Temporal Effects of Preceding Band-Pass and Band-Stop Noise on the Recognition of Voiced Stops

Teresa Cervera
University of Valencia, Avda. Blasco Ibanez 21, 46010 Valencia, Spain. Teresa.Cervera@uv.es

Julio Gonzalez-Alvarez
University Jaume I Castellon, Spain

Summary
A previous study showed that the scores on recognition of plosive-vowel syllables improve with noise preceding the speech, compared to noise gated on and off simultaneously with the signal [1]. To investigate this question further, the present study examined the influence of masker bandwidth, masker and signal spectral separation and masker level on the recognition scores of band-pass filtered speech stimuli for different durations of the preceding noise. When the signal onset was delayed 200 ms from the masker onset, its recognition improved, compared with gated noise. A smaller increase was observed in the performance for longer delays. The same amount of improvement in the recognition of the syllables (around 28%) was obtained for band-pass and band-stop preceding noise. The beneficial effects of these preceding maskers on the scores on recognition of the speech stimuli used in this study are similar to those obtained in the psychophysical studies on "overshoot" and "auditory enhancement" phenomena.

PACS no. 43.71.Gv

1. Introduction

Studies on speech perception have shown that background noise causes perceptual confusions [2, 3, 4, 5]. But when the masking noise, instead of being presented simultaneously, precedes the signal for some duration, it produces beneficial effects compared to the "gated" noise (noise gated on and off simultaneously with the signal). This result was found in previous studies [6, 7, 1].

AinsworthŠs study [7] used preceding noise (noise starting prior to the signal and continuing with it and offset simultaneously) durations of between 200 and 500 ms. Cervera and Ainsworth [1] used preceding noise of 100 to 800 ms in duration. The greatest benefit (obtained by comparing the recognition scores for preceding and gated noise) occurred for preceding noise of 200 ms. Longer delays did not continue to improve the recognition of the syllables.

The results found in these perceptual studies have some similarities to the "overshoot" effect found in psychophysical studies. [8, 9, 10, 11, 12]. The detection of a brief tonal signal improves as its onset is delayed (around 200 ms) relative to the onset of a longer masker (of around 400 ms). The overshoot effect has been studied with tonal [13, 12, 14, 15] and noise [8, 9, 10, 11] maskers.

On the other hand, the amount of overshoot is not only critically dependent on those components at the signal frequency, but it also depends on the off-frequency components [9, 16, 17, 18, 19, 20].

Another group of studies on the "enhancement phenomenon" have also studied the effects of the presentation of preceding maskers or "adaptors" [21, 22, 23, 24, 25]. These studies found that a target frequency region can be perceptually enhanced by previous exposure to a complex masker in which the component corresponding to the frequency of the target is omitted. The enhancement effect has been studied with non-speech stimuli [21, 22, 23, 24, 25] and speech-like stimuli [26, 27]. Viemeister and Bacon [25] interpreted this effect as an adaptation of suppression. That is, the suppressing effect on the stimuli by the precursor noise would adapt over time. However, Wright *et al.* [28] did not support for this hypothesis.

Whereas a considerable amount of research has studied the temporal effects of masking on the detectability of tonal signals, fewer studies have used spectrally complex and longer signals like speech in suprathreshold tasks. Among them, Summerfield *et al.* [26, 27] used speech-like vowels as stimuli and an identification task in their studies.

The present study, examined the phenomenon of "overshoot" on a suprathreshold level and with speech stimuli. This phenomenon was initially studied in a previous work [1]. The present study, investigates this question further in an experiment in which the influence of masker band-

Received 25 May 2006, revised 2 March 2007,
accepted 12 September 2007.

width, masker and signal spectral separation and masker level on the recognition scores of band-pass filtered speech stimuli were examined, for different durations of the preceding noise.

To achieve this objective, an identification task was used. This type of task has frequently been used in the speech perception field since the Miller and Nicely [2] study. The effects of preceding noise on speech recognition are worthy of consideration, because they can help to explain the robustness of speech perception in noise.

We performed different comparisons of the spectral characteristics of the noise maskers. First, we tested whether preceding band-pass and band-stop noise produce an advantage in identifying speech stimuli, similar to what has bee found for signal detection. We expect increases in the recognition scores for both band-pass and band-stop preceding noise, in light of our previous findings and the results from other previous psychophysical studies on the overshoot effect. In these studies, overshoot was found, not only for maskers whose components were at the signal frequency, but also for maskers whose components were remote from the signal frequency [9, 16, 17]. In addition, the studies on the enhancement phenomenon used band-stop precursors ([24] and [25, 26, 27]). Examining this question is of special interest because continuous background noise of different spectral characteristics (either at frequencies where the speech occurs or at other parts of the frequency spectrum) is commonly encountered in daily communication situations. Both types of noise could be equally responsible for the effects.

We also varied other characteristics of the band- pass and band-stop preceding noise, and the effects on the recognition scores were examined. The bandwidth of the band-pass noise was varied with the purpose of evaluating its effect. In several studies on overshoot [10, 8, 17, 18, 19] the effects of masker bandwidth on signal detectability have been examined. For maskers centered on the signal frequency, they found an increase in the overshoot effect as the masker bandwidth increased. However, in the present study the overall level of the noise was kept constant across the different bandwidths (meanwhile in the overshoot studies what was constant was the spectrum level). Thus, we expected different results.

Some aspects of the band-stop noise were studied. An examination was carried out of whether it would be necessary for the surrounding bands of the precursor noise to be adjacent to the test signal or not, in order to produce the improvement on the recognition scores. It was expected, following the Viemeister and Bacon [25] hypothesis, that the enhancement effect would be due to adaptation to suppression, and that non- adjacent bands would produce less enhancement and, therefore, less improvement on the recognition of the stimulus than adjacent bands. It was also expected that narrower bands of noise will produce less enhancement than complementary bands (that is, band-stop noise with the signal in the band-pass).

Furthermore, the question of whether different SNRs can produce differential effects on the recognition scores was examined. Thus, a higher SNR of 12 dB SNR was employed and compared to the 6 dB SNR condition.

