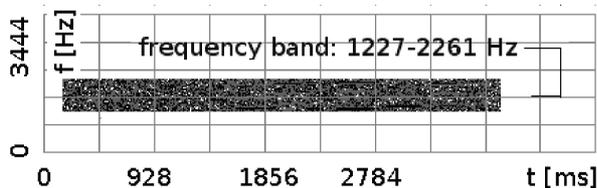


Fig. 7. The saxophone sound with all spectral components below and above the selected band removed, with the spectral “gap” on the left and noise added.

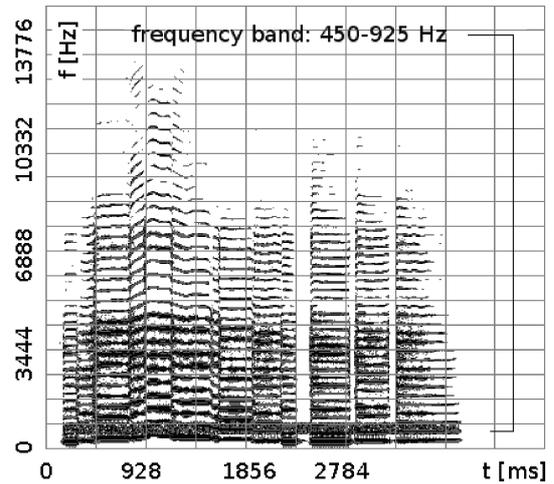


Fig. 8. The saxophone sound with a spectral “gap” (450–925 Hz) on the left and noise added.

and the second stimulus. The standard was selected randomly — it could be sample A as well as sample B. The first and the second stimuli were either A or B, presented in random order. The subject had to decide which of these two stimuli was different from the standard (Fig. 9).

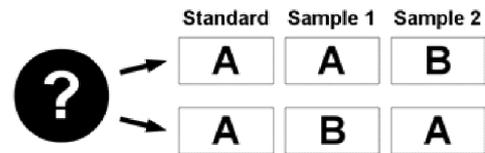


Fig. 9. The scheme of the modified ABX test.

The answer “I can’t decide” was allowed, but only in the case when the listener was sure that his/her choice would be random. The subjects could repeat listening to the ABX sequence for as many times as they wanted to.

2.3. The course of the experiment

The group of listeners consisted of 8 experts (2 women and 6 men) with experience in sound production. Listeners were in the age group of 20–50 and had normal hearing within 20 dB of hearing level measured at 6 standard octaves.

The stimuli were presented to the subjects monaurally. Listeners were free to choose the ear.

The stimuli were reproduced with the system consisting of a computer (MacBook), an audio interface (Lexicon Lambda) and headphones (Beyerdynamic DT 770 PRO). The connections in the system are shown in Fig. 10.

The initial level of the sound presented to the listener was *ca.* 80 dB (measured in a silenced room, at the distance of 5 cm from the headphone). In each series the sample with the lowest level of masking noise was presented first so that the subject could easily recognise the

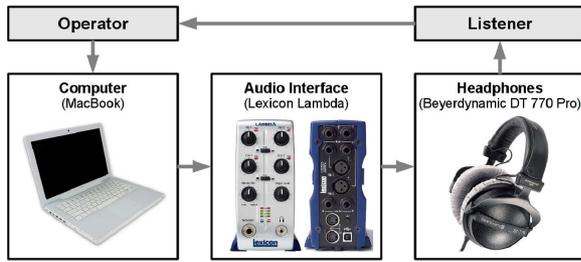


Fig. 10. Connections of elements in the listening system used in the experiments.

difference. Then the level of noise was increased step by step according to Table I. The level of the stimulus during the course of the test increased, due to the increase in the noise level, and the listeners could adjust the level until it was comfortable, while the test was going.

The subjects were communicating with the operator by pointing at one of the pieces of paper with a proper request or answer. They could ask for replaying the standard, the first stimulus, the second stimulus, all stimuli or give an answer. The subject could take a break at any moment of the test.

