ELSEVIER

Speech Communication 30 (2000) 197–206

SPEECH COMMUNICATION

www.elsevier.nl/locate/specom

Masking protection in the perception of auditory objects

Peter C. Gordon [1]

Department of Psychology, The University of North Carolina at Chapel Hill, CB# 3270, Davie Hall, Chapel Hill, NC 27599-3270, USA

Received 25 February 1999; received in revised form 28 June 1999; accepted 17 August 1999

Abstract

Three experiments demonstrate the phenomenon of masking protection, where the threshold for identifying a brief masked signal is lowered when that signal is presented in conjunction with other sounds that provide no information about the correct response and which are separated from the distinctive signal by more than a critical band (Gordon, 1997a,b). The first experiment shows that listeners' thresholds for distinguishing a low tone (375 Hz) from a high tone (625 Hz) is lower when those tones are accompanied by a synthetic speech sound that combines with the tones to give percepts of /ɪ/ or /ɛ/, respectively. This effect is reliable for individual listeners and involves a change in perceptual sensitivity. The second experiment shows that a similar lowering of identification thresholds is produced when the distinctive signals are combined with high-frequency, acoustic energy that does not prompt a speech percept. The third experiment shows that identification thresholds are elevated when the non-distinctive, high-frequency acoustic energy leads and lags the distinctive signals. The results of the experiments indicate that mechanisms of perceptual object formation that exploit the temporal alignment of energy changes across the spectrum can contribute to the accurate identification of speech and non-speech sounds. © 2000 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Drei Experimente demonstrieren das Phänomen des Maskierungsschutzes ("masking protection"). Hierbei sinkt die Erkennungsschwelle für ein kurzes maskiertes Signal, wenn das Signal in Gegenwart anderer akustischer Reize dargeboten wird, welche keinen Informationsgehalt bzgl. der korrekten Antwort besitzen und die vom Zielsignal weiter als eine kritische Bandbreite entfernt sind (Gordon, 1997a,b). Das erste Experiment zeigt, daß die Wahrnehmungsschwelle für den Unterschied zwischen einem tiefen (375 Hz) und einem hohen (625 Hz) Ton niedriger ist, wenn die Töne in Gegenwart eines synthetischen Sprachreizes dargeboten werden, der in Kombination mit den Tönen beim Hörer die Wahrnehmung eines /I/ bzw. /ɛ/-Klangs erzeugt. Der Effekt ist auf individueller Ebene reliabel und bedeutet eine Veränderung der Wahrnehmungssensitivität. Das zweite Experiment zeigt, daß eine ähnliche Senkung von Erkennungssschwellen erfolgt, wenn die zu diskriminierenden Signale zusammen mit hochfrequenter akustischer Energie dargeboten werden, welche keine Sprachwahrnehmung hervorruft. Das dritte Experiment zeigt, daß Erkennungsschwellen ansteigen, wenn die Beschallung mit unspezifischer hochfrequenter Energie den zu diskriminierenden Signalen vorhergeht oder folgt. Die Ergebnisse der Experimente weisen darauf hin, daß Mechanismen der

[1] Tel.: +1-919-962-2440; fax: +1-919-962-2537.
E-mail address: pcg@email.unc.edu (P.C. Gordon).

PII: S 0 1 6 7 - 6 3 9 3 (9 9) 0 0 0 5 3 - 9

perzeptuellen Objektbildung, die auf der zeitlichen Anordnung von Energieveränderungen über das Spektrum hinweg beruhen, zur Identifikation von sprachlichen und nichtsprachlichen Klängen beitragen.

Keywords: Auditory objects; Identification; Masking; Vowels; Thresholds

1. Introduction

This paper explores a new phenomenon, masking protection, as a means to better understanding the perception of speech and other types of auditory objects. Masking protection was recently demonstrated (Gordon, 1997a,b) by studying identification thresholds for masked sounds in two conditions. In the first condition, listeners had to identify a simple sound as one of two *target signals*. In the second condition, listeners had to identify a stimulus as one of two complex sounds; these complex sounds were created by adding an identical *cosignal* to each of the target signals. Identification thresholds were lower when listeners identified the complex sound created by pairing the target signal with the cosignal than when they identified the simple sound consisting of only the target signal. Thus the presence of the cosignal protected the target signal from some degree of masking.

