Enhancement, adaptation, and the binaural system

Maja Šerman, Catherine Semal, and Laurent Demany

Laboratoire Mouvement, Adaptation, Cognition (UMR CNRS 5227), BP 63, Université Bordeaux 2, 146 Rue Leo Saignat, F-33076 Bordeaux, France

(Received 7 March 2007; revised 29 February 2008; accepted 3 March 2008)

In a test sound consisting of a burst of pink noise, an arbitrarily selected target frequency band can be “enhanced” by the previous presentation of a similar noise with a spectral notch in the target frequency region. As a result of the enhancement, the test sound evokes a pitch sensation corresponding to the pitch of the target band. Here, a pitch comparison task was used to assess enhancement. In the first experiment, a stronger enhancement effect was found when the test sound and its precursor had the same interaural time difference (ITD) than when they had opposite ITDs. Two subsequent experiments were concerned with the audibility of an instance of dichotic pitch in binaural test sounds preceded by precursors. They showed that it is possible to enhance a frequency region on the sole basis of ITD manipulations, using spectrally identical test sounds and precursors. However, the observed effects were small. A major goal of this study was to test the hypothesis that enhancement originates at least in part from neural adaptation processes taking place at a central level of the auditory system. The data failed to provide strong support for this hypothesis.

© 2008 Acoustical Society of America. [DOI: 10.1121/1.2902177]

PACS number(s): 43.66.Mk, 43.66.Pn [RYL] Pages: 4412–4420

I. INTRODUCTION

A. Monaural enhancement

An auditory “enhancement” phenomenon occurs when a sound with a given power spectrum (A) is followed on the same ear by a second sound with a power spectrum consisting of A plus some additional frequency content, B. The presentation of the first sound (the “precursor”) appears to enhance the detectability, or the perceptual salience, of B in the second sound (the “test” sound). This is observable when, for instance, A is a sum of harmonics and B is another harmonic (Viemeister, 1980; Hartmann and Goupell, 2006), or when A is a broadband noise with spectral valleys filled by B (Wilson, 1970). Objective evidence for enhancement has been obtained in various experimental paradigms: simultaneous masking (e.g., Viemeister, 1980; Carlyon, 1989), forward masking (Viemeister and Bacon, 1982), pitch matching (Hartmann and Goupell, 2006), and vowel identification (Summerfield et al., 1984, 1987).

What is the origin of these enhancement effects? It has often been hypothesized that they stem from neural adaptation at a relatively peripheral level of the auditory system (Viemeister, 1980; Summerfield et al., 1987; Hicks and Bacon, 1992). The neural response to the components of the test sound that were already present in the precursor could be reduced following the presentation of the precursor, thus increasing the relative prominence of the novel part of the test sound. Palmer et al. (1995) found physiological support for this hypothesis in the auditory nerve of guinea pigs. However, this “simple” adaptation cannot account for the psychophysical fact that enhancing a tone increases its forward masking of a subsequent tone (Viemeister and Bacon, 1982).

To account for the latter observation, it has been supposed that the precursor adapts (i.e., reduces) the ability of the corresponding part of the test sound to “suppress” (or inhibit) the novel part of the test sound (Viemeister, 1980; Viemeister and Bacon, 1982). However, the physiological study of Palmer et al. (1995) did not provide support for that idea at the auditory nerve level. Moreover, the adaptation-of-suppression hypothesis is at odds with psychophysical results reported by Wright et al. (1993). Currently, therefore, the contribution of neural adaptation to the perceptual enhancement effects described above is not clear. Carlyon (1989) has argued that their main source is not adaptation. His alternative hypothesis and other ones will be considered later in this paper.

B. Binaural enhancement?

No significant enhancement occurs when the test sound and its precursor are presented to opposite ears (Viemeister, 1980; Summerfield et al., 1984, 1987; Carlyon, 1989; Kidd and Wright, 1994). As pointed out by Kidd and Wright (1994), this does not imply that the mechanism of enhancement is located at a peripheral level of the auditory system, below the level at which the two monaural pathways converge: It could be that the mechanism of enhancement has a central site and is sensitive to interaural relations. More specifically, enhancement phenomena might largely stem from a central form of neural adaptation (CNA). In the past few years, physiologists have uncovered previously unknown forms of adaptation in the auditory system. Ulansovskaya et al. (2003, 2004) described a highly stimulus-specific form of CNA in the primary auditory cortex of cats. In addition, McAlpine et al. (2000), Malone et al. (2002), and Furukawa et al. (2005) suggested that in the inferior colliculus or the auditory cortex of mammals, where neurons are (broadly)
tuned to specific interaural phase differences (IPDs) in a given frequency region, the neural response to a given IPD is extremely dependent on the IPDs presented in the recent past. The authors of these investigations emphasized that the CNA revealed by their work increased the contrast between the neural representations of slightly different sounds heard in succession.

