Coherence masking protection in brief noise complexes: Effects of temporal patterns

Peter C. Gordon[a)]
Department of Psychology, The University of North Carolina at Chapel Hill, CB# 3270, Davie Hall, Chapel Hill, North Carolina 27599-3270

(Received 25 April 1996; revised 2 November 1996; accepted 2 May 1997)

Three experiments examined listeners' thresholds for classifying the pitch of a target signal in a masking noise when it was presented alone as compared to when it was presented with a ''cosignal.'' The target signal was a narrow band of noise centered on either 375 or 625 Hz and the masker was noise low-pass filtered at 1000 Hz. The cosignal provided no information about the pitch of the target signal but could potentially combine with it to form an auditory object; it was spectrally well separated from the target signal, consisting of a band of noise ranging from 2200 to 2900 Hz. Experiment 1 showed that identification thresholds were lower when the target signal was paired with the cosignal than when it was presented alone if the onsets and offsets of the target signal and cosignal were temporally synchronous. This is an instance of ''coherence masking protection,'' a phenomenon that has previously been established in the perception of vowels [P. C. Gordon, Percept. Psychophys. **59**, 232–242 (1997)]. The effect disappears when the cosignal leads and lags the target signal by short durations, a finding that also matches that observed previously with vowels. The finding that temporal relations between the components of a stimulus have similar effects on the perception of nonspeech noise complexes and speech sounds suggests that speech perception makes use of general auditory mechanisms for perceptual integration of this sort. Experiments 2 and 3 examine further the role of temporal relations between the onsets and offsets of the target signal and the cosignal in producing coherence masking protection. The results show that either onset synchrony or offset synchrony is sufficient to produce the effect when the cosignal is of greater duration than the target signal, but that only onset synchrony produces the effect when the target signal has greater duration than the cosignal. This pattern indicates that the target signal and cosignal do not contribute equally to the formation of auditory objects. [S0001-4966(97)03609-6]

PACS numbers: 43.66.Dc, 43.66.Mk [RHD]

INTRODUCTION

Hypotheses about the processes underlying phonetic perception have frequently been tested and refined by comparing the perception of speech stimuli to the perception of nonspeech stimuli that mimic some properties of speech sounds (e.g., Liberman *et al.*, 1967; Mann and Liberman, 1983; Pisoni, 1977; Remez, 1980). Such comparisons have been made in order to determine whether characteristics of phonetic perception must be explained by speech-specific processes or whether they can be explained in terms of the operation of general auditory mechanisms. The rationale is that if phonetic perception differs from nonphonetic perception an appeal to specialized mechanisms is warranted, but a finding that phonetic and nonphonetic perception are very similar is most parsimoniously explained by appeal to general auditory mechanisms. The present paper applies this rationale to the integration of acoustic information in phonetic perception as it is shown by the phenomenon of coherence masking protection (Gordon, 1997).

Gordon (1997) demonstrated coherence masking protection (CMP) in speech sounds using a paradigm in which identification thresholds for speech sounds in noise were compared to identification thresholds for the acoustic information that distinguished the speech sounds when it was isolated from the remainder of the speech sound. Under certain conditions, identification thresholds were lower for the speech sounds than for the distinctive information alone, indicating that being part of a coherent speech object protected the distinctive information from masking. More specifically, Gordon (1997) had listeners classify a stimulus as /ɪ/ (as in ''bit'') or /ɛ/ (as in ''bet''), a distinction that can be minimally cued by the frequency of the first formant. When the signals were presented in a low-pass masking noise, identification thresholds for the vowels were lower than identification thresholds for the acoustic energy underlying the first formant even though that energy provided the only basis for distinguishing the vowels.

Development of the CMP paradigm was motivated in part by findings obtained in the comodulation masking release (CMR) paradigm (Hall *et al.*, 1984; Hall and Grose, 1988, 1990). In that paradigm, changes in detection thresholds for simple signals are studied as a function of the addition of energy bands to the masker at frequencies that are widely separated from the signal. When the amplitude modulation of the added energy bands has the same envelope as the on-signal masking band, thresholds are reduced. No effect on thresholds is observed when the envelopes of the

a)Electronic mail: pcg@gibbs.oit.unc.edu

added energy bands differ from that of the on-signal masker. The CMR paradigm provides a way of studying how factors promoting auditory coherence in a masker can release a signal from masking (Hall and Grose, 1990). CMP builds on this logic by examining how coherence within a signal may protect a signal from masking (Gordon, 1997).