The enhancement effect has also been observed with different masker intensities. But the signal-to noise relationships between the precursor and the signal have not been studied. According to the increased-gain hypothesis [25], the enhancement effect would occur not only because the exposure to precursor noise reduces its effective level and, therefore, its ability to mask, but also because the newly arriving signal is less susceptible to suppression, due to adaptation to suppression. These authors suggest that the enhancement effect not only reduces the effective auditory level of the pre-existing energy, but that the enhanced component also behaves as if it were increased in intensity. According to this hypothesis, the effective level in the not previously stimulated frequency region increases. If the SNR were increased or decreased, the effective level of the signal would be increased or decreased by the same proportion, and the signal detectability would not significantly change.

Finally, as three different syllables were used as speech stimuli on a perceptual task, it was considered necessary to assess the speech recognition behaviors of the listeners in more detail, by evaluating their perceptions of individual plosives across the different noise conditions. To achieve this, the data were arranged in confusion matrices, and the results were interpreted in relation to the acoustical characteristics of the individual stimuli. The confusions among consonants, using masking noise, were studied in the classic Miller and Nicely study [2] with the English consonants. In that study, the confusions among voiced stops showed that the /d/ and /g/ pair was more confused than /b/ with the other two voiced plosives. In the present study, it was expected that the pattern of confusions obtained in the recognition task, with the Spanish voiced stops, would not differ substantially from that previous study.

2. Method

2.1. Stimuli

Voiced-stop syllables were used in the present experiment, as well as in our previous study [1]. These consonants were chosen because they have common manner of articulation and voicing features, differing only in the place of articulation feature, which is a frequently confused feature [2]. In the present study, an attempt was made to limit the frequency region of the speech stimuli as much as possible, so that we could manipulate the effects of the presentation of the preceding noise either in the same frequency band as speech or in the “complementary” (where there was no speech) frequency regions. To accomplish this, a pilot study established that the voiced plosives combined with /a/ and band-pass filtered at 920 to 2000 Hz produced 100% correct identification in quiet.

The voiced-stop syllables were spoken by a native male speaker. They were recorded in a soundproof room using a Sennheiser HMD 224 microphone set at 25 cm from the lips and directly digitalized in the computer with a sampling rate of 11.025 KHz. The three syllables were made equal in duration. To isolate the stop-vowel syllable, 200 ms of the speech signal, beginning at the murmur preceding the plosive burst, were selected. The total 200 ms of duration of each stimulus included a cosine-squared onset/offset ramps of 10 ms. The amplitudes of the three stimuli were also adjusted to make them of equal amplitudes across the entire syllable.

The speech stimuli were then band-pass filtered at 920–2000 Hz. A 200th order FIR filter was used. This frequency band (920–2000 Hz) comprises 6 Bark (the edges of the band corresponded to the edges of the 9th to 14th Bark). The spectral characteristics of each plosive are described in section IV.

Maskers consisted of white noise (low-pass 5.5 kHz at a sampling frequency of 11.025 KHz) passed through a 200th order linear phase FIR filter, using the MATLAB *fir* function [29] with a Hanning window, to produce the different band-pass and band-stop filters, by varying the cut-off frequencies used in the different experimental conditions of this study. They were a) band-pass noise of different bandwidths. b) band-stop noise. c) maskers consisting of two flanking bands, were created by the combination of two band-pass filters.

In all the conditions, the noise began at 800 ms, 400 ms, 200 ms and 0 ms (gated noise) before the onset of the syllable and continued until the end of the signal (200 ms later). The onset and end of the noise were also subjected to a Hanning window. The different filtered noises were added to the signal to make the SNR during the syllable equal to 6 dB SPL (except in one of the conditions where a higher SNR of 12 dB was employed). The overall level of the noise was the same in all the conditions, allowing the spectrum level to vary in each condition.

The manipulations of both the speech signal and the maskers were performed using MATLAB [29] routines. All the stimuli (speech and noise) in the different experimental conditions were presented at 70 dB SPL

2.2. Listeners and procedure

A group of 11 listeners participated voluntarily in the experiment. All of them were students at the University of Valencia and native Spanish speakers. Their ages ranged from 22 to 31 years, and all of them had pure-tone air conduction thresholds of less than 20 dB SPL from 125 to 6000 Hz [30].

The listeners performed the experiment individually in a sound attenuated room. The stimuli were presented diotically through Sennheiser HD 435 headphones at the level of 70 dB SPL and controlled by a computer. Each time a stimulus was presented, the listener was instructed to press "B", "D" or "G" on the keyboard of the computer. After the key was pressed, the next stimulus was presented 2 s later.

The stimuli were presented in random order. The sessions were divided into blocks of 60 stimuli (3 syllables × 4 durations of the noise × 5 times). There were eight blocks corresponding to each condition of filtering characteristics of the noise. A practice block was presented at the beginning of the first session. In each session, different blocks were presented. The listeners performed the different blocks in different sessions in random order over a few weeks. Feedback was provided during the practice, but not during the experiment. Each condition was scored by counting the number of correct identifications. The differences between the percent scores obtained in gated and continuous 800 ms maskers provided a measure of the beneficial effects of the preceding noise.

2.3. Data analysis

The percent recognition scores, obtained in the perceptual task in each condition of filtering characteristic of the preceding noise (or "type" of noise) and duration of the noise by each listener were analyzed using a two-way intra-subjects ANOVA [31]. In this analysis, the main effects for both type of noise and duration of the noise were tested. Afterwards, in accordance with the objectives of the study, some comparisons among different filtering characteristics of the noise were examined separately by using the Bonferroni post-hoc test [32]. The differences among the levels of the duration of the noise were also tested.

The data were also arranged in confusion matrices to examine the perceptual errors. A confusion matrix was obtained for each type of masker and two of the durations of the masker (0 and 800 ms) . In the confusion matrices the first vertical line represents the stimulus, and the first horizontal line represents the response. The number of each cell is the frequency with which each stimulus-response pair was observed. Thus, the values in the diagonal represent the correct scores, and the off-diagonal values represent the confusions among the plosives.

The question arising from the confusion matrices was whether there were certain patterns of confusions (the errors were not scattered randomly), and, if so, whether these patterns of errors were influenced both by the duration of the noise and by the filtering characteristics of the noise.

To accomplish this objective, a log-linear analysis was conducted. The aim of this method is to explain the cell frequencies of the confusion matrix tables with the minimum number of terms, and then eliminate one term in each round through a process of backward elimination (backward hierarchical method). Each time a term was removed, the assessment of goodness-of-fit was calculated by means of a likelihood ratio chi-squared approximation, the G^2 statistic (or multivariate chi-square) [33]. This method was used for the analysis of consonant confusion matrices in a previous study by Bell *et al.* [34]. In this analysis, the diagonal values (correct scores) were not considered.

