Masking release based on comodulation of target and cue band

Angela Josupeit[1], Steven van de Par[2]

[1] CvO Universität Oldenburg, Institut für Physik, AG Akustik, Email: angela.josupeit@uni-oldenburg.de

[2] Email: steven.van.de.par@uni-oldenburg.de

Introduction

Comodulation of masker components can have opposite effects on detection performance: In a simple tone-in-noise task, by adding comodulated masker components outside of the critical band, the performance of tone detection is improved. This effect has been termed Comodulation Masking Release (CMR, [2]). Otherwise, if comodulated flanking bands are presented in a modulation detection task, the performance of modulation detection is degraded, an effect called Modulation Detection Interference (MDI, [7]).

CMR is often mentioned as one factor contributing to speech intelligibility in background noises. It has been suggested that, analogue to a single tone, speech signals can be detected better if the background signal is comodulated (e.g. [4]). In this specific context, it is usually not mentioned that also the speech signal itself is comodulated over a great frequency range due to envelope comodulation and due to common pitch. Possibly this comodulation also improves speech intelligibility. More specifically, it is the question whether this effect could hold for formant discrimination in the presence of a low-frequency noise masker. Assuming a CMR effect, high-frequency comodulated side bands would help to identify the modulation of formant-like sounds and therefore performance may be improved. However, based on reports on MDI, one might expect that the presence of the strong pitch-based modulation at high frequencies would diminish the ability to detect the modulations caused by the low-frequency formant bands. Therefore, in this experiment, the influence of high-frequency band comodulation on discriminability of narrow-band tone complexes which simulate speech formants is investigated.

Experiment

In this experiment, threshold levels were measured at which two bandpass filtered tone complexes with different center frequencies and with the same fundamental frequency $f_0 = 80$ Hz can be discriminated in the presence of a 3000-Hz lowpass noise masker. The time waveforms S(t) for the tone complexes were

$$S(t) = A \sum_{k=0}^{n} \sin\left(2\pi\left(f_l + k f_0\right) t\right), \qquad (1)$$

where A is the peak amplitude, $(n+1)$ is the number of sinusoids, here set to $(n+1) = 4$; f_l is the lowest frequency which is set to $f_l = 500$ Hz for the target interval and to $f_l = 700$ Hz for the reference intervals in the first experiment. The masker had a level of 60 dB SPL and a duration of 500 ms. Thresholds were estimated for tone

f_l	vowels	formant
500	ɜ, oe, ɛ	F1
700	æ, ɑ, aː, ʌ, ɒ	F1
1000	u, ɔ	F2
1200	ʊ	F2
1500	ʌ, ɜ, ae, a, œ	F2

Table 1: f_l values used in the experiments and formants of the associated vowels

complex durations of 30, 100, 200, and 500 ms. All signals were multiplied with a 15 ms raised cosine temporal window.

This basic setup formed the first experimental condition. In a second condition, a second tone complex was added to the signals, which had the same time waveform as described in 1, but setting $(n+1) = 60$, $f_l = 4000$ Hz and a level of 30 dB SPL per sinusoid component. In the following, this stimulus is labeled as "tone complex cue band". In a third condition, a 4000 Hz pure tone instead of a tone complex was added to the basic signals, below labeled as "single tone cue band". This tone had a level of 48 dB SPL. Both cue band types have the same starting phase as the target.

In a reference condition, the masked threshold for only the (200 ms) target tone complex was measured in the presence and absence of the tone complex cue band.

To further examine the effect of a comodulated cue band on formant discrimination, an additional experiment was implemented, indicating other formant regions. In both experiment conditions, the target tone complex had a lowest frequency of $f_l = 1000$ Hz. The lowest frequency of the reference tone complex was $f_l = 1200$ Hz in the first condition and $f_l = 1500$ Hz in the second condition. The other settings were the same as those in the first experiment.

The values of f_l of the target and reference tone complexes approximately represent either the first (F1) or second (F2) formants of certain vowels, see Tab. 1 (cf. [5, 1, 3]).

Thresholds were determined using a 3I-3AFC one-up two-down adaptive procedure, estimating the 70.7% correct point of the psychometric function. The initial step size of 8 dB was divided by 2 after every two reversals until a minimum step size of 1 dB was reached. The median of the final 8 reversal levels was taken as the estimate of the threshold for that run. The final threshold value was taken as the mean of 4 runs. Subjects received visual feedback indicating whether the response was correct.

