On the binding of successive sounds: Perceiving shifts in nonperceived pitches[a]

Laurent Demany[b] and Christophe Ramos
Laboratoire de Neurophysiologie, CNRS and Université Victor Segalen (UMR 5543), 146 rue Leo-Saignat, F-33076 Bordeaux, France

(Received 12 August 2004; revised 3 November 2004; accepted 30 November 2004)

It is difficult to hear out individually the components of a ''chord'' of equal-amplitude pure tones with synchronous onsets and offsets. In the present study, this was confirmed using 300-ms random (inharmonic) chords with components at least 1/2 octave apart. Following each chord, after a variable silent delay, listeners were presented with a single pure tone which was either identical to one component of the chord or halfway in frequency between two components. These two types of sequence could not be reliably discriminated from each other. However, it was also found that if the single tone following the chord was instead slightly (e.g., 1/12 octave) lower or higher in frequency than one of its components, the same listeners were sensitive to this relation. They could perceive a pitch shift in the corresponding direction. Thus, it is possible to perceive a shift in a nonperceived frequency/pitch. This paradoxical phenomenon provides psychophysical evidence for the existence of automatic ''frequency-shift detectors'' in the human auditory system. The data reported here suggest that such detectors operate at an early stage of auditory scene analysis but can be activated by a pair of sounds separated by a few seconds.
[DOI: 10.1121/1.1850209]

PACS numbers: 43.66.Mk, 43.66.Hg [JHG] Pages: 833–841

I. INTRODUCTION

The understanding of speech or the identification of a melody requires a perceptual binding of successive sounds that differ from each other. Little is known about the neural machinery responsible for this binding in the human brain. Remarkably, a melody retains its perceptual identity when it is transposed in the frequency domain, i.e., when the frequencies of the successive tones are multiplied by a common factor (Attneave and Olson, 1971; Divenyi and Hirsh, 1978). This shows that human listeners perceive melodies as patterns rather than mere concatenations of independent tones. From that point of view, a melody may not be completely equivalent to a sequence of tones varying in intensity rather than frequency: Curiously, binary sound sequences with a complex structure (e.g., ABAAABBAAB) are identified more accurately when their two components (A and B) differ in frequency than when they differ in intensity, independently of the magnitude of the difference (McFarland and Cacace, 1992). In order to account for the propensity of melodies to be perceived as patterns, it has been speculated by some authors (Deutsch, 1969; van Noorden, 1975; Anstis and Saida, 1985; Okada and Kashino, 2003) that the human auditory system contains automatic ''frequency-shift detectors'' which are sensitive to the direction of such shifts and can be activated by a pair of successive sounds even when these sounds are separated by a silent delay. The existence of shift detectors operating in the frequency or pitch domain might also account for an informal observation made by Davis *et al.* (1951) and Bilsen (2001) during their investigations of the weak pitch sensations produced by certain binaural interactions (Bilsen, 2001) or evoked by the periodicity of stimuli consisting of high-rank harmonics of a missing fundamental (Davis *et al.*, 1951). These authors noted that the audibility of pitch in a given sound can be markedly improved if, instead of being presented alone repeatedly, this sound alternates with a similar sound liable to evoke a somewhat different pitch.

In the auditory cortex of cats or monkeys, the response of many neurons to a pure tone can be strongly influenced by a preceding pure tone with a different frequency (Brosch and Schreiner, 1997, 2000; Weinberger and McKenna, 1988; McKenna *et al.*, 1989). However, this sensitivity to discrete frequency shifts seems to hold only when the interstimulus interval (ISI) is relatively short—less than 1 s. By contrast, the hypothesis that will be considered here is the existence of shift detectors (possibly more complex than single neurons) functioning even for ISIs lasting a few seconds. The hypothesis in question is consistent with the fact that the human ability to detect consciously small frequency shifts between temporally remote tones, and to identify the direction of such shifts, does not depend on the subject's mental activity or focus of attention during the ISI (Demany *et al.*, 2001; Clément, 2001). It is also worthy to note that this ability differs from the ability to detect intensity shifts with respect to their dependence on the ISI (Clément *et al.*, 1999).

The present paper stems from our accidental discovery of a paradoxical perceptual phenomenon which seems to lend strong support to the idea that automatic frequency-shift detectors exist in the human auditory system. This phenomenon is elicited by the successive presentation of: (i) a sum of N synchronous pure tones with equal amplitudes, forming an inharmonic ''chord;'' (ii) a single pure tone (''T''). The

[a]Portions of this work were presented at the 147th meeting of the Acoustical Society of America, New York, NY, May 2004.
[b]Electronic mail: laurent.demany@psyac.u-bordeaux2.fr

chord's components may have randomly drawn frequencies, but must be at least a few semitones apart in order to be separated in the cochlea (Plomp, 1964; Moore and Ohgushi, 1993; Shera *et al.*, 2002). Even so, if the chord is rather brief and N exceeds 2 or 3, it will be difficult for the listener to perceive the pitches of the chord's components. This difficulty reflects an "informational masking" effect (Neff and Green, 1987; Kidd *et al.*, 1994). A "fusion" (Bregman, 1990) of the chord's components takes place at a central level of the auditory system, so that typically the whole chord is heard as a single sound, with a tonal quality but only a vague global pitch. We observed indeed that if the following tone, T, is identical to one component of the chord, this identity generally goes unnoticed. However, we also observed that if T is instead slightly lower or higher in frequency than one component of the chord, many listeners are able, after a little practice, to perceive a pitch shift in the corresponding direction. To their surprise, they find that it is possible to hear an upward or downward shift in a pitch which was not heard in the chord (and is not heard retroactively when T is presented). Moreover, it appears that the direction of the melodic interval formed by T and its neighbor in the chord can be identified even when the chord and T are separated by a silent ISI of a few seconds. The experiments described below substantiate these counterintuitive observations.