Stimuli used in the experiment and levels of noise masker for each type of the stimulus.

TABLE I

Part of experiment	Masker investigated	Stimulus for sample B	Gap location in time [s]	Frequency band of noise and gap [Hz]	Noise level range [dB]		Steps
					from	to	
1	CBN	full sax with left lower spectral gap plus band of noise	0–1392	450–925	–11	16	two subsequent 3 dB steps at both extremes of a range, 2 dB steps otherwise
1	CBN	full sax with right lower spectral gap plus band of noise	1392–end	450–925	–11	16	
1	CBN	full sax with left higher spectral gap plus band of noise	0–1392	1227–2261	–25	18	
1	CBN	full sax with right higher spectral gap plus band of noise	1392–end	1227–2261	–25	18	
2	BN	lower band with left part removed plus band of noise	0–1392	450–925	–3	27	
2	BN	lower band with right part removed plus band of noise	1392–end	450–925	–3	27	
2	BN	higher band with left part removed plus band of noise	0–1392	1227–2261	–6	23	
2	BN	higher band with right part removed plus band of noise	1392–end	1227–2261	–6	23	

3. Results and discussion

The results are presented in Table II. The main finding is that there is a considerable difference between the thresholds of masking by, respectively, the band of noise

plus other spectral components of the sound, referred to as masker CBN in Introduction (upper two rows of Table II) and the band of noise alone, referred to as masker BN (lower two rows of Table II). The difference was 14.5 dB for the higher band and 16.5 dB for the lower

band. This demonstrates that masking from the spectral context is more effective by 15.5 dB (averaged for the two bands) than the purely energetic masking by noise inside the masked band. The values in the two investigated bands were close (only 2 dB of difference), they were also close when results from “left gap” stimuli are compared to the results for “right gap” stimuli. The “left gap” results were consistently higher, but the differences were not meaningful, from 1 to 4 dB.

TABLE II

Results of the experiment.

Masker	Band [Hz]	Masking threshold left gap with std. dev.	Masking threshold right gap with std. dev.	Averaged
CBN	1227–2261	2 ± 3.6	-2 ± 5.2	0
CBN	450–925	8 ± 3.5	6 ± 3	7
BN	1227–2261	16 ± 2.4	13 ± 2.7	14.5
BN	450–925	24 ± 3	23 ± 1.8	23.5

The higher band represents roughly harmonics no. 4 and 5. As different notes were played during the 3.7 s of the sample, at some parts of it the higher band consisted also of harmonics no. 3, 6 and 7. The width of the higher band was set so that it always contained the 4th and the 5th harmonics. The lower band contained mainly the 2nd harmonic, but as different notes were played its frequency changed. The width of the lower band was set so that it always contained the full energy of the 2nd harmonic.

4. Conclusions

When a band of frequencies in a saxophone sound playing a couple of notes legato is masked by noise, the efficiency of masking drops by about 15 dB when the entire spectral context below and above that band is removed. This shows that the spectral context itself provides substantial masking, equivalent to 15 dB of noise in the same band. Two bands were investigated, both about an octave wide. With such a width of the band the energetic masking from the harmonics outside of that band is low, suggesting that most of the masking caused by the spectral context is non-energetic or informational.

Acknowledgments

This work was supported by the National Centre for Research and Development, grant no. R02 0030 06/2009.

References

- [1] P. Kleczkowski, *Selective Mixing of Sounds*, 119th Conv. Audio Eng. Soc., New York 2005, Preprint 6552.
- [2] I. Pollack, *J. Acoust. Soc. Am. Suppl. 1* **57**, S5 (1975).
- [3] D. Kidd, Jr., Informational Masking, in: *Auditory Perception of Sound Sources*, Eds. W. Yost, A. Popper, R. Fay, Springer Science+Business Media, LLC, New York 2008, p. 143.
- [4] C.S. Watson, *Acust. Acta Acust.* **91**, 502 (2005).
- [5] H. Levitt, *J. Acoust. Soc. Am.* **49**, 467 (1971).