In the domain of binaural hearing, some psychophysical phenomena have already been interpreted as a consequence of CNA. A number of authors have reported “repulsive” aftereffects in sound localization or lateralization. Under certain conditions, after the presentation of a sound with some interaural time difference (ITD), the judged location of a second sound with the same spectrum but a different ITD is shifted in the direction opposite to the location of the previous sound (Thurlow and Jack, 1973; Kashino and Nishida, 1998; Phillips and Hall, 2005; Vigneault-MacLean et al., 2007). This might be a genuinely sensory phenomenon due to CNA, although an alternative possibility is a change in response criterion (i.e., a nonsensory bias). In the same vein, Kashino (1998) and Getzmann (2004) reported that the ability to detect a difference between the spatial positions of two successive sounds X and Y can be improved by the presentation of a precursor identical to X. This might again be due to CNA, although other interpretations are possible.

In the study reported here, we attempted to support the hypothesis that one source of enhancement phenomena is CNA. Given that if CNA does exist and has perceptual consequences, it should logically manifest itself in enhancement phenomena, this study was more generally a search for psychophysical correlates of CNA.

II. EXPERIMENT 1

A. Rationale

The goal of this initial experiment was to determine if the binaural system can have an influence on the perceptual enhancement of spectral energy in a given frequency region by a precursor with no energy in that frequency region. In one experimental condition, the test sounds and the precursors were binaural stimuli with the same ITD (either +600 or −600 μs). In another experimental condition, the precursor and test sounds were also binaural stimuli but had opposite ITDs (+600 and −600 μs). There was no difference at all between these two conditions with respect to the monaural components of the stimuli. Thus, if enhancement is of a purely peripheral origin, one would have expected it to have the same magnitude in the two conditions. If, on the other hand, enhancement is partially caused by an ITD-specific (as well as frequency-specific) form of neural adaptation, then its magnitude ought to have been larger in the “same-ITD” condition than in the “opposite-ITD” condition. Our index of enhancement magnitude was the salience of the pitch sensation evoked by the frequency band which was missing in the precursor but present in the test sound. The audibility of the corresponding pitch was assessed by means of a pitch comparison task.

**FIG. 1.** Schematic illustration of the relationships between precursor, test, and probe sounds in each experiment. The vertical dimension of each panel represents frequency. Differences in binaural lateralization are denoted by a contrast between black and gray areas or between continuous and dotted lines. In experiment 1, the precursor sound was a two-octave band of noise with a 1/3-octave notch, the test sound was the same noise band without the notch, and the probe sound was a 1/3-octave noise band, slightly lower or higher in frequency than the previous notch; in two out of four conditions the precursor differed in lateralization from the two following sounds. In experiment 2, the precursor sound was a two-octave band of noise, the test sound was identical to the precursor except for an ITD change in a 1/6-octave target band, and the probe sound was a 1/6-octave noise band, slightly lower or higher in frequency than the target band; on a given trial, the precursor could be presented four times, presented only once, or not presented. In experiment 3, the precursor sound was again a two-octave band of noise, the test sound was a sum of nine pure tones 1/4-octave apart, and the probe sound was a single pure tone; eight components of the test sound had the same ITD as the precursor sound; the remaining component, with a different ITD, was the target stimulus; the probe sound was matched in frequency with either the target stimulus or some other component of the test sound; on a given trial, the precursor could be either long (1200 ms), short (300 ms), or not presented.

B. Method

1. Stimuli and task

On each trial, subjects were successively presented with (1) a precursor sound (total duration: 1000 ms), (2) a test sound (600 ms), and (3) a probe sound (400 ms). These three sounds—as well as all those used in our subsequent experiments—were gated on and off with 5 ms cosinusoidal amplitude ramps. There was no silent interval (ISI) between the offset ramp of the precursor and the onset ramp of the test sound. A 1200 ms ISI separated the test sound from the probe.