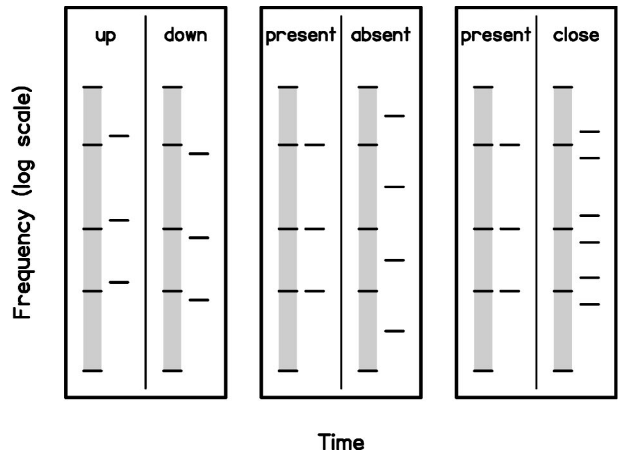

FIG. 1. Illustration of the stimulus configurations used in the "up/down," "present/absent," and "present/close" conditions. Each horizontal segment represents a pure tone and the shaded areas represent a possible chord. For a given chord, the frequency of the following tone, T, could take 6, 7, or 9 possible values, depending on the experimental condition. In this figure, for visual reasons, the T tones have been horizontally positioned close to the chords. In the experiments, actually, the time interval separating T from the chord always exceeded the duration of both stimuli.

II. EXPERIMENT 1

A. Method

On each trial run in this experiment (and the subsequent ones), the listener was presented with a chord consisting of five synchronous pure tones, spaced by intervals which were four independent random variables. The probability distribution of each interval was rectangular on a log-frequency scale and ranged from 6 semitones (1/2 octave) to 10 semitones (5/6 octave). The chord was randomly positioned within a 4-octave frequency range, 200–3200 Hz, and was followed by a single pure tone, T. The aim of experiment 1 was to compare perceptual performances in two conditions, respectively termed "up/down" and present/absent."

In the up/down condition, schematized in the leftmost panel of Fig. 1, T was positioned 1 semitone above or below one of the chord's three intermediate components. The component in question and the direction of the 1-semitone shift were selected at random on each trial. The listener knew that and had to judge if the 1-semitone frequency shift was made upward or downward.

In the present/absent condition (Fig. 1, central panel), T was equiprobably (i) identical to one of the chord's three intermediate components or (ii) positioned halfway in frequency between two components (frequency being scaled logarithmically). In either case, a random choice was made between the three or four possible options. The listener also knew that and had to judge if T was present in the chord or not.

In both conditions, the chord and T had a total duration of 300 ms and were gated on and off with 20-ms raised-cosine amplitude ramps. They were separated by a 500-ms silent ISI. Each T tone and component of the chord had a nominal sound-pressure level of 65 dB. The stimuli were heard binaurally (diotically), via earphones (Sennheiser HD265). They were generated via a 24-bit digital-to-analog converter (Echo Gina), at a sampling rate of 44.1 kHz. Each trial began with the presentation of a 2.1-s random melody, serving as a warning signal and consisting of seven 300-ms pure tones with frequencies drawn independently between 200 and 3200 Hz.[1] The chord was presented 600 ms after this melody. The listener, sitting in a double-walled sound-attenuating booth, had an unlimited response time after the presentation of T. His or her response, given by making a mouse-click on one of two labeled zones of a monitor screen, automatically triggered the next trial within blocks of 50 trials. Responses were not followed by an immediate feedback, but listeners were informed of their performance between successive blocks of trials. Each block was carried out in a fixed condition (up/down or present/absent); the two conditions alternated from block to block. Listeners were tested in three or four sessions, on different days, until 500 trials had been run in each condition.

The experiment was conducted on 11 normal-hearing listeners between the ages of 22 and 50 years. Two of them were the authors. Most of them had received a significant musical education. Six listeners had previously participated in other experiments concerning pitch perception (Demany *et al.*, 2001, 2004). For each listener, the experiment proper was preceded by a few training sessions—generally not more than two. The chords initially used in these training sessions consisted of only three pure tones, with a duration of 700 ms instead of 300 ms; they were similar to those employed by Demany *et al.* (2004). In addition to the 11 listeners who finally constituted the experimental group, three other listeners participated in training sessions. They were not included in the experimental group because of their apparent inability to exceed the chance level of performance in any condition.

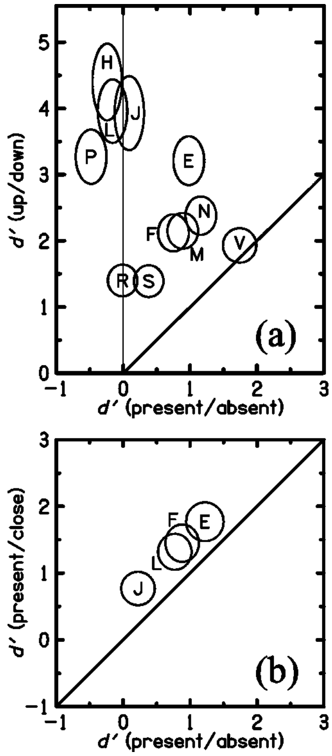

FIG. 2. Results of experiment 1 [panel (a)] and experiment 2 [panel (b)]. Each ellipse (or circle) is centered on the d' values measured in the two conditions for a given listener, and its surface represents a 95% confidence area. Oblique lines indicate where the ellipses could be centered if d' was identical in the two conditions. Listeners are identified by capital letters. Confidence intervals around d' were computed as suggested by Macmillan and Creelman (1991).