Gordon (1997) studied CMP in steady-state vowels. As Darwin (1981) has noted, there are two salient acoustic bases for coherence in such stimuli: synchrony of the onsets and offsets of the formants and the relation of the harmonics to a common fundamental. Research using techniques developed by Darwin (Darwin, 1984a, 1984b; Darwin and Gardner, 1986; Roberts and Moore, 1990, 1991) has shown that both of these factors play a role in determining whether acoustic energy contributes to the phonetic identification of sounds presented at suprathreshold levels. Gordon (1997) focused on synchrony of formants as a basis for the threshold-level coherence measured by CMP. Vowel sounds were created in which the harmonic structure at low frequencies was eliminated and the distinctive first formant was simulated by a narrow band of noise. CMP was observed with these stimuli if the higher formants and first formant were coterminous, but not when the higher formant began in advance and ended after the first formant. This result showed that synchrony of onsets and offsets was a sufficient basis for CMP in vowel sounds, even in the absence of a harmonic basis for coherence. The current experiments examine whether different types of synchrony provide a basis for CMP in nonspeech sounds. This serves two goals: to provide a basis for comparing perceptual integration in speech and nonspeech stimuli, and to understand better how synchrony of changes in energy across different parts of the spectrum influences the creation of auditory objects.

I. EXPERIMENT 1. CMP WITH SYNCHRONOUS COSIGNALS VERSUS FRINGING COSIGNALS

The nonspeech stimuli in the present experiment were designed to mimic some of the central properties of the stimuli used in the third experiment of Gordon (1997). That experiment studied identification of the vowels /ɪ/ and /ɛ/ that were constructed by combining a distinctive first formant consisting of a narrow band of noise with higher formants produced by the Klatt synthesizer. The noise-band first formant was 50 Hz wide and was centered on 375 Hz for /ɪ/ and 625 Hz for /ɛ/. The higher formants were identical for the two vowels; in particular, $F2$ was set at 2200 Hz and $F3$ was set at 2900 Hz. Identification thresholds in low-pass noise were determined for three types of targets. In the *synchronous-formants* condition, the higher formants and the noise band both had a duration of 40 ms and were gated on and off together. In the *fringing-formants* condition, the higher formants had a duration of 120 ms while the noise band had a duration of 40 ms; the higher formants began 40 ms before the noiseband and ended 40 ms after it. In the *no-formants* condition, only the noise band was presented. In the two conditions in which higher formants were presented, listeners identified the target stimulus as one of the two vowels (/ɪ/ vs /ɛ/). In the no-formants condition, listeners identified the noise band as a low- or high-pitched sound. Identification thresholds were lowest in the synchronous-formants condition; they did not differ significantly in the fringing- and no-formants conditions.

Nonspeech analogs of the synchronous-formants and fringing-formants conditions were created by replacing the formants with a bandlimited white noise that ranged from 2200 Hz to 2900 Hz; this bandlimited noise will be referred to as the *cosignal*. The cosignal was constructed so that it had energy in the frequency range of the second and third formants of the stimuli used by Gordon (1997). The cosignal did not prompt a phonetic percept in the judgment of the author. Listeners in the experiment were not told to identify the stimuli as speech, and none reported hearing them as such. Accordingly, if a speech-specific mechanism were responsible for integrating the higher formants with the first formant in the Gordon (1997) experiments, then integration of the cosignal with the first formant would not necessarily be expected in the current experiments. Alternatively, if general auditory mechanisms were responsible for the integration observed by Gordon (1997), then integration of the cosignal with the first formant would be expected in the current experiments.

The cosignal differed from the higher-formant stimulus in that it had a flat spectrum in the range of the second and third formants while the higher-formant stimulus contained two prominences in this range. Further, the higher-formant stimulus had a harmonic progression built on a fundamental of 125 Hz that began at 1200 Hz (due to the high-pass filtering that was used to eliminate information about the first formant) and extended to 4700 Hz (the cutoff of the anti-aliasing filter). These differences meant that while the formant stimuli had a pitch related to the fundamental of 125 Hz and a timbre reflecting the prominences of the formants, the cosignal sounded like a moderately high-frequency noise. For the higher-formant stimulus used by Gordon (1997), combination with the noise-band first formant produced a clear impression of a vowel, the identity of which was determined by the frequency of the target signal. For the cosignal used in the current study, this combination created the impression of a noise with a tone in it; the pitch of the tone was determined by the frequency of the target signal.

The current experiment examined identification thresholds for the target signals when they were paired with a synchronous cosignal, a fringing cosignal, or no cosignal, thereby matching the temporal patterns used by Gordon (1997) with the higher-formant stimulus.

A. METHOD

1. Subjects

Twelve subjects participated in a single session that lasted approximately an hour and a half. They were recruited with posted notices and were paid at a rate of $6/h. for their participation. To be included in the experiment, subjects had to meet a criterion of average identification thresholds of 64 dB SPL in the first six runs of the experiment. All subjects tested met this criterion.