The spectral relations of the precursor, test, and probe sounds are depicted in the upper panel of Fig. 1. Each test sound was essentially a band of pink noise with a width of two octaves. It was generated by adding together 121 synchronous pure tones with equal amplitudes and a frequency spacing of 1/60 octave. The frequencies of the lowest and highest tones were 200 and 800 Hz. Each tone had a nominal
sound pressure level (SPL) of 49 dB, resulting in an overall level of 70 dB SPL. Given that their component tones were very close in frequency (and had random initial phases, see below), the test sounds were perceived as nothing but noise in the absence of a precursor.

The power spectrum of the precursor sounds was identical to that of the test sound, except for the omission of 21 adjacent tones, forming a 1/3-octave frequency band, among the 121 tones making up the test sounds. The spectral position of the omitted frequency band (called the “target band” hereafter) varied randomly from trial to trial without any constraint.

The probe sound presented on a given trial consisted of 21 tones 1/60-octave apart. These tones formed a 1/3-octave band which was identical to the target band except for an overall frequency shift of plus or minus 1/12 octave. The subject’s task was to identify the direction of the frequency shift, which varied randomly from trial to trial. Due to the enhancement phenomenon, this pitch comparison task was relatively easy when, for instance, the precursor and test sounds were presented monaurally to the same ear. In contrast, we noted informally that the task became very hard when the test sound was removed.

The component tones of the precursor, test, and probe sounds had new random initial phases on each trial. Within a trial, moreover, the initial phase of a tone produced more than once varied randomly from presentation to presentation.

2. Conditions

Subjects were tested in four conditions. In the “monaural” condition, the test sound and its precursor were presented monaurally to the same ear. In the “contralateral” condition, the precursor and test sounds were also monaural stimuli but were presented to opposite ears. In the same-ITD condition, the precursor and test sounds were presented binaurally, with an ITD of 600 μs favoring the same ear; the ITD did not affect the onset and offsets of the stimuli, which were interaurally synchronous. In the opposite-ITD condition, finally, a 600 μs ongoing ITD was also present in the stimuli but favored opposite ears in the precursor and test sounds.

For each condition, the precursor sound was lateralized on the left or right in different subconditions. This created a total of eight subconditions, run in separate blocks of trials. With respect to lateralization, the probe sound was always identical to the test sound.

3. Procedure

The stimuli were digitally generated with a sampling rate of 20 kHz and a 24 bit amplitude quantization. They were presented to the subject through Sennheiser HD265 headphones in a double-walled sound-attenuating booth (Gisol, Bordeaux). Responses (“up” or “down”) were given by means of mouse clicks on two labeled virtual buttons and were immediately followed by visual feedback. The formal experiment (following a small number of training sessions) consisted of eight sessions, each including eight blocks of trials—one block for each of the eight subconditions, run in a random order. We wished to process only the data collected on trials in which the target band had been located in the middle part of the test sound’s spectrum, more precisely, the frequency region defined by the 61 most central tones (283–566 Hz). Thus, it was decided to terminate every block of trials after a fixed number of trials fulfilling this requirement; the number in question was 30. Overall, therefore, the analyzed data were collected in 30 × 8 × 2 = 480 trials per condition (each condition including two subconditions) and subject.

4. Subjects

Four listeners aged between 21 and 53 years (S1, S2, S3, and S4) were tested. All had normal pure-tone audiograms except for S1 who had a 25 dB dip in her left-ear audiogram between about 150 and 250 Hz. The results for this subject did not differ in trend from those of other subjects, neither in the present experiment nor in the subsequent experiments reported here. S1 and S2 were two of the authors.

Given that the experimental task consisted of ordinal pitch comparisons (up and down judgments), it was useful to check that the performance of each subject would not be limited by difficulty in making comparisons of that kind (Semal and Demany, 2006). This was checked by means of a preliminary test including 200 trials in which two successive sounds similar to the probe sounds used subsequently were presented. These two sounds differed by plus or minus (at random) 1/12 octave and the subject had to identify the direction of the corresponding change. S1, S2, S3, and S4 proved to be able to perform perfectly (100% correct) in this test.

C. Results and discussion

Figure 2 displays the percentage of correct responses obtained for each subject in each condition. It can be seen that the four subjects behaved similarly. As expected from previous research (e.g., Summerfield et al., 1987), performance was much better in the monaural condition (mean score: 79.9% correct) than in the contralateral condition (54.6% correct). In the latter case, performance was close to chance. In the same-ITD condition (mean score: 83.6% correct), subjects were slightly more efficient than in the mon-
aural condition. In the opposite-ITD condition (mean score: 70.4% correct), performance was well above chance but definitely poorer than that in the same-ITD condition.