B. Results

Performance was measured in terms of d' (Green and Swets, 1974). The results are displayed in the upper panel of Fig. 2, where 11 ellipses represent the 11 listeners' data. Each ellipse (actually a circle in some cases) is centered on the d' values measured in the two conditions for a given listener, and its surface represents a 95% confidence area. Listeners are identified by capital letters in order to permit within-subject comparisons of performance across experiments; ''E'' and ''F'' are the authors.

There was a considerable interindividual variability in performance. For each listener, however, d' was larger in the up/down condition than in the present/absent condition, and the corresponding difference was generally quite pronounced. Commonly, a sensory signal is said to be detectable if $d' > 1$. It can be seen that this threshold was exceeded by all listeners in the up/down condition, but by only two listeners in the present/absent condition. Four listeners (H, J, L, P) made only few errors (1.4–6.6 %) in the up/down condition while performing at the chance level, or even slightly below it, in the present/absent condition. Only one listener (V) had statistically equivalent performance levels in the two conditions. Perhaps not fortuitously, this was also the only listener endowed with ''absolute pitch'' (Ward and Burns, 1982).

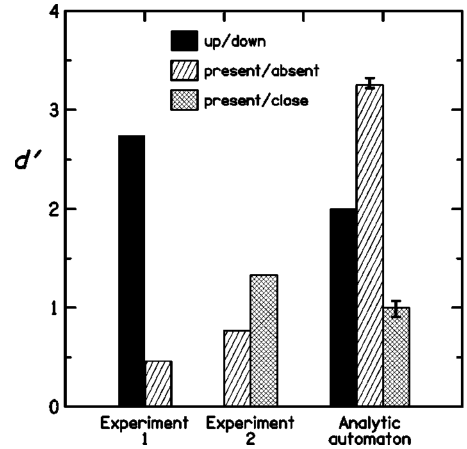

FIG. 3. Comparison of the average d' values measured in experiments 1 and 2 (left and center) with the d' values expected from the analytic automaton described in Sec. II C (right). The two small error bars represent a variability of d' resulting from the use of different values of δ_k by the automaton. The lowest and highest values of δ_k considered here (1.9–2.4 semitones for the present/absent condition, and 0.7–1.3 semitones for the present/close condition) are consistent with the largest response biases actually observed in experiments 1 and 2.

C. Discussion

Each participant in this experiment stated that the chords' components were very difficult or impossible to hear out individually. Let us consider, however, an ''analytic automaton'' Λ for which this is not the case. In both conditions, Λ takes on each trial an explicit (but relatively imprecise) measurement of the chord's component frequencies, and then relates these data to the frequency of T. In the up/down condition, the task of Λ is to determine if the musical interval formed by T and the closest component of the chord is equal to $+1$ or -1 semitone. In the present/absent condition, on the other hand, the task is to determine if the corresponding interval is equal to 0 or at least 3 semitones (given that the chords' components are separated by at least 6 semitones). Since the difference between $+1$ and -1 is smaller than the difference between at least 3 and 0, the performance of Λ in the present/absent condition should logically exceed its performance in the up/down condition. We confirmed this in a Monte Carlo simulation assuming that the strategy of Λ is optimal and that its performance is limited only by a source of noise in the frequency measurements. Our simulation specifically supposed that the measured frequency of each component of the chord is a Gaussian random variable with a mean equal to the true frequency and a standard deviation of 1 semitone. On each trial, Λ computes the differences between log transforms of the frequency of T and the measured frequencies of the chord's three intermediate components. Its response is then based on the difference δ which has the smallest absolute value. In the up/down condition, Λ responds ''up'' if and only if $\delta > 0$. In the present/absent condition, Λ responds ''present'' if and only if $|\delta| < \delta_k$, δ_k being a positive constant. The d' scores of Λ can be compared in Fig. 3 to the average d' values measured in experiment 1. In the present/absent condition, the d' score of Λ is somewhat dependent on δ_k, because $|\delta|$ is not a Gaussian variable; but, the influence of δ_k is small as long as this criterion takes reasonable values. Figure 3 makes clear that Λ is a completely inadequate model of real listeners: Whereas real listeners are much more successful in the up/down condition than in the present/absent condition, the opposite is true for Λ.

The analytic automaton Λ that we have just considered is an ideal listener, in several respects. In particular, there are no systematic errors (i.e., biases) in its frequency measurements, and its response criteria—especially δ_k—are strictly invariant across trials. In theory, the poor performance of real listeners in the present/absent condition could partly be ascribed to the difficulty of behaving like an *ideal* analytic listener rather than to an inability to hear out the individual components of the chords. This issue will be examined in the next section. Meanwhile, another issue should be considered. Although the bandwidths of the chords exceeded 2 octaves, each chord evoked as a whole a global pitch sensation, presumably related to its ''center of gravity'' in the frequency domain. This global pitch was imprecise, but it noticeably varied from trial to trial together with the chord itself. In the up/down condition, a conceivable strategy was to respond ''up'' if the pitch of T was higher than the global pitch evoked by the chord, and to respond ''down'' otherwise. To what extent was this holistic strategy profitable? The answer is that it could not be very efficient, even for a theoretical listener perceiving the global pitch in an optimal manner. Its use would have led to systematic errors when T was 1 semitone above the lower neighbor of the chord's median component, or 1 semitone below its higher neighbor. If the global pitch of a chord corresponded exactly to the geometric mean of its five component frequencies, and if the holistic strategy was used exclusively and perfectly, the d' value expected in the up/down condition was equal to 0.70. Much higher d' values were actually measured in experiment 1. For most listeners, the difference between the d' values obtained in the up/down and present/absent conditions was too large to be accounted for by the availability of holistic pitch cues in the up/down condition. It can actually be supposed that such cues were never used because each listener was initially instructed to ignore the global pitches of the chords.