2. Stimuli

Two 50-Hz-wide bands of noise, one centered on 375 Hz and the other on 625 Hz served as the target signals in the task. The noisebands were made by passing a broadband (0–2000 Hz), constant spectrum-level noise through a digital filter (IHR Universal) with extremely sharp spectral skirts and a noisefloor over 70 dB down. The sampling rate of the filter was 2500 Hz and the output was low-pass filtered at 1250 and recorded onto digital audio tape.[1] Playback of the tape was then redigitized at 10 kHz using a Kay Elemetrics CSL system. The noises were edited into 40-ms stimuli with 5-ms linear onset and offset ramps. Nine different 40-ms stimuli were made from each noise so that the fluctuations present in the narrow bands of noise would not be the same in each stimulus presentation; the starting level of the signal was 69 dB SPL. The cosignal consisted of a bandpass noise between 2200 and 2900 Hz; its starting level was 62 dB SPL and it began and ended with 5-ms linear ramps. The masking noise consisted of a 600-ms noise low-pass filtered at 1000 Hz, and it was presented at approximately 62 dB SPL. In the synchronous-cosignal condition, both the signal and cosignal began 420 ms into the masker. In the fringing-cosignal condition, the cosignal began 380 ms into the masker (ending 120 ms later), and the signal began 420 ms into the masker. In the no-cosignal condition, the 40-ms noise band began 420 ms into the masker.

3. Procedure and design

On each trial, a single stimulus consisting of a target signal and accompanying cosignal was presented in the masking noise; subjects were asked to identify it as a low-pitched or high-pitched sound by pressing the appropriate key. A one-up, three-down adaptive tracking procedure was used to determine listeners' thresholds. Both the level of the signal and cosignal were adjusted during tracking. After incorrect responses, a visual error message was presented to the subject. No overt message was presented after correct responses. The step size of the signal and cosignal adjustment was 8 dB for the first 2 reversals, 4 dB for the next 2 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. Subjects performed 18 runs, rotating through the conditions in the order: synchronous cosignal, fringing cosignal, and no cosignal. After every group of three runs, subjects were shown their identification threshold averaged over the preceding three runs and were encouraged to try as hard as possible to reduce this threshold in the remainder of the testing. This feedback served to increase subjects' motivation and to provide them with a way of tracking their performance without giving them information on their performance in the different experimental conditions. The first two runs in each condition were considered practice and were not included in the analysis.

B. Results

Table I shows the mean signal level at threshold in the three experimental conditions for individual subjects as well as the overall means and standard deviations. Analysis of variance showed that performance in the three conditions differed significantly, $F(2,22)=18.4$, $p<0.001$. Identification thresholds were lower in the synchronous-cosignal condition than in both the fringing-cosignal condition [$t(11)=4.73$, $p<0.001$] and the no-cosignal condition, $t(11)=4.28$, $p<0.002$. Identification thresholds did not differ significantly in the fringing-cosignal and no-cosignal conditions, $t(11)=0.44$, $p>0.25$.

TABLE I. Results of experiment 1. Mean signal level (dB SPL) at identification threshold for target signals with synchronous cosignals, fringing cosignals, and no cosignals.

Subject number	Synchronous cosignal	Fringing cosignal	No cosignal
1	54.6	55.6	57.9
2	57.4	57.9	55.0
3	60.3	63.4	57.5
4	54.4	57.7	57.9
5	54.1	57.0	56.8
6	58.5	60.7	61.3
7	52.3	57.2	56.0
8	59.0	60.7	59.3
9	56.1	58.7	60.8
10	55.1	58.7	60.0
11	58.9	63.7	61.9
12	53.4	56.1	58.2
Mean	56.2 (2.6)	58.9 (2.6)	58.6 (2.2)

C. Discussion

The results showed a significant CMP; identification thresholds were lower in the synchronous cosignal condition than in the no-cosignal condition, indicating that the identification of the target signal was facilitated by the presence of the cosignal which of itself provided no information about the frequency of the target signal. No CMP was observed for the fringing-cosignal condition, as shown by the lack of difference between that condition and the no-cosignal condition. This pattern of results for nonspeech stimuli exactly parallels the findings of Gordon (1997) for vowel stimuli with matched temporal patterns. In both cases, CMP was observed only when the high-frequency energy was synchronous with the distinctive signal. The finding of parallel results for speech and nonspeech stimuli is most parsimoniously explained by the idea that coherence of the sort that provides protection from masking derives from general processes of auditory perception that apply across domains.