The latter finding is the main outcome. It is at odds with results obtained by Kidd and Wright (1994) in an experiment which had basically the same goal as ours but was quite different methodologically. Kidd and Wright (1994) used a simultaneous masking paradigm in which the signal was a monaural tone burst with a fixed frequency (1000 Hz) and a very short duration (4 ms); the masker, a notched noise, was presented either ipsilaterally, contralaterally, or binaurally with an interaural level difference favoring the contralateral ear. Kidd and Wright’s (1994) data led them to suggest that enhancement is essentially a monaural, and thus a peripheral, phenomenon. Contrary to Kidd and Wright’s (1994) conclusion, our data indicate that enhancement is significantly dependent on binaural processing. It is presently difficult to see precisely why the two experiments led to discrepant conclusions because their methodologies differed in many ways.

The fact that subjects were more successful in the same-ITD condition than in the opposite-ITD condition shows unambiguously that central factors play a role in enhancement. It is possible that the central influence observed here had something to do with “attention:” The ITD change occurring in the opposite-ITD condition might have “distracted” the subject, perhaps strongly enough to produce the obtained 13.2% drop in performance relative to that in the same-ITD condition. However, this does not seem very likely since, on each trial run in the opposite-ITD condition, the change in ITD was predictable and could be anticipated. A more appealing interpretation of the advantage obtained in the Same-ITD condition is that this advantage originates from an ITD-specific CNA.

The fact that performance was better in the opposite-ITD condition than that in the contralateral condition may also have a central origin since the change in subjective lateralization produced in the opposite-ITD condition was smaller than the change produced in the contralateral condition. However, on the other hand, this result was of course expected under the hypothesis that one source of enhancement is peripheral. The latter hypothesis is neither ruled out nor clearly supported by our data. Some support for it can be found in the slight increase of performance from the monaural condition to the same-ITD condition since two ears offer a statistical advantage over a single ear with regard to the production of a significant monaural enhancement effect.

III. EXPERIMENT 2

A. Rationale

If one source of enhancement phenomena is CNA and if CNA can be ITD specific (as suggested by the physiological studies cited above), then it should be possible to produce an enhancement effect based entirely on ITD manipulations, using spectrally identical precursor and test stimuli. We attempted to do so in experiment 2. The test sounds were similar to those employed in the same-ITD and opposite-ITD conditions of experiment 1 except that within each test sound, a small set of adjacent tones, forming a narrow frequency band (the target band), had an ITD that differed from the ITD of the other tones (the “background” tones). The target band was thus liable to evoke a pitch sensation corresponding approximately to the pitch of its center frequency; previous studies on this instance of “dichotic pitch” were reported by Dougherty et al. (1998) and Akeroyd and Summerfield (2000). In the absence of a precursor, however, this pitch was not easily heard. We attempted to enhance it by a precursor consisting of the same tones as those forming the test sound but in which all tones had the ITD of the subsequent background tones, as illustrated in the middle panel of Fig. 1. The precursor was intended to adapt specifically the background components of the test sound and to increase in this way the salience of the target band’s pitch. We reasoned that if the benefit of the precursor was due to CNA, one would expect to obtain a larger benefit from repeating the precursor several times before the test sound than from presenting it only once. Indeed, the physiological studies of Ul- anovsky et al. (2003, 2004) indicate that even though a single stimulus repetition is sufficient to observe a substantial amount of CNA, the effect increases with the number of repetitions; CNA appears to be a cumulative process. This led us to use three conditions in which the precursor was, respectively, presented four times, presented only once, and not presented.

B. Method

1. Stimuli and task

In each test sound, consisting again of 121 synchronous pure tones spaced by 1/60-octave intervals and ranging from 200 to 800 Hz, 11 adjacent tones formed the target band (width: 1/6 octave). These tones had an ITD (ITDtarget) that differed from the ITD of the other tones (ITDbg). One of these two “ITDs” was in fact equal to zero; the other ITD favored the right ear and was subject dependent (more details in the next section); it did not affect the onsets and offsets of the stimuli. The spectral position of the target band varied randomly from trial to trial without any constraint. As before, each tone had a nominal SPL of 49 dB and a random initial phase (renewed from trial to trial) at a given ear. The test sounds had a total duration of 400 ms.

On each trial, after a 1200 ms ISI, the test sound was followed by a 400 ms probe sound which was, as in experiment 1, a transposition of the target band at a frequency distance of 1/12 octave; the ITD of this probe was equal to ITDtarget. Again, the subject’s task was to identify the direction of the frequency shift, this direction varying randomly from trial to trial, and visual feedback was provided following each response.