Masking protection occurs even though the target signal is the only basis for distinguishing any of the stimuli. Masking protection was first demonstrated for speech sounds where the target signal was energy associated with the first formant of a vowel (low frequency for the vowel /ɪ/ as in 'bit' and higher frequency for the vowel /ɛ/ as in 'bet') and the cosignal was energy associated with the higher formants of a vowel (Gordon, 1997a). It was subsequently demonstrated for non-speech stimuli consisting of brief noise bursts (Gordon, 1997b). For both speech sounds and noise bursts, masking protection was observed only when there was some synchrony in the temporal pattern of energy in the target signal and cosignal. Such temporal patterning of energy is an important basis of perceptual coherence in audition. Accordingly, the dependence of changes in identification thresholds on temporal structure led to the effect being called coherence masking protection.

Masking protection occurs even though the relevant spectral prominences in the stimulus are separated by more than a critical band indicating that there is some kind of interaction in the perceptual processing of acoustic energy at spectrally well-separated frequencies. An interaction of this sort would not be expected under traditional models of hearing that emphasized the operation of independent sensory channels tuned to narrow frequency bands. However, such an interaction is not a complete surprise given developments in psychoacoustics during the 1980s. The psychoacoustic paradigms of comodulation masking release (or CMR; Hall et al., 1984), modulation detection interference (or MDI; Yost et al., 1989) and profile analysis (Green, 1988) have had a major impact on the understanding of hearing by showing that across-frequency mechanisms operate early in perceptual processing. Masking protection is similar to the phenomena demonstrated in these psychoacoustic paradigms in providing evidence of across-frequency interactions in the thresholds of practiced listeners engaged in a simple task in which feedback is given on the accuracy of performance providing listeners with a basis for optimizing their performance.

2. Experiment 1

This experiment compared thresholds for identifying target signals consisting of tones to thresholds for identifying the same target signals when they were paired with a cosignal consisting of the second and third formants of a vowel. The characteristics of the target signals and the cosignal were selected so that when they were combined, they were perceived as the vowels /ɪ/ and /ɛ/. The stimuli were presented in noise that was low-pass filtered so that it spectrally overlapped the energy in the target signal, but was remote from the second and third formants of the vowel cosignal. Comparing performance on vowel stimuli to tone stimuli allowed us to determine whether identification of the tones was more resistant to masking based on their contribution to the vowel as compared to their identification in isolation.

2.1. Method

Stimuli. For the no-cosignal condition, the stimuli consisted of two target signals, sine waves at 375 and 625 Hz. For the vowel-cosignal condition, the stimulus consisted of energy above the first formant that was appropriate for either the vowel /ɛ/ or /ɪ/ . The vowel cosignal was created using the cascade configuration of the Klatt synthesizer. A token of the vowel /ɛ/ was synthesized based on parameters used by Darwin and Gardner (1986) (see also Gordon, 1997a). The vowel had the following characteristics: the fundamental frequency was 125 Hz, the first formant had a center frequency of 625 Hz with a bandwidth of 50 Hz, 3 dB down from the peak amplitude, the second formant had a center frequency of 2200 Hz with a bandwidth of 110 Hz, 3 dB down from the peak amplitude and the third formant had a center frequency of 2900 Hz with a bandwidth of 170 Hz, 3 dB down from the peak amplitude. The synthesized vowel was high-pass filtered at 1250 Hz using a digital filter (Kaiser and Reed, 1978) with a 50 dB stop-band attenuation and a 250-Hz transition band in order to create a sound that lacked the energy at the first formant. This high-pass portion of /ɛ/ was digitally mixed with the low tone (375 Hz) to create the vowel /ɪ/ and with the high tone (625 Hz) to create the vowel /ɛ/. The tones and vowel stimuli were shortened to 40 ms including 5-ms linear ramps at onset and offset using a waveform editor.

The masker was a 600-ms white noise, low-pass filtered at 1000 Hz and presented at 62 dB SPL. The target signal (tone or vowel) began 420 ms after the onset of the masking noise.