III. EXPERIMENT 2

A. Rationale and method

Introspectively, the difficulty of the present/absent condition of experiment 1 originated from the fact that the chords' components were very hard to perceive individually. In this condition, there were no obvious perceptual cues permitting to give correct responses without perceiving consciously the individual components of the chords (whereas the opposite was true for the up/down condition). In theory, however, the difficulty of the present/absent condition could be accounted for otherwise. Due to the inherent inaccuracy of any perceptual measurement, and also due to the possible existence of small contextual effects in the perception of pure tone pitch (e.g., Terhardt, 1970), a chord component physically identical to T might nevertheless be perceived as different from T. This implies that, in the present/absent condition, an ''analytic'' listener had to use a response rule involving a difference criterion, as discussed in the previous section. The difference criterion, δ_k, had to be larger than 0 and stable across trials. Crucially, δ_k had to be defined by the listener him-/herself since it was an *internal* reference. In the up/down condition, on the other hand, an internal reference was not *a priori* needed by an analytic listener; it could be sufficient to determine if T was higher or lower than the closest component of the chord (an ''external'' reference). A conceivable hypothesis, therefore, was that the difficulty of the present/absent condition stemmed from the need, in this condition, of a stable internal reference serving as a response criterion for the assessment of pitch differences.

In experiment 2, this hypothesis was tested by comparing the performances of four listeners in the present/absent condition and a new condition, termed ''present/close'' (Fig. 1, rightmost panel). In both conditions, T was equiprobably present or not present in the chord and the listener had to make a two-alternative judgment in this regard. The two conditions differed from each other only when T was not present in the chord. On the corresponding trials in the present/close condition, T was positioned exactly 1.5 semitone (1/8 octave) above or below one of the chord's three intermediate components; the six possible options were equiprobable. Note that, in this condition, T was always much closer in frequency to one component of the chord than to any other component since the components were spaced by intervals of at least 6 semitones. In contrast, when T was absent from the chord in the present/absent condition, T was at least 3 semitones away from any component since it was positioned halfway between two components. Thus, for an analytic listener, the present/close condition was more difficult than the present/absent condition. In Fig. 3, this is confirmed for the analytic automaton Λ defined in Sec. II C; its d' score in the present/close condition is about 1.0, much worse than its score of about 3.3 in the present/absent condition. Clearly, if the hypothesis that we intended to test in the present experiment were correct, performance could not be better in the present/close condition than in the present/absent condition, since the alleged problem of the internal reference serving as a response criterion existed in both conditions and was not less critical in the present/close condition. However, our prediction was on the contrary that performance would be better in the present/close condition than in the present/absent condition. This prediction rested on the assumption that, in the present/close condition, a clear upward or downward pitch shift would be generally audible on ''close'' trials, but not or less so on ''present'' trials. In contrast, the difficulty of the present/absent condition suggested that ''absent'' trials could not be reliably discriminated from ''present'' trials on the basis of the audibility of a pitch shift.

The procedure used in experiment 2 was essentially the same as that employed in experiment 1, except for the replacement of the up/down condition by the present/close condition. There was again a 500-ms ISI between each chord and the following T tone. Eight blocks of 50 trials were run in each of the two conditions; this required only two sessions. The experiment proper was preceded by a single and short training session. The four listeners who acted as subjects included the authors. Each listener had been previously tested in experiment 1.

B. Results and discussion

The individual d' scores obtained in each condition are displayed in the lower panel of Fig. 2, and their average values are plotted in Fig. 3. Each listener was significantly more successful in the present/close condition than in the present/absent condition. This outcome is consistent with our prediction and contradicts the hypothesis tested in the experiment. It seems clear that the difficulty of the present/absent condition did not stem from a difficulty to use an adequate response criterion in this condition. A more plausible explanation is that it was difficult to hear out the individual components of the chords, as suggested by the listeners' introspective reports.

Yet, d' had an average value of only 1.33 in the present/close condition, and was not dramatically lower in the present/absent condition. The modest values of d' in the present/close condition imply that it was not extremely easy to discriminate "close" trials from "present" trials on the basis of the audibility of a pitch shift. One might infer from this that pitch shifts were not easily heard on "close" trials. However, such a conclusion would be inconsistent with the listeners' introspective reports. From these reports, it appears instead that a pitch shift was, to some extent, liable to be heard on any type of trial—even a "present" trial. The fact that d' was lower in the present/close condition than in the up/down condition of experiment 1 is not, *per se*, surprising insofar as, in the present/close condition, the directions of the pitch shifts heard were not valid cues. In the absence of valid directional cues, the listeners had to base their judgments only on the salience and/or the magnitude of the pitch shifts heard.