A four-dimensional matrix (S×R×D×N) representing stimuli [S], response [R], duration of the noise [D] and type of noise [N] was used in each of the comparisons of the filtering characteristics of the noise (type of noise) used in the present experiment. The independent model was [N] [D] [S] [R], and the saturated model was the four-way interaction [NDSR], which represents the combined effects of type of noise and noise duration on the stimuli-response patterns. Among the two-way associations [ND] [NS] [NR] [DS] [DR] [SR], the [SR] association was of particular interest, as it represents the stimuli-response association or error patterns. Among the three-way interactions [NDS] [NDR] [DSR] [NSR], the [DSR] and the [NSR] associations were especially interesting, as they reflect the effects of the duration of the noise on the stimuli-response patterns and the effects of type of noise on the stimuli-response patterns, respectively.

The analysis was begun with the higher association or saturated model. Then the all three-way associations model was tested against the saturated model. If the G^2 value was non-significant, this model fitted the data (it did not differ from the saturated one, which represents the perfect fit) and, therefore, was accepted. Then, the all two-way association model was tested, thus eliminating as many interactions as possible, while maintaining an adequate fit between expected and observed frequencies.

3. Results. ANOVA

The scores were analyzed in a two-way ANOVA. Eight levels of type of the preceding noise (band-pass, band-stop, narrow band-pass, wider band-pass, non-adjacent bands, narrow adjacent bands, narrower adjacent bands, band-stop at 12 SNR) were considered in the analysis, and four levels (0 ms or coincident, 200 ms, 400 ms and 800 ms) of the duration on the preceding noise. The main effects for both type of noise and duration of the noise were significant (F = 15.26, df = 7, $p < 0.01$, and F = 20.12, df = 3, $p < 0.01$ respectively). The interaction was not significant. Differences among the levels of duration of the noise, by means of the Bonferroni test showed significant differences between 0 and 200 ms ($p < 0.01$), between 0 and 400 ms ($p < 0.01$), and between 0 and 800 ms ($p < 0.01$). The differences in the recognition scores among the levels of type of noise that were of interest in the present study were examined in more detail separately, as well as the perceptual errors.

3.1. Comparison of the band-pass and band-stop preceding noise

Figure 1, a and b, shows the means and standard deviations of the percent recognition scores calculated across all subjects for both band-pass and band-stop (920–2000 Hz) preceding noise, at the different durations of the noise. The trajectories of the two lines representing the recognition scores at the different noise durations show that the recognition increases as the precursor noise duration increases. The improvement on the intelligibility of the syllables (the difference between 800 ms and coincident noise conditions) was the same in both conditions (around 27.87%). The Bonferroni post-hoc test examining the differences between the scores in band-pass and band-stop conditions was significant ($p < 0.01$).

The percent scores were arranged in confusion matrices, shown in Table I, a and b. These data were analyzed by means of a likelihood ratio chi-squared approximation, or G^2 statistic, following the method described in the previous section. The data matrix consisted of a four-dimensional 3×3×2×2 matrix representing stimuli, response, duration and type of noise. The all two-way association model fitted the data (G^2 (df=18) = 13.73, $p > 0.01$). Therefore, it was selected over the higher-order association models. Among the two- way associations, it was especially important to test the [SR] term. The elimination of the [SR] term from the model did not produce a good fit. The conclusion can be drawn that, first, the pattern of confusions was systematic (not randomly distributed among the cells). Second, although the absolute number of correct scores was affected by the duration and type of noise (as the ANOVA and post-hoc tests showed), the patterns of errors were not affected by them. An inspection of the residuals and the confusion matrices showed that these patterns corresponded to the confusions between /da/ and /ga/. /ba/ was less confused, as expected.

3.2. Bandwidth effects of the band-pass preceding noise

The bandwidth effects of preceding band-pass noise centered on the speech signal on the recognition scores were examined. It should be noted that, in this experiment, the overall level of the masker was kept constant. Thus, as bandwidth is increased, the spectrum level is decreased. Three different bands of noise of different bandwidth were used: band-pass 920–2000 Hz, a narrower band-pass noise (1080–1720 Hz), that is, a band of noise of 1 Bark less above and below the band-pass signal; and a wider band of noise at 1 Bark above and below (770–2320 Hz) the band-pass signal.

The percent recognition scores for the three conditions of the bandwidth of the preceding noise are shown in Figure 1a, c, and d, across the different durations of the noise. The trajectories of the three lines representing the recognition scores throughout the duration of the preceding noise show that the recognition increases as the precursor noise duration increases The amount of increment after 800 ms of the precursor noise for the coincident band condition was, as mentioned in the preceding section, 27.87%. For the narrower band, the increment was 22.42%, and for the wider band, it was 21.21%.

Post-hoc contrast by means of the Bonferroni test showed significant differences between band-pass and the narrower band-pass ($p < 0.01$), and between the narrower band-pass and the wider band-pass ($p < 0.01$). There were non-significant differences between band-pass and wider band-pass conditions.

The data in the confusion matrices presented in Table I, a, c and d, were analyzed using the hierarchical log lineal approach with G^2 statistics. The data matrix consisted of a four-dimensional 3×3×2×3 matrix representing stimuli, response, duration and type of noise. The all two-way