Procedure and design. Eight subjects were tested in the experiment. On each trial, a single signal in noise was presented and subjects were asked to identify it by pressing the appropriate key. A one-up, three-down adaptive-tracking procedure was used to adjust the signal level in order to determine listeners' identification thresholds. The amplitude of both the target signal and the cosignal (when present) were adjusted by this method. The initial amplitude of the tone portion of the stimuli was 74 dB SPL. The step size of the signal adjustment was 8 dB for the first two reversals, 4 dB for the next two reversals and 2 dB for the final 12 reversals in a run. The average signal level of the last eight reversals was taken as the threshold for the run. An error indicator was presented after incorrect responses.

Identification thresholds were measured in two conditions. The first condition was identification of vowel stimuli (/ɛ/ or /ɪ/). Listeners were told that they were to decide whether the signal sounded like the vowel /ɛ/ as in 'bet' or /ɪ/ as in 'bit'. They were to press the key labeled '1' if it was /ɪ/ and '2' if it was /ɛ/. The second condition was identification of the tone. Listeners were told that they would hear either a tone with a high pitch or a tone with a low pitch. They were to use the keys labeled '1' for the low tone and '2' for the high tone in order to identify these sounds. Listeners performed nine runs in each condition, rotating through the conditions in the order listed above. The first three runs in each condition were considered practice and were not included in the analysis.

2.2. Results and discussion

Fig. 1 shows the mean signal level at threshold for both experimental conditions for the eight listeners tested in the study. Seven of the eight listeners showed significantly lower thresholds for the vowel-identification task than for the tone-identification task. Over all the listeners, thresholds were 3.4 dB lower in the vowel-cosignal condition than in the no-cosignal condition, $F(1,7) = 31.14$, $p < 0.001$. Thus, pairing the tone with the higher formants of the vowel provided protection from masking.

Adaptive-tracking procedures are typically used in multiple-interval tasks and not in single-interval tasks like the one used here (Macmillan and Creelman, 1991). The reason for this is that the

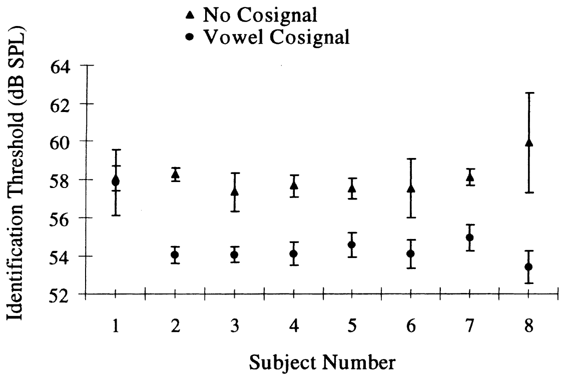

Fig. 1. The mean levels (dB SPL) of the target signals at threshold for identification are shown when the target signal was presented with no cosignal and with a cosignal consisting of the high-pass portion of the vowel. Error bars correspond to the standard error of the mean.

adaptive procedure tracks the signal level at which listeners achieve a certain percent correct (converging on 79.4% correct in a three-down, one-up procedure (Levitt, 1971)). Percent correct is a true measure of sensitivity when there is no response bias, but progressively underestimates sensitivity as response bias increases. Multiple interval tasks have tended not to show much response bias and therefore are generally assumed to be suitable for adaptive-testing methods (MacMillan and Creelman, 1991). There is no comparable empirical assurance that a given single-interval task will not show much response bias. Accordingly, it is important to show that the difference between performance in vowel identification and tone identification did not result from there being greater response bias in the tone-identification task (where there was no cosignal) than in the vowel-identification task (where there was a vowel cosignal). In order to do so, the data for each subject in each condition (vowel and tone identification) were partitioned by the amplitude of the signal and d' was calculated. This analysis was done for signal levels near the empirical thresholds where observations were present for both conditions.

Table 1 shows the results of the analysis. The d' measure was significantly higher in the vowel-identification task than in the tone-identification task for signal levels of 53, 55 and 57 dB SPL. Thus, a bias-free measure of sensitivity (d') shows that vowel identification is superior to tone identification in the present task.