II. EXPERIMENT 2: CMP WITH TEMPORALLY LEADING OR LAGGING COSIGNALS

The results of the first experiment demonstrate that the temporal relation between the target signal and the cosignal affects CMP. CMP is observed when their onsets and offsets are simultaneous but it is not observed when the cosignal leads and lags the target signal by 40 ms. Gordon (1997) employed temporal leads and lags of 40 ms in his study of CMP in speech sounds because studies by Darwin and his colleagues have shown that perceptual integration of acoustic components for purposes of phonetic and pitch perception is

influenced considerably by asynchrony of this magnitude (Darwin, 1984a, 1984b; Darwin and Sutherland, 1984; Hukin and Darwin, 1995; Roberts and Moore, 1991), though these findings have been obtained with signals of longer duration than have been studied in the CMP paradigm. This research has further shown that having asynchronous onsets disrupts perceptual integration to a greater degree than having asynchronous offsets. The current experiment examines the role of onset and offset synchrony in the perceptual integration process underlying CMP. It explores whether CMP occurs for *onset-synchronous* stimuli in which the target signal and cosignal begin at the same time but the cosignal extends 40 ms past the offset of the target signal, and whether it occurs for *offset-synchronous* stimuli in which the target signal and cosignal end at the same time but the cosignal begins 40 ms before the target signal. The stimuli in these conditions examine separately the two sources of asynchrony in the fringing stimuli used in the previous experiment.

TABLE II. Results of experiment 2. Mean signal level (dB SPL) at identification threshold for target signals with onset-synchronous cosignals, offset-synchronous cosignals, and no cosignals. Target signals are 40 ms and cosignals are 80 ms.

Subject number	Onset-synchronous cosignal	Offset-synchronous cosignal	No cosignal
1	54.2	56.0	57.5
2	56.2	57.0	57.5
3	56.4	57.1	59.4
4	55.4	54.8	58.0
5	57.0	58.1	60.3
6	53.1	55.4	58.6
7	56.0	55.8	65.4
8	57.2	56.8	57.3
9	58.1	58.3	59.6
10	58.1	56.8	57.3
11	55.9	55.3	60.2
12	55.7	56.8	57.0
Mean	56.1 (1.4)	56.5 (1.1)	59.0 (2.3)

A. Method

1. Subjects

Fifteen subjects from the same population as the previous study participated in the experiment. None of them had participated in the previous study. Three subjects failed to meet the criterion for inclusion in the study and were dismissed after the first six runs.

2. Stimuli, procedure, and design

The signals, cosignals, and masking noise were the same as in the previous experiment, except that the cosignals were shortened to 80 ms. In the onset-synchronous cosignal condition, both the signal and cosignal began 420 ms into the masker; the signal ended 40 ms later and the cosignal ended 80 ms later. In the offset-synchronous cosignal condition, the cosignal began 380 ms into the masker, the signal began 40 ms later. Both ended 460 ms into the masker. The procedure and design were the same as in the preceding experiment.

B. Results

Table II shows the mean signal level at threshold in the three experimental conditions for individual subjects as well as the overall means and standard deviations. Analysis of variance showed that performance in the three conditions differed significantly, $F(2,22)=11.5$, $p<0.001$. Identification thresholds were higher in the no-cosignal condition than in both the onset-synchronous cosignal condition [$t(11)=4.43$, $p<0.002$] and the offset-synchronous cosignal condition, $t(11)=3.81$, $p<0.005$. Identification thresholds did not differ significantly in the onset-synchronous cosignal and offset-synchronous cosignal conditions, $t(11) = 0.61$, $p > 0.25$.

C. Discussion

Significant CMPs were observed for both onset-synchronous and offset-synchronous stimuli. This indicates that synchrony either at the beginning or the end of a target signal embedded in a cosignal can provide a sufficient basis for perceptual integration but that neither onset-synchrony nor offset-synchrony is a necessary condition. The results of experiment 1 showed that perceptual integration of the sort underlying CMP does not occur when neither the onsets nor offsets of the target signal and cosignal are synchronous. With respect to the previous literature, this pattern offers one insight and creates one discrepancy.

The insight concerns the question of whether the effect of asynchronous onsets observed in studies of phonetic classification can be attributed to perceptual grouping or whether it results from perceptual adaptation (e.g., Darwin and Sutherland, 1984; Roberts and Moore, 1991). Previous studies of onset asynchrony have examined vowel (and pitch) identification in which an ''extraneous sound'' begins simultaneously with or in advance of some acoustic complex to be identified. The effect of the extraneous sound on identification of the complex typically decreases when the sound begins in advance of the complex. This finding can be explained by a perceptual grouping mechanism that integrates synchronous acoustic energy across the spectrum. Such a grouping mechanism receives independent support from studies of the effect of onset synchrony in auditory streaming (Bregman and Pinker, 1978). However, perceptual adaptation provides an alternative explanation of the effect of onset asynchrony in vowel and pitch identification. On this account, the early portion of the extraneous sound produces perceptual adaptation that reduces the contribution of the later portion of the extraneous sound to identification of the acoustic complex to which it is added. Perceptual adaptation has a well-established physiological basis (Kiang *et al.*, 1965) and it has been demonstrated in vowel identification studies through the phonetic classification of auditory afterimages (Summerfield *et al.*, 1984). Accordingly, perceptual grouping and perceptual adaptation constitute rival, though nonexclusive, accounts of why onset asynchrony reduces the contribution of an extraneous sound to the identification of an acoustic complex.