On a given trial, the test sound was preceded by either zero, one, or four presentations of a 200 ms precursor sound. The precursor did not differ from the test sound with respect to the power spectrum, but the ITD of all its component tones was equal to ITDbg. There was a 200 ms ISI between the final (or single) precursor presentation and the test sound. When the precursor was repeated, its presentations were also separated by 200 ms ISIs.
2. Procedure and subjects

The three conditions (zero, one, or four precursor presentations) were run in separate blocks of 50 trials. In contrast to experiment 1, we did not restrict data analysis to the trials on which the target band had been located in the middle part of the test sound’s spectrum. Every block included 25 trials in which ITDtarget was zero while ITDbackground was not zero and 25 trials in which the opposite was true. Any trial of one of these two types was always followed and/or preceded by a trial of the other type. This was done in order to avoid possible benefits of across-trial CNA in the blocks without precursor. In such blocks, had ITDbackground been fixed, the background components of the test sound presented on a given trial could have advantageously served as an adaptor to the background components of the test sound presented in the next trial.

In the experiment proper, each condition was run in eight blocks of trials (400 trials overall) per subject. The order of conditions was randomized within sessions. The experiment proper was preceded by preliminary training sessions during which, for each subject, the size of the nonzero ITD was varied in order to find an ITD value that avoided floor and ceiling effects. Except for this variation of the nonzero ITD, the preliminary training sessions did not differ from the following formal sessions.

Five listeners were tested. Three of them (S1, S2, and S3) also served as subjects in experiment 1. The two new subjects (S5 and S6) were in their 20s and had normal audiograms at both ears. Both of them performed without any error the preliminary pitch comparison test described in Sec. II B 4. The nonzero ITD value used in the experiment proper was 150 μs for S1, 500 μs for S2, and 200 μs for S3, S5, and S6.

C. Results and discussion

The scores obtained within blocks of trials were submitted to a three-way analysis of variance (ANOVA) [subject × (number of precursors) × ITDtarget (zero versus nonzero)]. This ANOVA indicated that the number of precursors had a highly significant main effect \(F(2,210)=23.5, P<0.001\) and did not interact significantly with the subject factor \(F(8,210)=1.2, P=0.32\) nor with ITDtarget \(F(2,210)<1\). The main effect of ITDtarget was at the limit of statistical significance \(F(1,210)=4.1, P=0.04\) and there was a significant interaction between this factor and the subject factor \(F(4,210)=8.1, P<0.001\). Figure 3 shows the effect of the number of precursors, with performance collapsed across the ITDtarget variable. The mean percentage of correct responses was 72.3% when no precursor was presented, 79.3% for one precursor presentation, and 80.6% for four presentations. So, although presenting a precursor had a benefit, this benefit was essentially the same for one and four presentations. Overall, performance was slightly better when ITDtarget differed from 0 (mean percentage of correct responses: 78.5%) than when ITDtarget was equal to 0 (mean percentage: 76.4%); this trend is consistent with observations by Hartmann and Zhang (2003) concerning the audibility of the “Huggins pitch” (Cramer and Huggins, 1958) for low-frequency target bands.

We can conclude from these results that it is possible to produce a perceptual enhancement effect based entirely on ITD manipulations, without adding new spectral energy to the precursor. The existence of this effect is consistent with the hypothesis that some aspects of human auditory perception are under the influence of CNA. However, CNA (like other forms of neural adaptation) is supposed to be a cumulative process (see, e.g., Ulanoisky et al., 2003, 2004). Thus, the fact that we did not observe a significantly stronger enhancement after four presentations of the precursor than after a single one puts into question the role of CNA in the present experimental situation.

It must also be noted that the benefit of the precursor was only modest. In this respect, a natural conjecture is that a stronger benefit could have been found if the ISI separating the final (or single) precursor presentation from the test sound had been shorter than 200 ms. The reason why we chose this ISI rather than a much shorter one (or no ISI at all) is that, if the ISI had been very short, the benefit of the precursor would have been likely to arise in part from factors unrelated to CNA. Culling (2000) pointed out that, given the well-known “sluggishness” of the binaural system (e.g., Grantham and Wightman, 1978), a rapid ITD change in some frequency region is detectable not as an ITD change per se but rather as a decorrelation of the two monaural inputs in that frequency region, leading to the perception of the corresponding pitch. In the present experiment, we wanted to prevent subjects from using such a cue in order to identify the target band when a precursor preceded the test sound.