The results of the experiment show that pairing the target tone with the higher formants of the vowel reduced thresholds for masked identification, complementing results obtained by Gordon (1997a). Thus, masking protection is an example of an interesting class of perceptual phenomena where a component of a complex stimulus is perceived more easily when it is part of the complex stimulus than when it is presented on its own. Perhaps the most studied example of this phenomenon is the *word superiority effect* where the threshold for recognizing a briefly presented visual letter is lower when that letter is presented in a word than when it is presented in isolation (Reicher, 1969). In the domain of speech perception, masking protection appears to be similar to the 'phonetic precedence' effect reported by Whalen and Liberman (1987). That effect is an instance of duplex perception, where acoustic energy contributes to the perceptual identity of a speech sound and simultaneously produces its own non-speech percept (Rand, 1974; Liberman et al., 1981). Whalen and Liberman showed that duplex perception could be induced with binaural stimulation by varying the intensity of the third-formant transition, a cue to the distinction between 'da' and 'ga' in their study. At high intensities, listeners heard a speech sound and also a chirp corresponding to the particular third formant transition. Manipulation of the intensities showed that subjects could identify the speech sounds below the threshold for detecting the chirp as an independent percept. The phenomenon reported by Whalen and Liberman is similar to masking protection because in both cases listeners are more sensitive to simple acoustic features based on the contribution of those simple features to the perception of a complex sound than based on the perceptions that those features generate independently.

While the phenomena of masking protection and phonetic precedence are similar, the method used to demonstrate masking protection overcomes some uncertainties surrounding the phonetic precedence effect that derive from the fact that duplex perception is a subjective state (Bailey and Herrman, 1993; Hall and Pastore, 1992).

Table 1
Sensitivity (d') for the no-cosignal and vowel-cosignal conditions at different signal levels for individual subjects[a]

	49 dB SPL		51 dB SPL		53 dB SPL		55 dB SPL		57 dB SPL		59 dB SPL	
Subject	No cosignal	Vowel cosignal	No cosignal	Vowel cosignal	No cosignal	Vowel cosignal	No cosignal	Vowel cosignal	No cosignal	Vowel cosignal	No cosignal	Vowel cosignal
1				0.42	0.72	0.88	1.07	1.35	1.41	2.04	1.60	
2				0.40		1.53	0.37	1.94	1.35	2.60	1.40	
3		0.23	1.11	0.97	0.49	1.50	1.54	2.62	0.74		1.93	
4			0.57	0.84	0.50	1.33	1.05	1.65	1.30	3.14	2.56	
5		1.11		1.48	0.68	1.53	0.57	1.71	1.26	2.15	2.65	
6		1.24	0.79	0.90	0.24	1.05	1.32	1.38	2.28	2.37		
7		1.22		0.80	0.82	1.17	0.35	1.91	1.06	2.38	2.09	2.24
8		0.68		1.31		1.45	0.81	1.88	1.16	2.64	2.58	
Mean					0.58 $t(5) = 4.91$, $p < 0.005$	1.24	0.88 $t(7) = 4.67$, $p < 0.005$	1.80	1.40 $t(6) = 4.86$, $p < 0.005$	2.48		

[a] Empty cells indicate conditions in which d' could not be calculated. At higher signal levels this occurred because performance was perfect for at least one of the stimuli (resulting in a d' of infinity). At lower signal levels this occurred because adaptive-tracking of performance caused there to be no observations at that level. The means and significance levels at the bottom of the table were computed based on subjects who had d' scores in both conditions for the indicated signal level.

Tasks that measure duplex perception require that listeners respond based on one aspect of their subjective experience while not using another aspect of their subjective experience; there is no objective way of verifying that listeners are able and willing to do so. In contrast, the masking protection paradigm demonstrated here does not rely on listeners' discernment about the appropriate subjective basis for responding. Listeners are presented with trial-to-trial feedback and are free to optimize their performance in any way that they can. When the distinctive acoustic information is presented alone performance is not as good as when it is presented with non-distinctive acoustic information with which it can be perceptually combined. This difference in performance can be seen in a bias-free measure of perceptual sensitivity.