Perceptual adaptation cannot be the basis of CMP be-

cause the paradigm involves comparison of exactly the same target signal, with and without a cosignal. The addition of the spectrally distant cosignal would not affect perceptual adaptation in the spectral region of the target signal. The results of the present experiment show CMP for the onset-synchronous (but offset-asynchronous) stimuli, while the fringing stimuli of the preceding experiment (in which neither onsets nor offsets were synchronous) did not show CMP. As noted above, this indicates that onset-synchrony is a sufficient acoustic basis for the kind of perceptual grouping that underlies the CMP effect. Therefore, the present results demonstrate that simultaneous onsets can form the basis for at least one kind of perceptual integration.

The discrepancy created by the current findings is that onset synchrony and offset synchrony produced CMP effects of indistinguishable magnitude whereas previous research using identification paradigms has shown that disrupting onset synchrony caused a greater decrease in the contribution of the extraneous sound than did disrupting offset synchrony (e.g., Darwin, 1984a; Roberts and Moore, 1991). Before addressing this discrepancy at a conceptual level, an important difference should be noted in the arrangement of the parts of the stimulus in the current experiment as compared to earlier research that has looked at the role of asynchrony in perceptual integration.

III. EXPERIMENT 3: CMP WITH TEMPORALLY LEADING OR LAGGING TARGET SIGNALS

The temporal patterns within the stimuli used in experiment 2 were chosen to change single dimensions of the fringing stimuli used in experiment 1. As such, asynchrony was created by having the duration of the high-frequency cosignal exceed that of the distinctive, lower-frequency target signal. Previous research has taken the opposite tack and has used distinctive signals of greater duration than the acoustic complexes into which they were to be integrated (e.g., Darwin, 1984a; Roberts and Moore, 1991). The present study takes this latter approach: Onset asynchrony (with offset synchrony) is created by having the onset of the target signal lead the onset of the cosignal by 40 ms. Offset asynchrony (with onset synchrony) is created by having the offset of the target signal lag the offset of the cosignal by 40 ms. Identification thresholds for both of these conditions were compared to those for identifying the target signal alone.

A. Method

1. Subjects

Fifteen subjects from the same population as the previous study participated in the experiment. None of them had participated in either of the previous studies. Three subjects failed to meet the criterion for inclusion and were dismissed after the first six runs.

2. Stimuli, procedure, and design

The signals were the same as in the previous two experiments, except that they were 80 ms in duration (including onset and offset ramps) as opposed to the 40 ms used previously. The cosignals were the same as before, except they were now 40 ms in duration. In the onset-synchronous cosignal condition, both the signal and cosignal began 420 ms into the masker; the cosignal ended 40 ms later and the signal ended 80 ms later. In the offset-synchronous cosignal condition, the signal began 380 ms into the masker, the cosignal began 40 ms later. Both ended 460 ms into the masker. The procedure and design were the same as in the preceding two experiments.

TABLE III. Results of experiment 3. Mean signal level (dB SPL) at identification threshold for target signals with onset-synchronous cosignals, offset-synchronous cosignals, and no cosignals. Target signals are 80 ms and cosignals are 40 ms.

Subject number	Onset-synchronous cosignal	Offset-synchronous cosignal	No cosignal
1	52.1	54.9	54.9
2	55.5	55.2	56.7
3	51.2	52.6	54.7
4	53.5	55.6	53.4
5	56.6	57.2	57.0
6	52.0	54.3	54.9
7	52.8	53.3	55.3
8	51.5	54.3	52.5
9	52.6	55.0	53.7
10	52.8	54.6	55.6
11	52.6	55.0	53.9
12	55.8	57.4	54.0
Mean	53.2 (1.8)	55.0 (1.4)	54.7 (1.3)

B. Results

Table III shows the mean signal level at threshold in the three experimental conditions for individual subjects as well as the overall means and standard deviations. Analysis of variance showed that performance in the three conditions differed significantly, $F(2,22)=9.6,\ p<0.005$. Identification thresholds were lower in the onset-synchronous cosignal condition than in both the offset-synchronous cosignal condition $[t(11)=4.04,\ p<0.001]$ and the no-cosignal condition, $t(11)=3.5,\ p<0.01$. Identification thresholds did not differ significantly in the offset-synchronous cosignal and no-cosignal conditions, $t(11)=0.54,\ p>0.25$.

C. Discussion

The results of the experiment show that onset synchrony makes a greater contribution to CMP than does offset synchrony. A significant CMP effect was observed when the target signal and cosignal were onset synchronous but offset asynchronous. No CMP effect was observed when the target signal and cosignal were offset synchronous but onset asynchronous. This finding is consistent with previous research showing that onset asynchrony causes a greater reduction in perceptual integration than does offset asynchrony (e.g., Darwin, 1984a; Roberts and Moore, 1991). This shows that synchrony of onsets and offsets has a consistent effect on perceptual integration as studied by identification of both suprathreshold and threshold-level complex sounds.