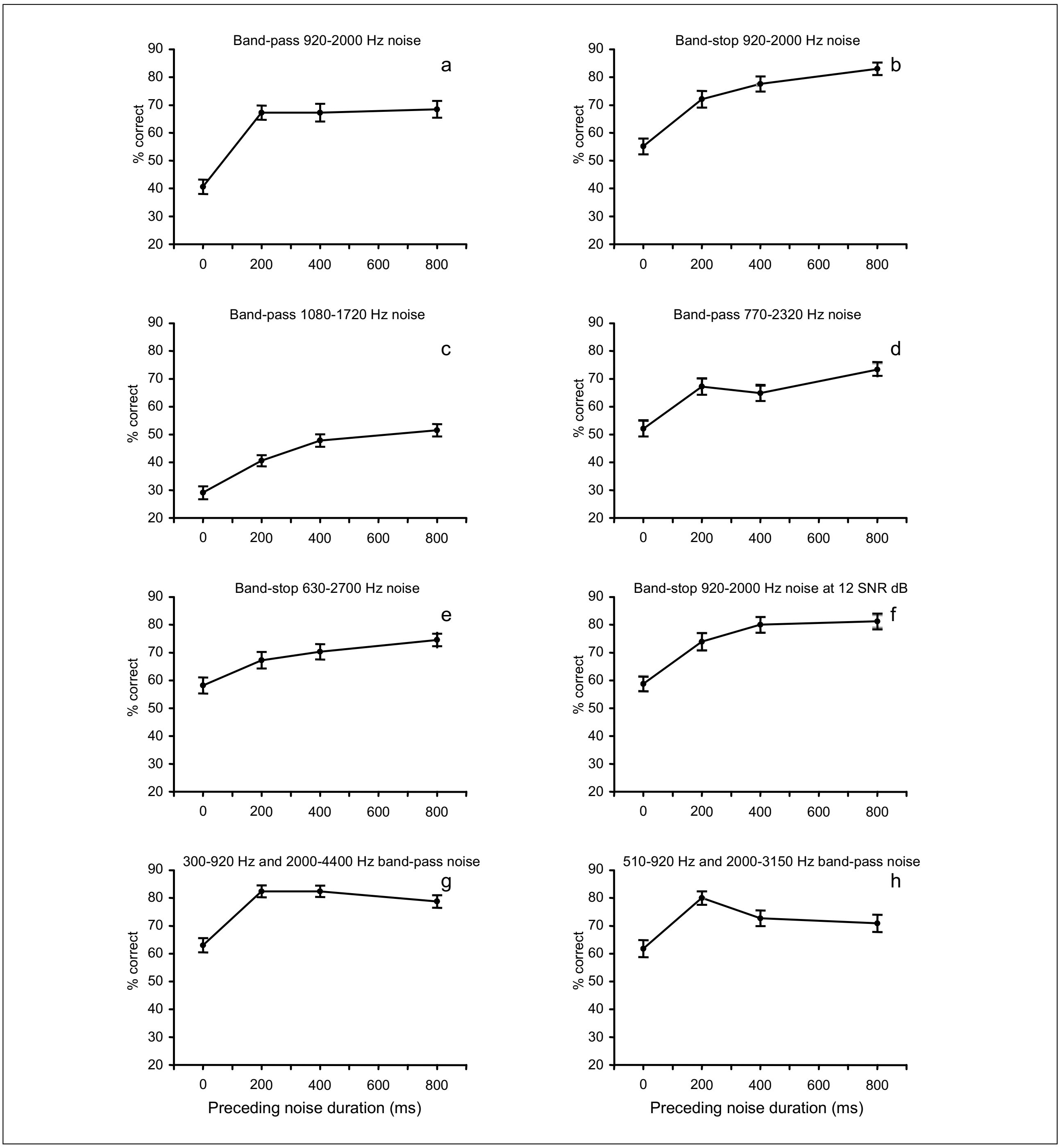

Figure 1. Recognition scores (% plosives correct) and their standard deviations as a function of the duration of the preceding noise for a) band-pass 920–2000 Hz noise, b) band-stop 920–2000 Hz noise, c) band-pass 1080–1720 Hz noise (narrower band-pass), d) band-pass 770–2320 Hz noise (wider band-pass), e) band-stop 630–2700 Hz noise (non adjacent), f) band-stop 920–2000 Hz noise at 12 SNR, g) 300-920 Hz and 2000–4400 Hz band-pass noise (two narrow adjacent bands), and h) 510–920 Hz and 2000–3150 Hz noise (two narrower adjacent bands).

association model fit the data well and, therefore, was accepted [$G^2 = 23.75$ (df=30) ($p > 0.05$)]. The elimination of the [SR] term did not produce a good fit. The conclusion drawn is that, although the total number of correct scores varied with the type of the noise, the patterns of errors did not. An inspection of the residuals and the confusion matrices in Table I a, c and d showed that /da/ and /ga/ were more confused than /ba/ was with them.

3.3. Effects of adjacent and non-adjacent bands of preceding noise surrounding speech

Two different types of noise were examined. In the “adjacent bands” condition (band-stop 920–2000 Hz), the precursor noise was presented in the part of the frequency spectrum complementary to the signal, that is, the part where the signal had no energy. In the “non-adjacent bands” condition (band-stop 630–2700 Hz), the precursor

Table I. Confusion matrices of /ba/ /da/ and /ga/ in 0 ms and 800 ms noise in different filtered noise conditions. The first column represents the stimuli and the first row represents the response.

	0 ms			800 ms		
	/ba/	/da/	/ga/	/ba/	/da/	/ga/
a. Band-pass 920–2000 Hz						
/ba/	54.5	30.9	14.5	78.2	3.6	18.2
/da/	5.4	41.8	52.7	7.3	50.9	41.8
/ga/	23.6	50.9	25.4	1.8	21.8	76.4
b. Band-stop 920–2000 Hz						
/ba/	78.2	14.5	7.3	94.5	3.6	1.8
/da/	9.1	34.5	56.4	3.6	61.8	34.5
/ga/	5.4	41.8	52.7	1.8	5.4	92.7
c. Band-pass 1080–1720 Hz						
/ba/	27.3	38.2	34.2	60	10.9	29.1
/da/	20	21.8	58.2	21.8	38.2	40
/ga/	20	41.8	38.2	18.2	25.4	56.4
d. Band-pass 770–2320 Hz						
/ba/	76.4	14.5	9.1	92.7	3.6	3.6
/da/	1.8	49.1	50.9	0	61.8	38.2
/ga/	7.3	61.8	30.9	1.8	32.7	65.4
e. Band-stop 630–2700 Hz						
/ba/	76.4	16.4	7.3	80	12.7	7.3
/da/	7.3	45.4	47.3	0	63.6	36.4
/ga/	10.9	34.5	54.5	1.8	18.2	80
f. Band-stop 920–2000 Hz at 12 SNR						
/ba/	60	23.6	18.2	90.9	5.4	3.6
/da/	3.6	56.4	38.2	3.6	54.5	41.8
/ga/	5.4	32.7	60	0	1.8	98.2
g. 300–920 Hz and 2000–4400 Hz band-pass noise						
/ba/	89.1	10.9	0	94.5	5.4	1.8
/da/	3.6	47.3	49.1	0	56.4	47.3
/ga/	7.3	40	52.7	1.8	14.5	83.6
h. 510–920 Hz and 2000–3150 Hz band-pass noise						
/ba/	81.8	10.9	7.2	83.6	9.1	7.3
/da/	1.8	43.6	54.5	5.4	45.4	49.1
/ga/	3.6	36.4	60	1.8	14.4	83.6

bands of noise were set at two Bark from the edges of the speech band (920–2000 Hz).