IV. EXPERIMENT 3

If, in the present/absent condition, performance was poor owing to a difficulty to use an adequate response criterion and not because the chords' components were hard to hear out individually, then performance should be poor not only when T follows the chord but also when T precedes the chord. By contrast, it could be expected that presenting T before the chord rather than after it would increase performance by improving the audibility of T within the chord on "present" trials. This rationale led us to perform a short experiment testing listeners in two variants of the present/absent condition: a "chord-T" variant, replicating what had been done in experiments 1 and 2, and a "T-chord" variant in which T preceded the chord instead of following it. In both variants, there was a 500-ms ISI between the chord and T. For each subject, in the experiment proper, eight blocks of 50 trials were run in each of the two variants, alternating from block to block. The four listeners who acted as subjects had previously participated in experiment 1; two of them were the authors. Before the experiment proper, each listener received some training in the T-chord variant.

The results are displayed in Fig. 4. For three listeners, d' was much higher in the T-chord variant than in the chord-T

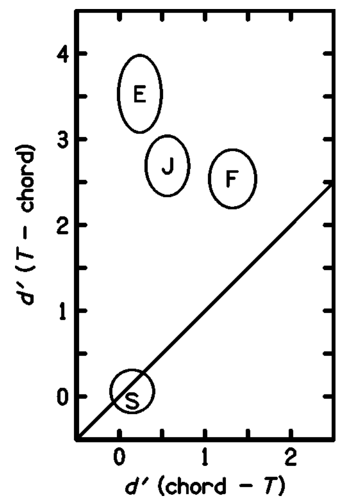

FIG. 4. Results of experiment 3, in the same format as Fig. 2.

variant; the fourth listener (S) performed at the chance level in the two variants.[2] Overall, these data confirm the view that, in the present/absent condition of experiments 1 and 2, performance was limited by an informational masking effect rather than by a difficulty to use an adequate response criterion. For at least some listeners, the informational masking of T within the chord can be efficiently opposed by presenting T before the chord rather than after it. The source of this benefit is presumably a focusing of the listener's attention on the appropriate spectral region of the chord during the chord presentation.

V. EXPERIMENT 4

This final experiment was an extension of experiment 1. Its aim was to determine if the up/down condition remains easier than the present/absent condition when the ISI separating T from the chord is made much longer than the 500-ms ISI used in experiment 1. A positive answer was expected because preliminary listening sessions revealed that the 1-semitone frequency shifts occurring in the up/down condition were still liable to evoke sensations of pitch shift for an ISI of a few seconds.

Four listeners served as subjects. Two of them were the authors, and all of them had previously served as subjects in experiment 1. Each listener was at first tested only in the up/down condition, using ISIs of 0.5, 1, 2, 4, and 8 s. Each of these ISIs was used in a single block of 50 trials per session, until 400 trials had been run for every ISI; the order in which the ISIs were used was counterbalanced across sessions. Then, two final sessions included a total of 400 trials in the present/absent condition with an ISI fixed at 4 s.

Figure 5 shows the results. In the up/down condition, as the ISI increased, d' decreased with a negative acceleration. When the ISI was 4 s, d' was still well above 0 in this condition; its average value was 0.94. For the same ISI, in the present/absent condition, d' was definitely lower; its average value was 0.33.

These data call for comments relating to a hypothesis put forth by van Noorden (1975; see also Anstis and Saida, 1985) on the perception of pitch shifts between successive

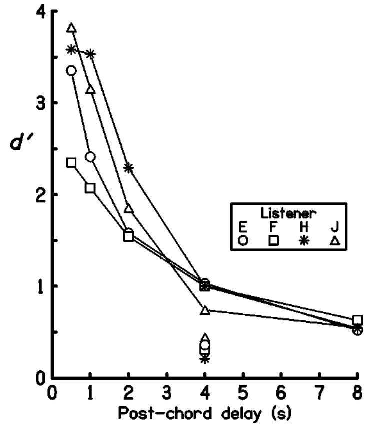

FIG. 5. Results of experiment 4: d' values measured in the up/down condition (symbols connected by lines) and the present/absent condition (nonconnected symbols), as a function of the silent interval separating the chord from T.

tones. van Noorden argued that such shifts are perceived in the same way as discontinuous displacements of visual objects. In the visual domain, a discontinuous displacement is able to evoke a sensation of motion, called ''phi'' motion, when the ISI amounts to 100–300 ms, but not when the ISI is much longer. Likewise, according to van Noorden, a pair of successive tones with different frequencies can evoke a sensation of ''pitch motion'' when the ISI is, e.g., 200 ms, but not when the ISI is much longer. If this idea is correct, then the present data imply that the mechanism underlying listeners' good performance in the up/down condition is independent of the neural processing of rapid frequency shifts. However, van Noorden's hypothesis may be wrong. To the best of our knowledge, whereas there are objective psychophysical data (i.e., measures of performance) supporting the distinction between percepts of ''succession with motion'' and ''succession without motion'' in the visual domain (Palmer, 1986), this is not the case in the domain of pitch.

VI. GENERAL DISCUSSION

A. Summary of the findings

For at least some fraction of human listeners,[3] it is possible to identify the direction of a discrete frequency shift in an informational masking context preventing a conscious perception of the initial frequency. This was demonstrated using sound sequences consisting of a random chord of pure tones followed by a single pure tone. When the single tone was positioned slightly above or below a randomly selected component of the chord, a pitch shift in the corresponding direction could be heard by 11 listeners. The position of the single tone relative to the closest component of the chord could be reliably identified even when the chord and the single tone were separated by a silent ISI of a few seconds. However, for most of the tested listeners, it was very difficult to discriminate between sequences in which the single tone was respectively identical to one component of the chord or halfway in frequency between two components. The pattern of listeners' performances in the three discrimination tasks employed was very different from the pattern expected from an ideal ''analytic'' discriminator, hearing out the individual components of the chords. Moreover, the listeners' pattern of performance could not be produced either by a nonideal analytic discriminator, hearing out the individual components of the chords but using a defective response criterion.