The present finding concerning the relative importance of onset and offset synchrony differs from the pattern found

in experiment 2. It appears that perceptual integration of the target signal into the cosignal depends not only on synchrony of onsets and offsets, but also on whether the target signal leads the cosignal, or the cosignal leads the target signal; the former disrupts CMP while the latter does not. This difference indicates that the target signal and cosignal do not contribute symmetrically to perceptual integration as measured in the CMP paradigm, in that a leading target signal is perceptually segregated from the cosignal while a leading cosignal is not perceptually segregated from the target signal. This asymmetry could be due to a number of factors: (1) The target signal conveys the distinctive information necessary to perform the identification, so listeners likely focus more attention in the frequency region of the target signal than in that of the cosignal. (2) The target signal is close to its masked threshold, but the cosignal is not. (3) The target signal has a narrow bandwidth and is at a relatively low frequency, while the cosignal has a broader bandwidth and is at a higher frequency. Additional studies are required to determine the extent to which any of these factors are responsible for the asymmetry in the roles of the target signal and cosignal in producing CMP.

IV. GENERAL DISCUSSION

The results of the three experiments show that identification thresholds for masked noise bands can be reduced by the addition of acoustic energy that is spectrally well separated from the target signal, a phenomenon that has been dubbed ''coherence masking protection'' (CMP) in studies of speech perception (Gordon, 1997). Experiment 1 showed that CMP in nonspeech stimuli was influenced by the temporal arrangement of the components of the sound in a manner that closely matched that observed with speech stimuli. The results of experiments 2 and 3 tease apart the contribution of the synchrony of stimulus onsets and offsets to CMP. Below, the implications of these results are discussed with respect to two issues: specialized versus general processes in speech perception and possible mechanisms underlying CMP.

A. Specialized versus general processes in speech perception

Gordon (1997) demonstrated CMP in the perception of vowels and showed that it could be disrupted by certain asynchronies between the first formant and higher formants. This finding could be attributed either to specialized mechanisms for phonetic perception that exploit temporal regularities inherent in the production of speech or to general mechanisms of auditory perception that exploit temporal regularities that are often characteristic of events in the world. Experiment 1 of the current paper showed that CMP in the perception of nonspeech sounds was influenced by the temporal relations between low-frequency and high-frequency energy in a manner that exactly matched that observed for the temporal relation between the first formant and higher formants in experiment 3 of Gordon (1997). While it is possible that different mechanisms underlie the effect in speech and nonspeech sounds, the more parsimonious explanation is that CMP in speech sounds (and nonspeech sounds) emerges from the operation of general mechanisms of perceptual integration that can be applied to sounds irrespective of their origin.

The contention that phonetic perception uses specialized mechanisms arose early in the study of speech perception (Liberman, 1982 for a review), and has continued to have ardent supporters (e.g., Remez *et al.*, 1994). Over the last 15 years, a critical arena in which this contention has been debated is the integration of acoustic energy into coherent percepts. Support for the view that speech makes use of specialized mechanisms for perceptual integration has been claimed based on the phenomenon of duplex perception (e.g., Liberman *et al.*, 1981; Whalen and Liberman, 1987; cf. Bailey and Herrmann, 1993) and on the ability to recognize sine-wave replicas of speech (Remez *et al.*, 1994). Support for the view that speech makes use of general mechanisms for perceptual integration has come from studies showing that the phonetic contribution of acoustic energy is strongly influenced by factors (synchrony, harmonic relations and streaming) that contribute to perceptual integration in nonspeech sounds (Ciocca and Bregman, 1989; Darwin, 1984a), and by studies showing nonspeech stimuli can show duplex perception (Fowler and Rosenblum, 1990). Bregman (1990) has presented a two-stage model of perceptual integration of acoustic energy in speech perception; the first stage uses general processes of auditory segregation while the second stage uses speech-specific schemas.

CMP is an effect on the identification thresholds of fairly simple masked signals. Historically, detection thresholds for simple masked signals formed the basis of the critical-band model and were assumed to reflect very early stages of auditory processing, in part because of the simplicity of the tasks and in part because of the match between psychoacoustic data and recordings in the peripheral nervous system (Moore, 1993). Phenomena such as comodulation masking release (CMR; Hall *et al.*, 1984) have shown that the critical-band model cannot account completely for psychoacoustic data on masked thresholds. To some extent this means that effects on masked thresholds cannot necessarily be attributed to early stages of perceptual processing based on the relationship between psychoacoustic and neurophysiological data. However, there is still good reason for believing that effects such as CMR and CMP emerge from basic processes of perceptual organization and not from strategic decision processes. In these paradigms, listeners are presented with a simple task in which they are given considerable practice with feedback, features that could be expected to optimize strategic decision processes. However, performance is improved by the addition of acoustic energy that does not in a straightforward way increase the signal-to-noise ratio in the spectral region of the target signal, but which does provide a basis for perceptual reorganization of the stimulus. This suggests that CMP should be attributed to an early stage of perceptual processing like the first stage of perceptual segregation/integration proposed by Bregman (1990).