Figure 1 b and e, shows the recognition scores for the two different types of noise, as a function of the duration of the noise. It can be seen that the scores increased as the duration of the noise increased The amount of the improvement on the recognition scores after 800 ms of preceding noise was 27.88% for the adjacent bands (band-stop noise) and 16% for the non-adjacent bands conditions. The Bonferroni post-hoc test examining the differences between the scores in adjacent and non-adjacent bands conditions was not significant.

The data obtained from the perceptual task were arranged in confusion matrices in Table I b and e and submitted to a log-linear analysis. A four-dimensional 3×3×2×2 matrix, representing stimuli, response, duration and type of noise, was used to find the best unsaturated model. The all two-way association model fitted the data [$G^2 = 10.78$ (df=20), $p > 0.05$]. The elimination of the [SR] term did not produce a good fit to the data. It was concluded that, although the number of correct scores was significantly different across the different duration of the noise, the pattern of errors was not. At the same time, the two types of noise produced similar pattern of errors. This pattern corresponded to the confusions between /da/ and /ga/, as in the previous experimental conditions.

3.4. Effects of adjacent bands of different bandwidths

Surrounding bands of preceding noise adjacent to the speech band were examined using three different bandwidths. One of them was band-stop noise 920–2000 Hz. This corresponds to approximately 7 Bark above and below the band-pass signal. The other precursors consisted of two bands of noise at 5 Bark above and below the signal (300–920 Hz and 2000–4400 Hz) (“narrow adjacent bands”), and at 3 Bark above and below the signal (510–920 Hz and 2000–3150 Hz) (“narrower adjacent bands”).

Figure 1 b, g and h, show the percent recognition scores of these three conditions of type of noise a function of the duration of the noise. It can be seen that the identification scores increase with the duration of the three types of noise. The amount of improvement of the recognition scores was 27.88% for the band-stop, and 15.75% and 9.1% for the narrower bands 5 Bark and 3 Bark above and below the speech band, respectively. The post-hoc comparisons showed significant differences between band-stop and narrower adjacent bands ($p < 0.01$), but not between band-stop and narrow adjacent bands, or between narrow adjacent bands and narrower adjacent bands.

The confusion matrices obtained corresponding to these type of noise are shown in Table I, b, g and h. The patterns of errors obtained in these matrices were examined across type of noise and duration of the noise, using log-linear analysis. A four-dimensional 3×3×2×3 matrix, representing stimuli, response, duration of the noise and filtering noise, was used. The all two-way association model fitted the data [$G^2 = 4.04$ (df= 28) ($p > 0.05$)]. The elimination of the [SR] did not fit the data. The conclusion follows that the pattern of errors was not significantly different in the two durations of the noise or in the three types of noise. As in the previous experimental conditions, it corresponded to the confusions between /da/ and /ga/

3.5. Effects of the SNR of the band-stop masker

Two different SNRs of the band-stop (920-2000 Hz) precursor were compared: 6 and 12 dB SNRs. Figure 1 b and f, show the recognition scores of the two conditions, as a function of the duration of the precursor noise. The post- hoc comparisons by means of the Bonferroni test between band-stop noise at 6 SNR and 12 SNR showed non-significant differences.

The data in the confusion matrices in Table I b and f, were submitted to a log-linear analysis. A matrix of 3×3×2×2, representing stimuli, response, duration and filtering characteristics of the noise, was used. The all three-way association model fitted the data [G^2 (df=4) = 6.13, (p > 0.05)]. Therefore, the interaction of duration and filtering characteristics of the noise on the pattern of stimuli-response can be deleted from the model. Then the two-way association model was tested against the three-way association model, and it did not fit the data. Therefore, the all three-way association model could not be rejected. Finally, the [DER] term was tested, which represents the effects of duration on the pattern of errors, and the term [NER], which represents the effects of filtering noise on the pattern of errors, successively, against the all three-way association model. In the first case, a G^2 was obtained that did not fit the data. In the second case, the model fitted the data [G^2 (df=4) = 4.23 (p > 0.05)]. The conclusion was that the pattern of errors was influenced by the duration of the noise, but not by the type of noise. An inspection of the residuals and the confusion matrices shows that at 0 ms or gated noise, the most frequent errors corresponded to the confusions between /da/ and /ga/, but at 800 ms noise, only the confusions of /ga/ with /da/ are maintained, but not those of /da/ with /ga/.

4. Acoustic analysis of individual plosives

The pattern of errors shown by the confusion matrices needed to be interpreted in relation to the acoustical characteristics of the different plosives. As the three speech stimuli (/ba/, /da/, and /ga/) differ in the place of articulation feature, this cue would have to be used by the listeners for correct identification.

As has been shown, both the spectrum of the burst [35] and the characteristics of the formant transitions [36, 37] are important cues to the perception of plosives. Thus, linear prediction spectra were computed at the burst of each plosive (approximately 10-ms from the onset) and approximately 45 ms later (approximately at or before the release of voiced plosives). Each plosive burst was analyzed by computing a 4th order LPC spectrum with a 10 ms window. LPC spectra were also computed 45 ms later. From the two spectra, the time course of the formant transitions can be estimated. These spectra are shown in Figure 2 as solid and dashed lines, respectively.

It can be seen that the burst of /ba/ has a peak at about 1100 Hz and at 1750 Hz, /da/ has a peak at 1400 Hz and at 1750 Hz, and /ga/ shows a main peak at 1450 and another peak at about 1850 Hz. The structure of the spectra obtained 45 ms later shows that /ba/ has a main peak at about 1080 Hz and at 1800 Hz, whereas /da/ and /ga/ have peaks at more similar frequency regions; that is, /da/ has a peak at about 1410 and at 1700 Hz, and for /ga/ there is a main peak at about 1350 Hz and at 1750 Hz. Thus, the second transition of /ba/ rises (from 1750 Hz to 1800 Hz), but it falls for /da/ (from 1700 Hz to 1700 Hz) and /ga/ (from 1850 Hz to 1750 Hz). The direction of the transitions of the plosives was as expected [36].