IV. EXPERIMENT 3

A. Rationale

In experiment 2, the precursor sounds had only a weak influence on performance even when they were repeated before the test sounds. This may have been due to their rather short duration (200 ms). We had thought that the repetition of a short precursor could be more efficient than only one presentation of a longer precursor because the binaural system is particularly sensitive to the onset of sounds (see, e.g.,
However, this may have been a wrong idea. In experiment 3, we used longer precursors with the hope of observing a stronger effect.

Experiment 3 was also motivated by the fact that, in experiment 2, the benefit of a precursor could, at least in theory, originate from factors unrelated to CNA. One such factor could be “timbral cuing.” The precursor indicated precisely to the subject, just before the test stimulus how the test stimulus, would sound if all its spectral components had the same ITD; the precursor thus provided a potentially useful timbral reference in auditory memory. Hence, this renders the ITD of the target. Since they were not resolvable by the auditory system, the precursor sounds were perceived as noise, whereas the test sounds were perceived as tonal chords. Therefore, when a test sound was preceded by a precursor, the background components of the test sound were not liable to be perceived as a repetition (or quasirepetition) of a portion of the precursor. In addition, since the test sound was separated from the precursor by a 100 ms silent ISI, the sequence was not likely to induce a “retrospective continuity illusion” (Carlyon et al., 2005).

The precursor sounds and test sounds had the same overall level, 70 dB SPL. Thus, whereas the component tones of the precursors had (as before) a SPL of 49 dB, the SPL of the test sounds’ component tones was 60.5 dB. Test sounds and probe tones had a duration of 200 ms. As mentioned above, the ISI between precursor and test sounds was 100 ms. The ISI between test and probe stimuli was 1200 ms.

On a given trial, the test sound was preceded by either a “long” precursor (1200 ms), a “short” precursor (300 ms), or no precursor at all. These three conditions were run in separate blocks of 40 trials (12 blocks for each condition in the experiment proper). Within each block, there were 20 trials in which ITDbackground = 0 and ITDtarget ≠ 0 (the right ear leading) and 20 trials in which the opposite was true. As in experiment 2 (and for the same reason), any trial of one of these two types was always followed and/or preceded by a trial of the other type. The order of conditions was randomized within sessions.

Five listeners were tested, among whom four (S1, S2, S3, and S4) had participated as subjects in at least one of the previous experiments. The fifth subject, S7, was in his 20s and had a normal audiogram at each ear. Following some preliminary sessions in which the nonzero ITD value was varied adaptively, this parameter was set to 110 μs for S1, 600 μs for S2, 200 μs for S3, 350 μs for S4, and 150 μs for S7.

B. Method

New test sounds were used, as well as a new task. As illustrated in the bottom panel of Fig. 1, each test sound was the sum of nine equal-amplitude pure tones, spaced by intervals of 1/60 octave and ranging in frequency from 200 to 800 Hz. The spacing of the tones was such that they were resolvable by the auditory system. The test sounds were thus perceived as tonal stimuli (chords). On each trial, eight components of the presented test sound had the same ITD (ITDbackground) and the remaining component—the target—had a different ITD (ITDtarget). The target tone was selected at random among the seven “inner” components (i.e., the components with frequencies differing from 200 and 800 Hz). The test sound was followed by a probe which was a single pure tone, equiprobably identical to the target tone or matched in frequency to some other inner component of the test sound (selected at random among the six candidates). The ITD of the probe was always equal to ITDtarget. The task was to indicate if the probe tone differed from the target tone or not.

As in experiment 2, the precursor sounds consisted of 121 tones spaced by 1/60 octave and ranging from 200 to 800 Hz. All these tones had again an ITD equal to ITDbackground. Since they were not resolvable by the auditory system, the precursor sounds were perceived as noise, whereas the test sounds were perceived as tonal chords. Therefore, when a test sound was preceded by a precursor, the background components of the test sound were not liable to be perceived as a repetition (or quasirepetition) of a portion of the precursor. In addition, since the test sound was separated from the precursor by a 100 ms silent ISI, the sequence was not likely to induce a “retrospective continuity illusion” (Carlyon et al., 2005).