3. Experiment 2

This experiment examined whether masking protection occurs for complex sounds that lack some of the critical acoustic features of speech sounds, but which preserve the temporal relations that exist between energy in different parts of the spectrum in the speech sounds. A finding that masking protection occurs for non-speech sounds in addition to speech sounds would be most parsimoniously explained by the operation of a general auditory mechanism for perceptual integration rather than by separate mechanisms, one for speech sounds and one for sounds in general.

3.1. Method

Stimuli, design and procedure. The stimuli for the experiment included the no-cosignal and vowel-cosignal stimuli from the previous experiment as well as two additional types of stimuli that changed features of the high-pass component of the vowels. The *complex-tone* cosignal consisted of a series of seven, equal-amplitude harmonics progressing from 2250 to 3000 Hz. This cosignal was similar to the high-pass portion of the vowel in that both consisted of a harmonic progression related to the same fundamental frequency of 125 Hz; the target signal in both cases was also a harmonic of the same fundamental. The complex tone also provided energy in the same range of the spectrum as did the high-pass portion of the vowel. The complex tone differed from the high-pass portion of the vowel in that all of its harmonics had the same amplitude whereas the formant structure in the high-pass portion of the vowel resulted in two prominences in the 2200–2900 Hz region. The stimuli also differed in that energy in the high-pass portion of the vowel extended below and above the frequencies in which energy was present in the complex-tone cosignal. The *noise-band* cosignal consisted of white noise band-pass filtered between 2200 and 2900 Hz. Thus, this cosignal had energy in the same part of the spectrum as the complex-tone cosignal, but it lacked the harmonic structure of the complex-tone cosignal.

The general procedure was the same as in the previous experiment. For the tone-identification task and the vowel-identification task, the characterization of the stimuli given to the listeners was the same as in the previous experiment (i.e., the stimuli were characterized as tones and vowels, respectively). For both the complex-tone and noise-band cosignals, the stimuli were characterized as complex sounds which subjects were instructed to identify on the basis of whether they contained a low-pitch or high-pitch sound. Four listeners performed eight runs in each condition, rotating through the conditions in the order: vowel cosignal, noise-band cosignal, complex-tone cosignal and no cosignal. The first run in each condition was considered practice and was not included in the analysis.

3.2. Results and discussion

Fig. 2 shows the mean signal level at identification threshold for each of the experimental conditions for each listener. A repeated-measures analysis of variance showed that there were significant differences among the four conditions, $F(3,9) = 30.7$, $p < 0.001$. A planned contrast showed that the difference between the no-cosignal condition and the average of the cosignal conditions was significant across subjects,

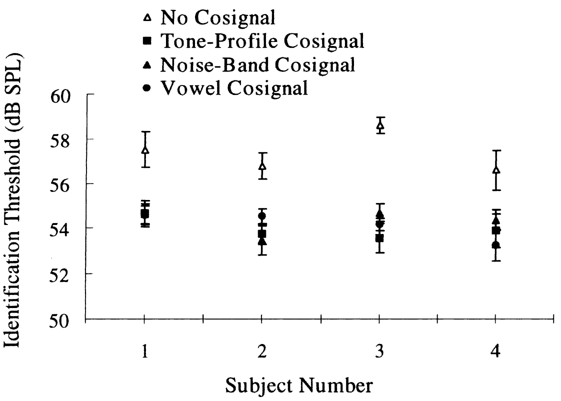

Fig. 2. The mean signal levels (dB SPL) of the target tones at threshold for identification are shown when the target tone was presented with no cosignal, the vowel cosignal, the complex-tone cosignal and the noise-band cosignal. Error bars correspond to the standard error of the mean.

$t(3) = 10.74$, $p < 0.002$. Contrasts showed that were no significant differences among the cosignal-present conditions.

The results show that masking protection of comparable magnitude occurs for speech sounds and non-speech sounds that are matched to the speech sounds on certain dimensions. The complex-tone and noise-band cosignals lacked some of the important features of the high-pass vowel cosignal. The complex-tone lacks the energy contour that the formant structure of the vowel imposes on the harmonics. The noise-band cosignal is aperiodic. The stimuli that resulted from combination of the complex-tone and noise-band cosignals with the target signal did not sound like speech. The finding that masking protection occurs for both speech sounds and non-speech sounds is most parsimoniously explained by operation of general auditory mechanisms.