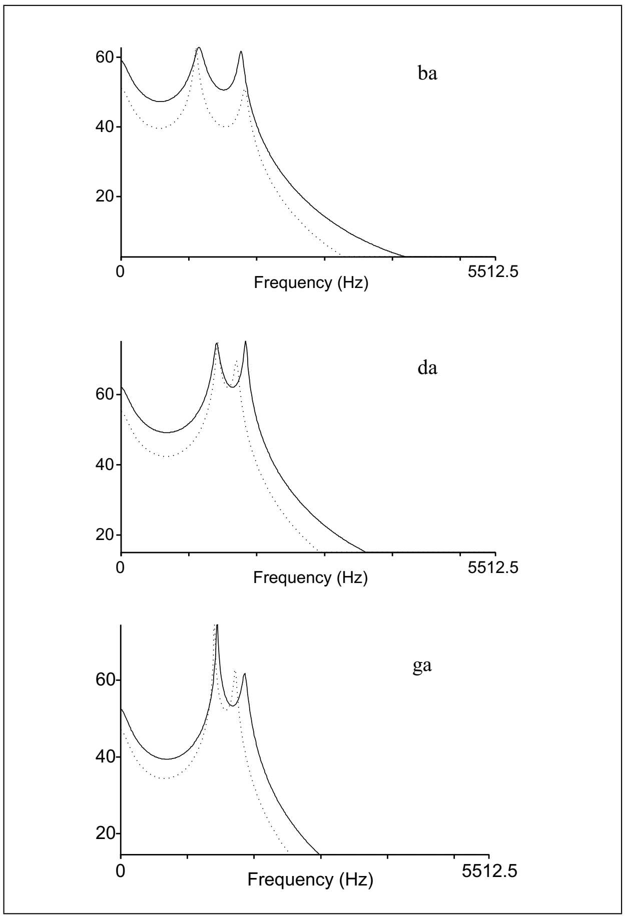

Figure 2. Spectrum of band-pass filtered /ba/ (top), /da/ (middle) and /ga/ (bottom) at the plosive burst (solid line) and 45 ms later (dashed line).

5. Discussion

The most important finding in the present study was that, when the noise preceded the syllable with a duration of 200 ms or more, the recognition of the syllables improved, compared to the recognition in gated noise. This result was consistent throughout the different comparisons of the filtering characteristics of the noise, in this study. The results from the present study are not directly comparable with the previous data in the overshoot experiments. In these previous studies, the improvement of signal threshold detectability (in dB) was assessed as a measure of the effect. In the present study, the scores on recognition of consonants in an identification task was measured. Both the task and the stimuli are different. However, some similarities between the psychophysical studies and the present study have been found.

Gated or simultaneous noise, either presented in the same frequency region as speech or in remote frequencies, produced negative effects on the recognition of the syllables employed in this study, as expected. The most deleterious effect was found for band-pass noises overlapping the speech signal. On the other hand, flanking bands of noise, either adjacent or non adjacent to the speech band, caused the least amount of masking.

5.1. Effects of band-pass noise

Among the three different band-pass noises used in the study, the narrower noise produced more masking than band-pass noise spectrally coincidental with the speech band, and band-pass noise wider than the speech band. It seems that, as the masker bandwidth increases, the masking effect decreases. This happens by keeping the overall level of the masker constant, but allowing the spectrum level to vary.

On the other hand, the increases in the recognition scores through the increasing duration of the preceding noise showed similar time course in the three maskers.

These results do not agree with the findings of the psychophysical studies on the overshoot effect. In those studies in which the bandwidth was manipulated [10, 17, 18, 8, 19], masker bandwidth increases produced increases in the overshoot effect. This occurred because masker bandwidth increases involved increases in the overall level. This confound was pointed out by Wright [19], who used constant spectrum levels for band-pass maskers and constant overall levels for band-stop maskers. The choice to keep the overall level constant across the different band- pass noises, used in the present study, was made in order to be consistent throughout the different experimental conditions (except when the SNR was manipulated). The SNR in the speech band (920–2000 Hz) varied when the bandwidth of the band-pass noise varied, but not in the other experimental conditions, in which the noise does not overlap the speech band.

Thus, it is possible that the narrower band, which has a greater density level than the wider bands, may more easily obscure those perceptual cues contained in the 1080–1720 Hz band that might be relevant to the perception of the syllables.

However, the results in the present study are not directly comparable with those of the overshoot studies because the spectrum level varied with the variations in bandwidth, and also because of the complexity of the speech stimuli extended over a broader band (compared with tonal signals employed in the psychophysical studies).

5.2. Effects of band-stop noise

Preceding band-stop noise, presented at frequencies complementary to the speech (band-stop 920–2000), produces important beneficial effects (around 28% improvement) on the recognition scores (compared to gated noise), similar to 920–2000 band-pass noise. These results were obtained for maskers presented at 6 dB SNR. For a band-stop masker presented at a lower level (12 dB SNR), 22.43% improvement was found, which was not significantly different from the noise at 6 dB SNR. This result suggests that for these favorable SNRs, the SNR does not seem to be very important, although additional experiments including other SNR conditions will be needed in order to reach a conclusion.

Preceding narrow adjacent bands, presented outside the speech band, also produced beneficial effects even with bandwidth changes, with between 9% and 15% improvement (narrower and narrow adjacent bands, respectively), as well as non-adjacent bands (with around 16% improvement).

These results agrees with the findings in the classical overshoot studies. When the maskers were presented at frequencies not overlapping the signal frequency they also produced a decrease in the signal detection threshold [8, 9]. Wright [19] used maskers that did or did not spectrally overlap the signal. Carlyon [38] showed that "notched noise" produced a "release of masking" on tonal signals of different frequencies. Carlyon and White [18] used maskers close to and "remote" from the signal frequency. Bacon and Savel [20] found beneficial temporal effects using "off-frequency" narrow-band maskers. And the studies by Viemeister [24], Viemeister and Bacon [25], and Summerfield *et al.* [27] also used band-stop maskers to produce the enhancement effect.