B. Mechanisms of CMP

Gordon (1997) discusses two distinct models of the CMP phenomenon, both based on ideas developed in the CMR literature. The results of experiments 2 and 3 of the current paper provide challenges to both these models.

The first model elaborated by Gordon (1997), called ''peak listening,'' is based on ''listening in the valleys'' or ''dip listening'' accounts of CMR (Buus, 1985) which state that listeners use changes in energy of the comodulated flanking bands to locate energy minima in the on-signal masking band, thus finding the optimal signal-to-masker ratio. In the peak-listening model of CMP, the clearly audible cosignal (or higher formants) are seen as marking the temporal location of the target signal (or first formant) in the masking noise, thereby indicating the temporal location of the optimal signal-to-masker ratio. Because a fringing cosignal marks the target signal but no CMP is observed, Gordon (1997) considered a modified peak-listening model in which the signal-to-masker ratio is averaged over the interval in which the cosignal is present. With fringing cosignals, this interval includes time when the target signal is not on, thereby eliminating the CMP effect. However, this modified peak-listening model is challenged by the present results. The onset-synchronous and offset-synchronous stimuli of experiment 2 produced CMPs of 2.9 and 2.4 dB, respectively, while the synchronous stimuli of experiment 1 produced a CMP of 2.4 dB. The onset-synchronous and offset-synchronous stimuli include intervals in which the cosignal is on but the target is off; therefore, computing signal-to-masker ratios over the interval of the cosignal should be less effective than it would be with synchronous cosignals. Thus the finding in experiment 2 that CMP occurs for onset-synchronous and offset-synchronous stimuli appears to undercut the modified peak-listening model.

The second model elaborated by Gordon (1997) involves two processes, both of which build on prominent constructs in the study of the perception of complex sounds. The first is auditory grouping as it has been related to CMR (Hall and Grose, 1990) and the second is comparative perceptual evaluation, as it has been developed in profile analysis (Green, 1988). The auditory grouping process responds to the simultaneous energy changes at the frequencies of the target signal and cosignal that occur when the onset and offset of the signals occur at the same time. Given the very brief signals (40 and 80 ms) used in the present experiments, these energy changes occur at a rate where substantial CMR is observed with periodic modulation of masking bands (Hall and Haggard, 1983). Because the CMR paradigm involves comodulation over a relatively long interval, the masking bands could group auditorily based on many instances of simultaneous changes in energy. In contrast, such grouping in CMP could only be based on the simultaneous energy changes that occur due to the onsets or offsets of the target signal and cosignal. The perceptual comparison process is engaged by the perceptually coherent object and enables listeners to be more sensitive to the identification of the target signal because the cosignal provides a concurrent perceptual basis for estimating the expected energy level at the frequencies of the target signal. No audible comparative basis is available when there is no cosignal, the target signal must be identified by comparing energy at the two target-signal frequencies or by comparing energy at those frequencies to the memory of the energy level earlier in the masker. The combination of the target signal and the cosignal into a coherent perceptual object could potentially allow listeners to identify the stimulus based on timbre, the distribution of energy across the spectrum, as well as on pitch.

The results of experiments 2 and 3 suggest that auditory grouping as measured by CMP can be based on synchrony of specific energy changes in different parts of the spectrum. Experiment 2 shows this for both the onset and offset of energy. Experiment 3 shows that auditory grouping can be based on the synchrony of onsets, but shows that synchrony of offsets is not sufficient to produce grouping if the onset of the target signal precedes that of the cosignal. The contrast between the results of experiments 2 and 3 indicates that the target signal and cosignal do not contribute in an equivalent manner to the formation of an auditory object. The discussion of experiment 3 indicates several factors—attentional focus, masking, frequency, and bandwidth of the signals—that might explain this difference. Exploration of these factors may provide further insight into the processes that integrate acoustic energy into coherent auditory objects.

ACKNOWLEDGMENTS

The research reported in this paper was supported by grants from the Stephenson Faculty Fund and by the University Research Council of the University of North Carolina at Chapel Hill. I would like to thank Tom Carrell, John Grose, Joe Hall, and Bob Peters for helpful discussions of this research. I would also like to thank Tim Fiscus and Shelley Poovey for assistance in testing subjects. Direct correspondence to Peter C. Gordon, Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3270.

[1]My thanks to John Grose for making the noise bands.