The precursor sounds and test sounds had the same overall level, 70 dB SPL. Thus, whereas the component tones of the precursors had (as before) a SPL of 49 dB, the SPL of the test sounds’ component tones was 60.5 dB. Test sounds and probe tones had a duration of 200 ms. As mentioned above, the ISI between precursor and test sounds was 100 ms. The ISI between test and probe stimuli was 1200 ms.

On a given trial, the test sound was preceded by either a “long” precursor (1200 ms), a “short” precursor (300 ms), or no precursor at all. These three conditions were run in separate blocks of 40 trials (12 blocks for each condition in the experiment proper). Within each block, there were 20 trials in which ITDbackground = 0 and ITDtarget ≠ 0 (the right ear leading) and 20 trials in which the opposite was true. As in experiment 2 (and for the same reason), any trial of one of these two types was always followed and/or preceded by a trial of the other type. The order of conditions was randomized within sessions.

Five listeners were tested, among whom four (S1, S2, S3, and S4) had participated as subjects in at least one of the previous experiments. The fifth subject, S7, was in his 20s and had a normal audiogram at each ear. Following some preliminary sessions in which the nonzero ITD value was varied adaptively, this parameter was set to 110 μs for S1, 600 μs for S2, 200 μs for S3, 350 μs for S4, and 150 μs for S7.

C. Results

A three-way ANOVA of the scores measured within blocks of trials [subject×precursor duration (0,300, or 1200 ms)×ITDtarget (zero versus nonzero)] revealed significant main effects of precursor duration [F(2,330)=23.0, P < 0.001] and ITDtarget [F(1,330)=15.1, P < 0.001] but no significant interaction between these two factors [F(2,330) < 1]. There was a significant interaction between the subject and precursor duration factors [F(8,330)=2.2, P=0.03] and a significant three-way interaction [F(8,330)=2.0, P=0.05]. As in experiment 2, overall, performance was better when ITDtarget differed from 0 (mean percentage of correct responses: 78.6%) than when ITDtarget was equal to 0 (mean percentage: 74.4%). The results concerning precursor duration, collapsed across the ITDtarget variable, are displayed in Fig. 4. Comparisons between the three conditions with the Holm-Sidak method showed that the mean score obtained for each condition differed significantly (P < 0.05) from the mean score obtained for each of the other two conditions.
These differences were in the direction predicted by the CNA hypothesis. Adding long (1200 ms) precursors to the test sounds improved performance by 8.8% on average. Performance was also globally better in the long precursor condition than in the short precursor condition; however, this difference was quite small (3.2%).

V. GENERAL DISCUSSION

For any living organism, it is crucially important to detect changes in the environment. The existence of sensory mechanisms enhancing the internal representation of a new stimulus in a background of previously perceived stimuli is thus clearly profitable. The present research shows that, in humans, the binaural system—more specifically the processing of ITDs—can play a role in this enhancement. Experiment 1 indicated that it can modulate the enhancement of newly arriving spectral energy. Moreover, experiments 2 and 3 indicated that it is possible to enhance a pitch percept on the sole basis of ITD manipulations, without any spectral change. In the latter case, however, the observed effects were small.

It has been found in several previous studies (McFadden, 1966; Robinson and Trahiotis, 1972; Yost, 1985; Bernstein et al., 2006) that the binaural detection of an interaurally antiphase is easier when and are preceded by a “forward fringe” consisting of a sample of alone than when and are simply pulsed synchronously in the absence of a fringe. One might think that this observation is closely related to what we found in experiments 2 and 3, but that is probably not the case. In the just-mentioned masking experiments, the signal had a fixed frequency across trials and it was superimposed on an independent noise, thus producing a time-varying ITD; it appeared that a forward fringe improved signal detection significantly only when the signal was very short. In contrast, the target sounds that we used in experiments 2 and 3 varied randomly from trial to trial and all the components of the test sounds had a steady ITD; moreover, we found informally that the duration of the test sounds was not a critical parameter with respect to enhancement. It should also be noted that CNA was never considered as a possible explanation for the benefit of forward fringes in the masking experiments using configurations; Bernstein et al. (2006) proposed a quite different scenario.