5.3. Comparison of band-pass and band-stop preceding noise

The overshoot effects obtained with maskers overlapping the signal have been traditionally interpreted as supporting the adaptation hypotheses, [39, 14, 38]. The similarity with the physiological studies by Evans [40] and Gibson *et al.* [41] suggests this possibility. The results obtained in the present study with speech stimuli and band-pass maskers would be consistent with this interpretation as well.

On the other hand, the effects of the maskers not overlapping the signal, more specifically, presented at complementary speech frequency regions relative to the frequency signal, in the studies on enhancement phenomena [21, 22, 23, 24, 25, 27] have been interpreted as an adaptation of suppression [25].

The results obtained in the present study using band-stop maskers could be related to the same mechanisms underlying the auditory enhancement phenomenon. That is, the subsequent presentation of the speech signal in that part of the frequency spectrum where there was no previous energy could produce an enhancement of it.

The temporal effects of masking noise on the recognition of plosive-vowel syllables, shown in this study using different band-pass and band-stop maskers, suggest that more than one mechanism may be involved. Both effects can be observed under a variety of circumstances. The effects of band-pass continuous noise can be observed, even if the noise does not completely coincide in the same frequency region as speech. The effects of off-frequency continuous noise can also be observed under different circumstances.

The time course of masking seems to be similar for all the conditions used in this study. Most of the increments in the intelligibility of the syllables, produced by the different types of precursor noise, were produced in the first 200 ms, and then the increments were smaller and tended to stabilize. This result is congruent with findings obtained in a previous study [1] and with the results obtained in the psychophysical studies. In the classical overshoot experiments, typically maskers of 400 ms were used, and the greatest effects were found when the signals were delayed 200 ms from the onset of the masker. Zwicker [9] found a " steady-state condition" at the 200 ms delay of the signal, because longer durations of the masker did not further decrease the thresholds of signal detectability. Carlyon [38] and Carlyon and White [18] obtained maximum "release from masking" at about 300 ms. Bacon and Healy [42] and Bacon and Liu [43], using "precursor" maskers from 50 to 400 ms, obtained maximum effectiveness when the precursor had a duration of about 200 ms.

In the enhancement studies, the same time course was found, although the point at which the effect was maximal can vary, depending on the study, from around 50 to 500 ms [24, 26, 27].

On the other hand, some physiological studies that examine the role of the efferent system in the phenomenon of overshoot suggest that its activation is relatively slow and requires about 200 ms to reach its maximum effectiveness [44, 45]. This time duration agrees, in general, with the time duration found in the psychophysical studies, as well as in the present study.

5.4. Analysis of the confusions among consonants

The patterns of errors observed in the confusion matrices were consistent across both the duration of the noise and the filtering characteristics of the noise (an exception must be noted in the case of band-stop noise, which presented different patterns of confusions between the 0 and 800 ms conditions). The most frequent errors were the confusions of /da/ with /ga/ and /ga/ with /da/, with /ba/ being more easily distinguished than the other two. The acoustical analysis of the syllables provides the basis for an interpretation of the data. Both the burst spectrum and the spectrum at the transition of both /da/ and /ga/ showed spectral peaks in closer proximity than those of /ba/, which shows a first peak (both the one corresponding to the burst and the one obtained at the transition) in a lower frequency region than /da/ and /ga/. In addition, the second formant transition rises for /ba/, and it falls for /da/ and /ga/. This may be the reason /ba/ was less confused with the other plosives than /da/ and /ga. Consequently, /ba/ was easier to distinguish. This same pattern of errors was observed in previous studies of perception of consonants in noise [2].

6. Conclusions

The presentation of a noise, either band-pass or band-stop, prior to the speech stimuli, produced an improvement in its recognition, compared with noise gated on and off with the signal. The effect can be obtained with different conditions of masker bandwidth, masker and signal spectral separation, and masker level. The greatest benefit was obtained when the preceding noise was presented either in the same frequency region as the speech, or in the complementary frequency regions, producing around 28% improvement in the recognition scores. Similarly to what occurred in the psychophysical studies of the overshoot effect, a duration of 200 ms of a precursor masker seems to be critical in obtaining this effect. Although recognition is different from signal detection, they probably involve some of the same processing, and similar beneficial effects obtained with detection of signals can be found with identification of speech stimuli.

The results obtained in the present study also suggest that both within-and across-channel processes seem to be involved in the improvement on the recognition of stop-vowel syllables as a consequence of lengthening the duration of the masker. These processes would play an important role in the perception of speech in natural conditions, where background continuous noise of different spectral characteristics is often present. By means of these processing mechanisms, the auditory system would attenuate background noises and, at the same time, preserve the ability to represent changes in amplitudes, enhancing the perceptual representation of the speech signal.

References

[1] T. Cervera, W. A. Ainsworth: Effects of preceding noise on the perception of voiced plosives. Acta Acustica united with Acustica **91** (2005) 132–144.

[2] G. A. Miller, P. E. Nicely: An analysis of perceptual confusions among some english consonants. J. Acoust. Soc. Am. **27** (1955) 338–352.

[3] S. Singh, J. V. Black: Study of twenty six intervocalic consonants as spoken and recognized by four language groups. J. Acoust. Soc. Am. **39** (1966) 372–387.

[4] M. D. Wang, R. C. Bilger: Consonant confusions in noise: a study of perceptual features. J. Acoust. Soc. Am. **54** (1973) 1248–1266.

[5] T. S. Bell, D. D. Dirks, E. C. Carterette: Interactive factors in consonant confusion patterns. J. Acoust. Soc. Am. **85** (1989) 339–346.

[6] W. A. Ainsworth, M. G. F.: Recognition of plosive syllables in noise: comparison of an auditory model with human performance. J. Acoust. Soc. Am. **96** (1994) 687–694.

[7] W. A. Ainsworth: Effects of preceding noise duration on the perception of voiced plosives and vowels. Proc. Eurospeech'95, Madrid, 1995, Vol. 2, 971–974.

[8] L. L. Elliot: Changes in the simultaneous masked threshold of brief tones. J. Acoust. Soc. Am. **38** (1965) 738–746.

[9] E. Zwicker: Temporal effects in simultaneous masking white noise burst. J. Acoust. Soc. Am. **37** (1965) 653–663.

[10] E. Zwicker: Temporal effects in simultaneous masking and loudness. J. Acoust. Soc. Am. **38** (1965) 132–141.