The results of this experiment are consistent with those of Gordon (1997b). That paper presents a series of three experiments in which the noise-band cosignal used in the present experiment was combined with a target signal consisting of a narrow band of noise. Masking protection was exhibited with those stimuli by listeners who were not presented with any speech stimuli during the experiments and were not instructed to identify any stimuli as speech. Thus, masking protection in those experiments is very unlikely to have occurred because listeners attempted to hear the non-speech sounds in a speech mode of listening. The current findings add to those reported by Gordon (1997b) by extending the range of non-speech stimuli for which masking protection has been observed and by showing that masking protection in such stimuli is reliable for individual listeners.

The results of the current experiment along with those of Gordon (1997b) are broadly similar to other findings where adding noise (or other non-informative acoustic energy) facilitates the identification of a signal. Spectral restoration of speech, as shown by Warren et al. (1997), is an effect where the intelligibility of speech that has been filtered into two very narrow bands is enhanced when broadband noise is inserted between the two speech bands. The spectral restoration effect has been studied using sentence-length speech stimuli and has been found to be sensitive to the degree of linguistic redundancy in the stimuli. This contrasts with the masking protection paradigm which uses very brief stimuli that lack any linguistic redundancy. Thus, different mechanisms may underlie these two effects though both reflect listener's ability to perceive speech in challenging acoustic environments. Another finding, by Shriberg (1992), shows that noise can induce perceptual restoration in filtered vowels and in doing so enhances vowel identification. Shriberg's stimuli are similar in duration and linguistic redundancy to those used in the masking protection paradigm suggesting that similar processing mechanisms may underlie the two effects.

4. Experiment 3

This experiment examined the dependence of identification thresholds on the temporal alignment of energy in the cosignal and target signal. A finding that identification thresholds were influenced by temporal alignment of parts of the stimulus would indicate that masking protection is due in part to the configuration of the parts of the stimulus. Research by Darwin and his colleagues (Darwin, 1984a,b; Darwin and Cioca, 1992; Darwin and Sutherland, 1984; Hill and Darwin, 1996) on supra-threshold vowel identification, pitch perception and binaural fusion has shown that asynchronies in the onsets and offsets of acoustic energy can influence the degree to which acoustic energy contributes to phonetic perception. Further, research in the streaming paradigm (Bregman and Pinker, 1978) has shown that alignment of the onsets and offsets of tones plays a role in whether they are perceived as part of the same or different auditory streams. The masking protection paradigm complements these methods in important ways. Feedback can be given to the listener in masking protection because the task involves an objectively correct answer. Providing a listener with substantial practice and feedback promotes the development of optimal (or near-optimal) task strategies, which increases the likelihood that effects in the masking protection task reveal basic perceptual processes rather than higher-level strategic or decision processes. In contrast, current paradigms for suprathreshold identification do not involve objectively correct answers and feedback, therefore results may reflect either basic perceptual processes or decision processes.

4.1. Method

Stimuli, design and procedure. The stimuli varied in the temporal relation between the cosignal (consisting of the high-pass portion of the vowel) and the target signal (consisting of a tone of either 375 or 625 Hz). In the *synchronous*-cosignal condition, the cosignal and target signal were gated on and off simultaneously; the stimuli in this condition matched those used in the vowel-cosignal conditions of Experiments 1 and 2. In the *fringing*-cosignal condition, the cosignal began 40 ms before the target signal and ended 40 ms after the target signal. Asynchronies of this magnitude have been shown to decrease the contribution of acoustic energy into a phonetic percept in suprathreshold identification tasks (Darwin, 1984a,b; Darwin and Sutherland, 1984).

The general procedure was the same as in the previous experiment. Ten subjects were instructed to identify both the synchronous and fringing stimuli as vowels. They performed six runs in each condition, alternating between the conditions. The first two runs in each condition were considered practice. Two listeners (Subjects 1 and 2) in the experiment also participated in the previous experiment.