B. A schematic model of frequency-shift detectors

Our results can be understood by assuming the existence, in the human auditory system, of automatic ''frequency-shift detectors'' playing a role in the binding of successive sounds and sensitive to memory traces left in some primitive ''echoic store'' (Kubovy and Howard, 1976). More specifically, we hypothesize that these detectors have the following properties: (i) some of them are activated only by upward frequency shifts, while others are activated only by downward shifts; (ii) within each subset, the detectors' response is stronger for small shifts than for large shifts, but some minimum magnitude of shift is required to elicit a response; (iii) when detectors of upward shifts and downward shifts are simultaneously activated—this was presumably the case in our experiments—the dominantly perceived shift is in the direction preferred by the subset of detectors with the stronger activation. According to the corresponding model, performance in our up/down condition ought to be relatively good because ''up'' and ''down'' trials dominantly activated different subsets of shift detectors. By contrast, on ''present'' trials as well as on ''absent'' trials, the two subsets of detectors were in theory activated with approximately the same strength, which explains why it was more difficult to discriminate between these two types of trial. Finally, regarding the present/close condition, the model implies that one subset of detectors tended to be more strongly activated than the other subset on ''close'' trials, whereas this was less true on ''present'' trials; thus, performance in this condition ought to be better than in the present/absent condition, but still poorer than in the up/down condition.

It should be noted that the three hypotheses presented above tally with those proposed by Allik *et al.* (1989) in order to account for data concerning the perception of complex sequences of chords.[4] A more precise and quantitative model is certainly desirable, but probably premature at present. Further psychophysical studies are needed to this aim.

C. Physiological considerations

In the visual cortex of mammals, continuous motions of objects are detected by direction-selective neurons. In the auditory cortex, similarly, certain neurons selectively respond to either upward or downward frequency glides (Whitfield and Evans, 1965; Zhang *et al.*, 2003). It has been hypothesized that, in humans, such neurons govern the perceptual experience of rapid and continuous frequency modulations (Gardner and Wilson, 1979; Tansley and Regan, 1979; Wilson *et al.*, 1994; Shu *et al.*, 1993), and also play a role in the perception of successive steady tones separated by a short ISI (van Noorden, 1975; Okada and Kashino, 2003). At present, however, both of these points remain speculative (Wakefield and Viemeister, 1984; Moody *et al.*, 1984). Whereas several studies showed that the responses of individual visual neurons to motion provide information which determines the subject's ability to identify behaviorally the direction of motion (e.g., Ditterich *et al.*, 2003), analogous demonstrations have not been reported in the auditory domain. According to Weinberger and McKenna (1988) and McKenna *et al.* (1989), the response of a single cortical neuron to a given tone preceded by a higher- or lower-frequency tone can depend on the direction of the shift even if the two tones are separated by more than 500 ms; but, the corresponding evidence is limited. At least for ISIs exceeding 500 ms, it is possibly erroneous to search for a physiological substratum of the perceptual sensitivity to frequency shifts in the discharge rates of individual neurons. This substratum might in fact lie in the cross correlations of separate neurons' responses, as suggested by Espinosa and Gerstein (1988). In other words, the shift detectors might not be reducible to single neurons and consist instead of neuronal assemblies.

Functional imaging (fMRI) studies have recently indicated that, in humans, the cerebral activity induced by melodies differs from that induced by repetitions of a single tone (Griffiths *et al.*, 2001; Patterson *et al.*, 2002; Warren *et al.*, 2003) or by sequences of sounds with changing spatial locations (Warren and Griffiths, 2003). However, these studies suggest that melodies engage specific neural processes only beyond the primary auditory cortex. More precisely, on the basis of both functional imaging data (Patterson *et al.*, 2002) and behavioral data relating to the consequences of brain lesions on pitch perception (Johnsrude *et al.*, 2000), it has been surmised that the lateral part of Heschl's gyrus, in the right hemisphere, is crucially involved in the perception of frequency shifts. The special sound sequences used in the present investigation could be profitably reemployed in future imaging and lesion studies: they seem to be particularly appropriate tools for the localization of brain structures responsible for the sensitivity to frequency shift *per se*.

We argue here that the human auditory system is equipped with automatic detectors of frequency shifts. From a more general point of view, the idea that modifications in a sound can be detected automatically within the auditory system has already been supported by a large amount of research concerning an event-related brain potential termed "mismatch negativity" or MMN (Näätänen and Winkler, 1999; Schröger, 1997). This brain potential, elicited by any type of acoustic novelty or "deviance" in a sound sequence, is observable when the subject does not pay attention to the sequence. However, Cowan *et al.* (1993) reported that no MMN can be elicited by a sequence of only two sounds, which would imply that the change-detection process reflected by the MMN differs from the one investigated here. Recently, Jääskeläinen *et al.* (2004) claimed that a sequence of only two sounds is actually able to elicit an MMN. According to these authors, and also May *et al.* (1999), the MMN does not originate from a neural structure responding to auditory change *per se*, contrary to a commonly held idea; instead, it is produced by the neural population responding to the deviant stimulus itself and is due to transient adaptation of feature-specific neurons. Note that a change detector based on adaptation phenomena should logically lead its possessor to be more successful in our present/absent condition than in our up/down condition, whereas the opposite is true for real listeners. Thus, the thesis advocated by Jääskeläinen *et al.* and May *et al.* would still imply that the change-detection mechanism reflected by the MMN differs from the one investigated here.