[11] H. Fastl: Temporal masking effects. I. Broadband noise. Acustica **35** (1976) 287–331.

[12] C. C. Wier, D. M. Green: Detection of a tone burst in continuous-and noise maskers. Effects of signal frequency, duration, and masker level. J. Acoust. Soc. Am. **61** (1977) 1298–1300.

[13] D. N. Green: Masking with continuous and pulsed sinusoids. J. Acoust. Soc. Am. **46** (1969) 939–946.

[14] S. P. Bacon, N. F. Viemeister: Simultaneous masking by gated and continuous sinusoidal maskers. J. Acoust. Soc. Am. **78** (1985) 1220–1230.

[15] S. P. Bacon, B. C. J. Moore: Temporal effects in simultaneous pure tone masking. effects of signal frequency, masker/signal frequency ratio, and masker levels. Hear. Res. **23** (1986) 257–266.

[16] D. McFadden: Spectral differences in the ability of temporal gaps to reset the mechanism underlying overshoot. J. Acoust. Soc. Am. **85** (1989) 254–261.

[17] S. P. Bacon, M. A. Smith: Spectral, intensive, and temporal factors influencing overshoot. J. Exp. Psychol. A **43** (1991) 373–399.

[18] R. P. Carlyon, L. J. White: Effect of signal frequency and masker level on the frequency regions responsible for the overshoot effect. J. Acoust. Soc. Am. **91** (1991) 1034–1041.

[19] B. A. Wright: Detectability of simultaneous masked signals as a function of masker bandwidth and configuration for different signal delays. J. Acoust. Soc. Am. **101** (1997) 420–429.

[20] S. P. Bacon, S. Savel: Temporal effects in simultaneous making with on-and off-frequency noise maskers: Effects of signal frequency and masker level. J. Acoust. Soc. Am. **115** (2004) 1674–1683.

[21] J. F. Schouten: The residue, a new component in subjective analysis. Proc. Kon. Ned. Akam. Wetensch. **43** (1940) 356–365.

[22] B. L. Cardozo: Ohm's law and masking. J. Acoust. Soc. Am. **42** (1967) 1193.

[23] M. Kubovy: The sound of silence: a new pitch segregation phenomenon. Bull. Pscychon. Soc. **8** (1976) 256.

[24] N. F. Viemeister: Adaptation of masking. – In: Psychophysical, Physiological and Behavioural Studies in Hearing. G. van den Brink, F. A. Bilsen (eds.). Delft YU. P., The Netherlands, 1980, 190–199.

[25] N. F. Viemeister, S. Bacon: Forward masking by enhanced components in harmonic complexes. J. Acoust. Soc. Am. **71** (1982) 1502–1507.

[26] Q. Summerfield, M. P. Haggard, J. Foster, S. Gray: Perceiving vowels from uniform spectra: phonetic exploration of an auditory after-effect. Percept. Psychophys. **35** (1984) 203–213.

[27] Q. Summerfield, A. Sidwell, T. Nelson: Auditory enhancement of changes in spectral amplitude. J. Acoust. Soc, Am. **81** (1987) 700–708.

[28] B. A. Wright, D. McFadden, C. A. Champlin: Adaptation of suppression as an explanation of enhancement effects. J. Acoust. Soc. Am. **94** (1993) 72–82.

[29] Math Works, Inc.: 386-Matlab user's guide. Math Woks, Inc., South Natick, MA, 1990.

[30] American National Standards Institute: American national standard specification for audiometers (ANSI S3.6-1996). Author, New York, 1996.

[31] B. J. Winner: Statistical principles in experimental design. 2nd ed. McGraw-Hill, New York, 1971.

[32] C. E. Bonferroni: Teoria statistica delle classi e calcolo delle probabilitá. Publicación del R Istituto Superiore di Ciencia Economiche e Comerciali di Firenze **8** (1936) 3–62.

[33] Y. M. Bishop, S. E. Fienberg, P. W. Holland: Discrete multivariate analysis. Theory and practice. MIT Press, Cambridge, MA, 1975.

[34] T. S. Bell, D. D. Dirks, H. Levitt, J. R. Dubno: Log-linear modeling of consonant confusion data. J. Acoust. Soc. Am. **79** (1985) 518–525.

[35] S. E. Blumstein, K. N. Stevens: Perceptual invariance and onset spectra for stop consonants in different vowel environments. J. Acoust. Soc. Am. **67** (1980) 648–662.

[36] A. M. Liberman, P. C. Delattre, F. S. Cooper, J. L. Gerstman: The role of consonant-vowel transitions in the perception of stop and nasal consonants. Psych. Monographs **68** (1954) 1–13.

[37] D. Kewley-Port: Time-varying features as correlates of place of articulation in stop consonants. J. Acoust. Soc. Am. **73** (1983) 322–335.

[38] R. P. Carlyon: A release from masking by continuous, random notched noise. J. Acoust. Soc. Am. **81** (1987) 418–426.

[39] R. L. Smith, J. J. Zwislocki: Short-term adaptation and incremental response of single auditory-nerve fibers. Biol. Cybernet. **17** (1975) 169–182.

[40] E. F. Evans: Auditory frequency selectivity and the cochlear nerve. – In: Facts and Models of Hearing. E. Zwicker, E. Terhard (eds.). Springer-Verlag, Berlin, 1975, 118–132.

[41] D. J. Gibson, E. D. Young, J. A. Costalupes: Similarity of dynamic range adjustment in auditory nerve and cochlear nuclei. J. Neurol. **53** (1985) 940–958.

[42] S. P. Bacon, E. W. Healy: Effects of ipsilateral and contralateral precursors on the temporal effect in simultaneous masking with pure tones. J. Acoust. Soc. Am. **107** (2000) 1589–1597.

[43] S. P. Bacon, L. Liu: Effects of ipsilateral and contralateral precursors on overshoot. J. Acoust. Soc. Am. **108** (2000) 1811–1818.

[44] E. H. Warren III, M. C. Liberman: Effects of contralateral sound on auditory-nerve responses. I. Contributions of cochlear efferents. Hear. Res. **37** (1989) 89–104.

[45] C. W. Turner, K. A. Doherty: Temporal masking and the active process in normal and hearing-impaired listeners. – In: Modeling Sensoneural Hearing Loss. W. Jestead (ed.). Erlbaum, Hillsdale, NJ, 1997, 387–396.