Bailey, P. J., and Herrman, P. (**1993**). ''A reexamination of duplex perception evoked by intensity differences,'' Percept. Psychophys. **54**, 20–32.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Bregman, A. S., and Pinker, S. (**1978**). ''Auditory streaming and the building of timbre,'' Can. J. Psychol. **32**, 19–31.

Buus, S. (**1985**). ''Release from masking caused by envelope fluctuations,'' J. Acoust. Soc. Am. **78**, 1958–1965.

Ciocca, V., and Bregman, A. S. (**1989**). ''The effects of auditory streaming on duplex perception,'' Percept. Psychophys. **46**, 39–48.

Darwin, C. J. (**1981**). ''Perceptual grouping of speech components differing in fundamental frequency and onset-time,'' Q. J. Exp. Psychol. **33A**, 185–207.

Darwin, C. J. (**1984a**). ''Auditory processing and speech perception,'' in *Attention and Performance X: Control of Language Processes*, edited by H. Bouma and D. G. Bouwhuis (Erlbaum, Hillsdale, NJ), pp. 197–210.

Darwin, C. J. (**1984b**). ''Perceiving vowels in the presence of another sound: Constraints on formant perception,'' J. Acoust. Soc. Am. **76**, 1636–1647.

Darwin, C. J., and Gardner, R. B. (**1986**). ''Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality,'' J. Acoust. Soc. Am. **79**, 838–845.

Darwin, C. J., and Sutherland, N. S. (**1984**). ''Grouping frequency components of vowels: When is a harmonic not a harmonic?'' Q. J. Exp. Psychol. **36A**, 193–208.

Fowler, C. A., and Rosenblum, L. D. (**1990**). ''Duplex perception: A comparison of monosyllables and slamming doors,'' J. Exp. Psychol: Human Perception and Performance **16**, 742–754.
Gordon, P. C. (**1997**). ''Coherence masking protection in speech sounds: The role of formant synchrony,'' Percept. Psychophys. **59**, 232–242.
Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U.P., New York).
Hall, J. W., III, and Grose, J. H. (**1988**). ''Comodulation masking release: Evidence for multiple cues,'' J. Acoust. Soc. Am. **84**, 1669–1675.
Hall, J. W., III, and Grose, J. H. (**1990**). ''Comodulation masking release and auditory grouping,'' J. Acoust. Soc. Am. **88**, 119–125.
Hall, J. W., III, and Haggard, M. P. (**1983**). ''Co-modulation—A principle for auditory pattern analysis in speech,'' Proceedings of the 11th ICA **4**, 69–71.
Hall, J. W., III, Haggard, M. P., and Fernandes, M. A. (**1984**). ''Detection in noise by spectrotemporal pattern analysis,'' J. Acoust. Soc. Am. **76**, 50–56.
Hukin, R. W., and Darwin, C. J. (**1995**). ''Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification,'' Percept. Psychophys. **57**, 191–196.
Kiang, N. Y.-S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT, Cambridge, MA).
Liberman, A. M. (**1982**). ''On finding that speech is special,'' Am. Psychol. **37**, 148–167.
Liberman, A. M., and Mattingly, I. G. (**1989**). ''A specialization for speech perception,'' Science **243**, 489–494.
Liberman, A. M., Isenberg, D., and Rackerd, B. (**1981**). ''Duplex perception for stop consonants: Evidence for a phonetic mode,'' Percept. Psychophys. **30**, 133–143.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). ''Perception of the speech code,'' Psychol. Rev. **74**, 431–460.
Mann, V. A., and Liberman, A. M. (**1983**). ''Some differences between phonetic and auditory modes of perception,'' Cognition **14**, 211–235.
Moore, B. C. J. (**1993**). ''Frequency analysis and pitch perception'' in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York).
Pisoni, D. B. (**1977**). ''Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops,'' J. Acoust. Soc. Am. **61**, 1352–1361.
Remez, R. E. (**1980**). ''Susceptibility of a stop consonant to adaptation on a speech-nonspeech continuum: Further evidence against feature detectors in speech perception,'' Percept. Psychophys. **27**, 17–23.
Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (**1994**). ''On the perceptual organization of speech,'' Psychol. Rev. **101**, 129–156.
Roberts, B., and Moore, B. C. J. (**1990**). ''The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels,'' J. Acoust. Soc. Am. **88**, 2571–2583.
Roberts, B., and Moore, B. C. J. (**1991**). ''The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels under conditions of asynchrony,'' J. Acoust. Soc. Am. **89**, 2922–2932.
Summerfield, Q., Haggard, M. P., Foster, J., and Gray, S. (**1984**). ''Perceiving vowels from uniform spectra: phonetic exploration of an auditory after-effect,'' Percept. Psychophys. **35**, 203–213.
Whalen, D. H., and Liberman, A. M. (**1987**). ''Speech perception takes precedence over nonspeech perception,'' Science **237**, 169–171.