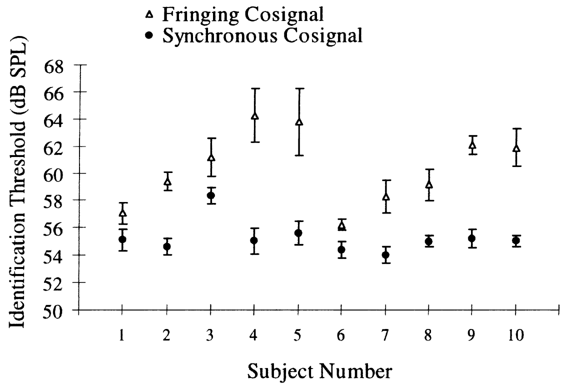

Fig. 3. The mean signal levels (dB SPL) of the target tones at threshold for identification are shown when the target tone was presented with a synchronous vowel cosignal and with a fringing cosignal which began 40 ms before the target signal and ended 40 ms after the target signal. Error bars correspond to the standard error of the mean.

4.2. Results and discussion

Fig. 3 shows the mean signal level at identification threshold for both experimental conditions. Listeners showed lower thresholds, by an average of 7.1 dB, for the synchronous stimuli compared to the fringing stimuli. This difference was significant as shown by a repeated-measures analysis of variance, $F(1,9) = 38.49,\ p < 0.001$.

These results show that identification thresholds for the vowels are significantly higher when the vowel cosignal both leads and trails the target signal by 40 ms (fringing condition) than when the cosignal and target signal are gated on and off simultaneously (synchronous condition), a finding that shows that threshold-level identification is not simply influenced by having the cosignal present simultaneously with the target signal. This finding is consistent with findings by Darwin (Darwin, 1984a,b; Darwin and Sutherland, 1984) that onset and offset asynchronies between a tone and a vowel reduce the contribution of the tone to the identification of the vowel. The present finding shows that such asynchronies influence threshold-level identification in addition to their influence in the suprathreshold task used by Darwin, indicating that in this case the two types of methods give converging results. Nonetheless, the masking protection paradigm provides an important check on results obtained with suprathreshold identification and it addresses a facet of perceptual integration – protection from masking – that is not assessed by suprathreshold identification.

The results of the experiment are also consistent with previous findings on masking protection in the perception of vowels (Gordon, 1997a) and of noise bursts (Gordon, 1997b). Gordon (1997a) measured identification thresholds for noise-band target signals (50-Hz wide centered on 375 and 625 Hz) in three conditions: synchronous vowel cosignal, fringing vowel cosignal and no cosignal. Gordon (1997b) measured identification thresholds for the same target signals, substituting the noise-band cosignal for the vowel cosignal. In both studies, identification thresholds were lowest in the synchronous-cosignal condition. Identification thresholds did not differ significantly in the fringing-cosignal and no-cosignal conditions. Those findings show that masking protection that is present with synchronous cosignals is eliminated with fringing cosignals for both speech sounds and non-speech sounds.

5. Summary

Masking protection occurs when the threshold for identifying a masked signal is reduced by the presence of spectrally distant acoustic energy with which the signal can form a coherent perceptual object (Gordon, 1997a,b). The present paper extends previous research on masking protection in the following ways. (1) Masking protection was shown to occur for a broader range of stimulus types than have been used before. In particular, masking protection was observed for target signals consisting of tones whereas previous work had shown it for target signals consisting of noise bands and synthesized first formants. Further, masking protection was shown with cosignals consisting of complex tones of equal amplitude tones in addition to the noise bands and synthesized higher formants that have been used previously. (2) Masking protection was demonstrated to be reliable for individual listeners. (3) Masking protection was shown to involve a change in listeners' sensitivity in identifying the target signal and not to result from criterion shifts. In particular, in a critical range of signal levels, d' was shown to be greater in the target-signal-plus-cosignal condition than in the target-signal-alone condition.

The present work in combination with earlier work (Gordon, 1997a,b) establish masking protection as a clear example of where the perception of a component of a stimulus is enhanced by its being part of a more complex stimulus despite substantial spectral separation between the parts of the stimulus. Gordon (1997a,b) discusses two possible mechanisms for this enhancement. The first mechanism is that the cosignal serves as a temporal marker that helps the listener locate the target signal in time, thereby facilitating its recognition. The second mechanism is based on the idea that the cosignal makes the listener more sensitive to the frequency of the target signal because it provides a concurrent perceptual basis for estimating the expected energy level at the possible frequencies of the target signal. Similar mechanisms have been advanced to explain the perception of complex acoustic signals in other tasks.