D. Change detection and scene analysis

The change-detection mechanism investigated here appears to be more primitive than the one reflected by the MMN. An MMN originates from the processing of high-level sound representations which are quite similar to the conscious percepts evoked by the stimuli. Indeed, it has been shown that the MMN is sensitive to the perceptual fission of rapid melodic sequences into separate and concomitant sound streams (Sussman *et al.*, 1999), to the auditory continuity illusion (Micheyl *et al.*, 2003), to a change in the virtual pitch of complex tones with randomly varying spectral contents (Winkler *et al.*, 1995), and even to the listener's linguistic background (Näätänen *et al.*, 1997). In our up/down condition, by contrast, the *T* tone was put in relation with a tone which could not be consciously perceived because it was grouped with synchronous tones producing an informational masking effect. A representation of the masked tone's frequency existed at the cochlear level and certainly beyond in the auditory pathway. However, there was no representation of this specific frequency in the outcome of auditory scene analysis since the corresponding pitch could not be heard.

It would be unwarranted to infer from the latter point that the perceptual phenomenon described here is unrelated to auditory scene analysis. On the contrary, we suppose that the auditory system makes use of automatic frequency-shift detectors in the global scene analysis process. They are probably useful for the creation of links between temporally separate acoustic events which differ from each other in spectral content, but were nonetheless produced by one and the same sound source and should thus be integrated into a common perceptual stream.

ACKNOWLEDGMENTS

The authors thank R.P. Carlyon, A. de Cheveigné, J.M. Edeline, C. Lorenzi, D. Pressnitzer, C. Semal, N. Versfeld, and especially A. Gorea and C. Micheyl for helpful discussions or comments on the manuscript. This work was supported by a grant from the CNRS (Programme Interdisciplinaire "Cognition et Traitement de l'Information").

[1]This random melody not only served as a warning signal: It was also intended to "erase" the memory trace of the *T* tone used on the previous trial. Without the melody, in the up/down condition, a given *T* tone was liable to be mistakenly compared with the previous *T* tone rather than with a component of the chord. Informal pilot studies suggested to us that such confusions were indeed easily made. In contrast, when each trial began with a random melody, we felt that the *T* tone was not liable to be mistakenly compared with the last component tone of the melody, because all the component tones of the melody were perceptually grouped into a separate entity. This was not formally checked, however. The ideal procedure would be to replace the melody by a visual warning signal and to insert a very long silent interval between consecutive trials in order to minimize the disrupting effect of irrelevant memory traces.

[2]In the chord-T variant, Listener E was significantly less successful here than during experiments 1 and 2. By contrast, F and J were slightly more successful than before. In view of these changes in performance, it is worthy to note that experiment 3 was conducted 1 year after experiment 2 (and actually after the experiment numbered 4 in the present paper).

[3]It was mentioned in Sec. II A that three volunteers for experiment 1 were eventually dismissed because they did not succeed at all in either condition. These listeners might have been more successful in the up/down condition after a longer, and/or different, training phase. However, the converse is also true: Some normal-hearing people may be completely unable to experience the perceptual phenomenon permitting success in the up/down condition. In this respect, the influence of musical education is still unclear. The three listeners that we had to dismiss were not more impervious to music than those who provided the data. Indeed, one of them was a professional orchestra conductor.

[4]In the experiments of Allik *et al.* (1989), the subjects' task was to identify the direction of pitch motion in stochastic sequences of chords made up of either pure or complex tones (with frequency ratios which were generally much smaller than those employed in the present research). The results were accounted for by a "dipole contribution model." This model is not directly usable here but it is based on assumptions similar to those we make. Essentially, Allik *et al.* argued that the perceived direction of pitch motion was determined by the sum of the contributions of direction-sensitive dipoles consisting of pairs of consecutive tones *close in pitch*.

Allik, J., Dzhafarov, E.N., Houtsma, A.J.M., Ross, J., and Versfeld, H.J. (**1989**). "Pitch motion with random chord sequences," Percept. Psychophys. **46**, 513–527.

Anstis, S., and Saida, S. (**1985**). "Adaptation to auditory streaming of frequency-modulated tones," J. Exp. Psychol. Hum. Percept. Perform. **11**, 257–271.

Attneave, F., and Olson, R.K. (**1971**). "Pitch as a medium: A new approach to psychophysical scaling," Am. J. Psychol. **84**, 147–166.

Bilsen, F.A. (**2001**). "The case of the missing central spectra," in *Physiological and Psychophysical Bases of Auditory Function*, edited by D.J. Breebart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs, and R. Schoonoven (Shaker, Maastricht, The Netherlands), pp. 145–152.

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

Brosch, M., and Schreiner, C.E. (**1997**). "Time course of forward masking tuning curves in cat primary auditory cortex," J. Neurophysiol. **77**, 923–943.

Brosch, M., and Schreiner, C.E. (**2000**). "Sequence sensitivity of neurons in cat primary auditory cortex," Cereb. Cortex **10**, 1155–1167.

Clément, S. (**2001**). "La mémoire auditive humaine : Psychophysique et neuroimagerie fonctionnelle," Ph.D. dissertation (Université Victor Segalen, Bordeaux, France).