Five normal-hearing subjects participated in the first experiment. Three of them also participated in the sec-

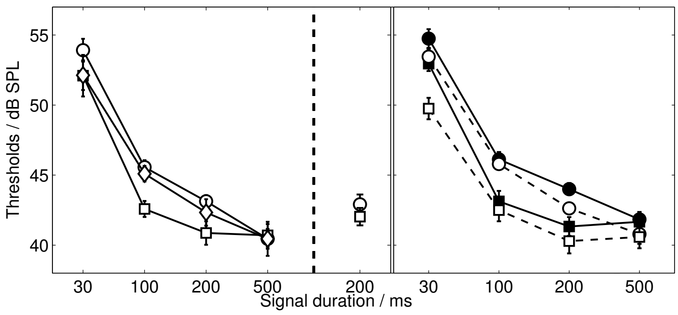

Figure 1: Discrimination and detection thresholds for lowest frequency discrimination of tone complexes in the presence of a 60 dB SPL 3000-Hz low pass masker as a function of signal duration (circles: no cue band; squares: tone complex cue band; diamonds: single tone cue band). Left: f_l = 500 Hz versus f_l = 700 Hz (left side of this panel: discrimination thresholds; right side: detection threshold for a tone complex with f_l = 500 Hz). Right: f_l = 1000 Hz versus f_l = 1200 Hz (closed symbols) and f_l = 1000 Hz versus f_l = 1500 Hz (open symbols)

ond experiment. All had previous experience in psychoacoustic tasks.

Results

The results of the first experiment are shown in the left panel of Fig. 1, which plots the threshold for lowest frequency discrimination as a function of time averaged over 5 subjects. In addition, thresholds for tone complex detection (f_l = 500 Hz and 200 ms duration) are plotted in the right side of the left panel. The results of the second experiment are shown in the right panel of Fig. 1, averaged over 3 subjects. The circles represent thresholds for conditions with no cue band presented. The squares and diamonds indicate thresholds with a tone complex cue band and a single tone cue band respectively.
In all conditions, discrimination thresholds decrease with increasing target duration, which is in line with earlier results on duration effects for detecting tones in noise (e.g. [6]).
In the first experiment, for target durations of 100 and 200 ms, thresholds decreased significantly ($p < 0.05$ in an ANOVA) by about 2-3 dB if the tone complex cue band is presented. In the second experiment, similar results were also found for a tone complex duration of 30 ms.
On average, there is no significant difference between thresholds obtained with a single tone cue band and those obtained with no cue band for target durations of 100, 200, and 500 ms.
The detection threshold is equal to the discrimination threshold at 200 ms with no tone complex cue band presented. Thus, it can be assumed that once a tone complex signal is detected it can also be discriminated with respect to its lowest frequency.

Discussion and Conclusion

For target durations of 200 ms or less, it is observed that the presence of a comodulated cue band can improve the ability to discriminate between the lowest frequencies of tone complexes. This effect is shown for several frequency regions that represent different formants.
A possible beneficial mechanism could be that the timing (i.e. start and end time) of the target signal is signaled by the cue band. This assumption is tested with the single tone cue band, which does not lead to a performance improvement.
Since thresholds decrease if a tone complex cue band is presented, it is suggested that some type of CMR mechanism plays a role. An MDI effect can be supposed to be irrelevant for the explanation of the results since an MDI effect would increase thresholds in the presence of the cue band.
The effects observed in the presented experiment can be relevant for a better understanding of the factors influencing speech intelligibility in noise. That is, the results indicate that high frequency modulation that is related to low frequency modulation could improve formant discriminability of speech signals. This finding is relevant for hearing impaired with high-frequency hearing loss. Although the effect found in this experiment is rather small, it may still be regarded significant in the context of speech intelligibility improvement.

References

[1] D. Deterding. The formants of monophthong vowels in standard southern british english pronunciation. J Int. Phonetic Assoc., 27:47–55, 1997.

[2] J. W. Hall, M. P. Haggard, and M. A. Fernandes. Detection in noise by spectro-temporal pattern analysis. J. Acoust. Soc. Am., 76(1):50–56, Jul 1984.

[3] A. Iivonen. Zur regionalen Variation der betonten Vokale im gehobenen deutsch. Neophilologica Fennica. Neuphilologischer Verein 100 Jahre. Mémoires de la Société Néophilologique de Helsinki XLV:87–119, 1987.

[4] B. J. Kwon. Comodulation masking release in consonant recognition. J. Acoust. Soc. Am., 112(2):634–641, Aug 2002.

[5] G.E. Peterson and H.L. Barney. Control methods used in a study of the vowels. J. Acoust. Soc. Am., 24:175–184, 1952.

[6] C. S. Watson and R. W. Gengel. Signal duration and signal frequency in relation to auditory sensitivity. J. Acoust. Soc. Am., 46(4):989–997, Oct 1969.

[7] W. A. Yost and S. Sheft. Across-critical-band processing of amplitude-modulated tones. J. Acoust. Soc. Am., 85(2): 848–857, Feb 1989.