Another previous study, by Kubovy and Howard (1976), is more closely related to the present work. Kubovy and Howard (1976) generated sequences of binaural chords in which each chord consisted of six synchronous pure tones with different IPDs. The tones had the same frequencies, ranging from 392 to 659 Hz and forming a diatonic musical scale, in all chords. In the initial chord, the IPDs were an arbitrary function of frequency. Each of the subsequent chords was identical to the first chord, except for a modification in the IPD of a single tone. The tone with the modified IPD changed from chord to chord in a sawtooth manner, going gradually from 392 to 659 Hz in some sequences and vice versa in other sequences. It appeared that, in such sequences, listeners were able to track the tone with the modified IPD from chord to chord; as a result, they perceived an ascending or descending cyclical melodic pattern. For a majority of the tested listeners, this task was feasible even when the chords were separated by an ISI as long as 1 s (but not much longer than that). An important point is that the modified IPDs had no special characteristic which would allow the listener to identify the target component of a chord in this chord alone; some kind of memory of the initial IPDs (generally repeated from chord to chord) was necessary in order to identify and track the target tones.

Using chord sequences which were similar but not identical to the sequences of Kubovy and Howard (1976), Culling (2000) obtained different results. He found that ascending and descending progressions of the target tone (with the modified IPD) were easy to discriminate from each other when the ISI was very short (0–20 ms), but that for most listeners, discrimination became nearly impossible for an ISI of 160 ms. The reason why discrepant results were obtained in these two experiments is not clear. In any case, the results of Kubovy and Howard (1976) are trustworthy. Just before the present study, we replicated their experiment with exactly the same stimuli. The outcome was consistent with their results. We tested five listeners, all of whom gave more than 70% of correct responses (“ascending” versus “descending”) for a 200 ms ISI.

The perceptual phenomenon described by Kubovy and Howard (1976) is, we believe, a case of binaural enhancement (in which each chord, except for the very first one, serves both as a precursor sound and as a test sound). However, their experiment does not clearly demonstrate the existence of binaural enhancement. The mere fact that in their sequences of chord, it is possible to discriminate between ascending and descending progressions of the target tone does not prove that, within a chord, the target tone is more audible than the other tones. An alternative interpretation is that all tones can be heard out equally clearly and that the change in IPD simply serves as a pointer to the tone on which the listener’s attention should be focused. In our experiments 2 and 3, by contrast, the audibility of the target sounds was assessed in the absence as well as the presence of precursor sounds, and a comparison was made between the two corresponding data sets.
In fact, Kubovy and Howard (1976) did not describe their finding as a case of enhancement and did not explicitly consider CNA as a possible explanation. Their conclusion was instead—or more vaguely—that there exists a form of auditory memory that is “in the service of perception, a percept-forming memory or in particular a pitch-segregating memory” (pp. 536–537). The present research, on the other hand, was primarily intended to test the idea that CNA is a source of enhancement and thus plays a role in human auditory perception. Our results provide some support to that idea but not a strong one; moreover, they suggest that, if it does exist, CNA has only small effects.

As already pointed out above, the presentation of a precursor sound before a target sound may improve the perception of the target sound for a number of reasons that have nothing to do with CNA. Generally speaking, the precursor may serve as an attentional cue. It may also be automatically grouped with the background of the target sound, thus isolating the target itself from irrelevant stimulation (see in this regard Best et al., 2007). Yet another possibility, specific to binaural stimuli, is the interaural decorrelation process hypothesized by Culling (2000) and mentioned in Sec. III C. Our best evidence for a role of CNA is probably that obtained in experiment 3 because here, timbral cuing was ruled out, as well as decorrelation (due to the 100 ms ISI) and sequential grouping (the background components of the test sound could not be perceived as a repetition or an extension of part of the precursor). In that experiment, however, the effect of 1200 ms precursors was barely different from the effect of precursors which were four times shorter (300 ms), a problematic fact for the CNA hypothesis.

It is conceivable, of course, that we would have observed stronger binaural enhancement effects if we had used precursors lasting several seconds. Very long “adapter” (i.e., precursor) stimuli have been employed in the experiments concerning the effect of a sound on the subjective lateralization of a subsequent sound (e.g., Kashino and Nishida, 1998). However, very long precursors are not required in order to obtain monaural enhancement of newly arriving spectral energy (see, e.g., Viemeister, 1980). Moreover, if ITD-specific CNA is an extremely slow process, then this form of CNA is unlikely to play an important role in ordinary listening situations because people often move their head or their whole body.

ACKNOWLEDGMENTS

The research reported here was carried out while M.S. held a position of “chercheur associé étranger” at the CNRS; this author is now affiliated with Siemens Audiological Engineering Group (Erlangen, Germany). We thank Dr. Constantine Trahiotis for a very helpful discussion. Dr. Wesley Grantham and two anonymous reviewers made constructive comments on a previous version of this article.


Šerman et al.: Enhancement, adaptation, and the binaural system 4419