Acknowledgements

The research reported in this paper was supported in part by NSF grant IIS-9811129. I would like to thank Tom Carrell, John Grose, Joe Hall, Bob Peters, Art Samuel and two anonymous reviewers for helpful discussions of this research and/or for comments on this paper.

References

Bailey, P.J., Herrman, P., 1993. A reexamination of duplex perception evoked by intensity differences. Perception & Psychophysics 54, 20–32.

Bregman, A.S., Pinker, S., 1978. Auditory streaming and the building of timbre. Canadian Journal of Psychology 32, 19–31.

Darwin, C.J., 1984a. Auditory processing and speech perception. In: Bouma, H., Bouwhuis, D.G. (Eds.), Attention and

Performance X: Control of Language Processes. Erlbaum Associates, Hillsdale, NJ, pp. 197–210.

Darwin, C.J., 1984b. Perceiving vowels in the presence of another sound: Constraints on formant perception. Journal of the Acoustical Society of America 76, 1636–1647.

Darwin, C.J., Cioca, V., 1992. Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned harmonic. Journal of the Acoustical Society of America 91, 3381–3390.

Darwin, C.J., Gardner, R.B., 1986. Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. Journal of the Acoustical Society of America 79, 838–845.

Darwin, C.J., Sutherland, N.S., 1984. Grouping frequency components of vowels: When is a harmonic not a harmonic?. Quarterly Journal of Experimental Psychology 36A, 193–208.

Gordon, P.C., 1997a. Coherence masking protection in speech sounds: The role of formant synchrony. Perception & Psychophysics 59, 232–242.

Gordon, P.C., 1997b. Coherence masking protection in brief noise complexes: Effects of temporal alignment. Journal of the Acoustical Society of America 102 (4), 2276–2283.

Green, D.M., 1988. Profile analysis: Auditory intensity discrimination. Oxford University Press, New York, NY.

Hall, M.D., Pastore, R.E., 1992. Musical duplex perception: Perception of figurally good chords with subliminal distinguishing tones. Journal of Experimental Psychology: Human Perception and Performance 18, 752–762.

Hall III, J.W., Haggard, M.P., Fernandes, M.A., 1984. Detection in noise by spectro-temporal pattern analysis. Journal of the Acoustical Society of America 76, 50–56.

Hill, N.I., Darwin, C.J., 1996. Lateralization of a perturbed harmonic: effects of onset asynchrony and mistuning. Journal of the Acoustical Society of America 100, 2352–2364.

Kaiser, J.F., Reed, W.A., 1978. Bandpass (bandstop) digital filter design routine. Review of Scientific Instrumentation 48, 1103–1106.

Levitt, H., 1971. Transformed up-down methods in psychoacoustics. Journal of the Acoustical Society of America 49, 467–477.

Liberman, A.M., Isenberg, D., Rackerd, B., 1981. Duplex perception for stop consonants: Evidence for a phonetic mode. Perception & Psychophysics 30, 133–143.

Macmillan, N.A., Creelman, C.D., 1991. Detection theory: A user's guide. Cambridge University Press, New York, NY.

Rand, T.C., 1974. Dichotic release from masking for speech. Journal of the Acoustical Society of America 55, 678–680.

Reicher, G.M., 1969. Perceptual recognition as a function of meaningfulness of stimulus materials. Journal of Experimental Psychology 81, 275–280.

Shriberg, E.E., 1992. Perceptual restoration of filtered vowels with added noise. Language and Speech 35, 127–136.

Warren, R.M., Hainsworth, K.R., Brubaker, B.S., Bashford, J.A., Healy, E.W., 1997. Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps. Perception & Psychophysics 59, 275–283.

Whalen, D.H., Liberman, A.M., 1987. Speech perception takes precedence over nonspeech perception. Science 237, 169–171.

Yost, W.A., Sheft, S., Opie, J., 1989. Modulation detection interference: Effect of modulation frequency. Journal of the Acoustical Society of America 86, 2138–2148.