Clément, S., Demany, L., and Semal, C. (**1999**). "Memory for pitch versus memory for loudness," J. Acoust. Soc. Am. **106**, 2805–2811.

Cowan, N., Winkler, I., Teder, W., and Näätänen, R. (**1993**). "Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP)," J. Exp. Psychol. Learn. Mem. Cogn. **19**, 909–921.

Davis, H., Silverman, S.R., and McAuliffe, D.R. (**1951**). "Some observations on pitch and frequency," J. Acoust. Soc. Am. **23**, 40–42.

Demany, L., Clément, S., and Semal, C. (**2001**). "Does auditory memory depend on attention?," in *Physiological and Psychophysical Bases of Auditory Function*, edited by D.J. Breebart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs, and R. Schoonoven (Shaker, Maastricht, The Netherlands), pp. 461–467.

Demany, L., Montandon, G., and Semal, C. (**2004**). "Pitch perception and retention: two cumulative benefits of selective attention," Percept. Psychophys. **66**, 609–617.

Deutsch, D. (**1969**). "Music recognition," Psychol. Rev. **76**, 300–307.

Ditterich, J., Mazurek, M.E., and Shadlen, M.N. (**2003**). "Microstimulation of visual cortex affects the speed of perceptual decisions," Nat. Neurosci. **6**, 891–898.

Divenyi, P.L., and Hirsh, I.J. (**1978**). "Some figural properties of auditory patterns," J. Acoust. Soc. Am. **64**, 1369–1385.

Espinosa, I.E., and Gerstein, G.L. (**1988**). "Cortical auditory neuron interactions during presentation of 3-tone sequences: Effective connectivity," Brain Res. **450**, 39–50.

Gardner, R.B., and Wilson, J.P. (**1979**). "Evidence for direction-specific channels in the processing of frequency modulation," J. Acoust. Soc. Am. **66**, 704–709.

Green, D.M., and Swets, J.A. (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, New York).

Griffiths, T.D., Uppenkamp, S., Johnsrude, I.S., Josephs, O., and Patterson, R.D. (**2001**). "Encoding of the temporal regularity of sound in the human brainstem," Nat. Neurosci. **4**, 633–637.

Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levänen, S., Lin, F.H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., and Belliveau, J.W. (**2004**). "Human posterior auditory cortex gates novel sounds to consciousness," Proc. Natl. Acad. Sci. U.S.A. **101**, 6809–6814.

Johnsrude, I.S., Penhune, V.B., and Zatorre, R.J. (**2000**). "Functional specificity in the right human auditory cortex for perceiving pitch direction," Brain **123**, 155–163.

Kidd, G., Mason, C.R., Deliwala, P.S., Woods, W.S., and Colburn, H.S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.

Kubovy, M., and Howard, F.P. (**1976**). "Persistence of a pitch-segregating echoic memory," J. Exp. Psychol. Hum. Percept. Perform. **2**, 531–537.

Macmillan, N.A., and Creelman, C.D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge, MA).

May, P., Tiitinen, H., Ilmoniemi, R.J., Nyman, G., Taylor, J.G., and Näätänen, R. (**1999**). "Frequency change detection in human auditory cortex," J. Comput. Neurosci. **6**, 99–120.

McFarland, D.J., and Cacace, A.T. (**1992**). "Aspects of short-term acoustic recognition memory: Modality and serial position effects," Audiology **31**, 342–352.

McKenna, T.M., Weinberger, N.M., and Diamond, D.M. (**1989**). "Responses of single auditory cortical neurons to tone sequences," Brain Res. **481**, 142–153.

Micheyl, C., Carlyon, R.P., Shtyrov, Y., Hauk, O., Dodson, T., and Pullvermüller, F. (**2003**). "The neurophysiological basis of the auditory continuity illusion: A mismatch negativity study," J. Cogn Neurosci. **15**, 747–758.

Moody, D.B., Cole, D., Davidson, L.M., and Stebbins, W.C. (**1984**). "Evidence for a reappraisal of the psychophysical selective adaptation paradigm," J. Acoust. Soc. Am. **76**, 1076–1079.

Moore, B.C.J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour-Luhtanen, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. (**1997**). "Language-specific phoneme representations revealed by electric and magnetic brain responses," Nature (London) **385**, 432–434.

Näätänen, R., and Winkler, I. (**1999**). "The concept of auditory stimulus representation in cognitive neuroscience," Psychol. Bull. **125**, 826–859.

Neff, D.L., and Green, D.M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.

Okada, M., and Kashino, M. (**2003**). "The role of spectral change detectors in temporal order judgment of tones," NeuroReport **14**, 261–264.

Palmer, J. (**1986**). "Mechanisms of displacement discrimination with and without perceived movement," J. Exp. Psychol. Hum. Percept. Perform. **12**, 411–421.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., and Griffiths, T.D. (**2002**). "The processing of temporal pitch and melody information in auditory cortex," Neuron **36**, 767–776.

Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.

Schröger, E. (**1997**). "On the detection of auditory deviations: a pre-attentive activation model," Psychophysiology **34**, 245–257.

Shera, C.A., Guinan, J.J., and Oxenham, A.J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.

Shu, Z.J., Swindale, N.V., and Cynader, M.S. (**1993**). "Spectral motion produces an auditory after-effect," Nature (London) **364**, 721–723.

Sussman, E., Ritter, W., and Vaughan, H.G. (**1999**). "An investigation of the auditory streaming effect using event-related brain potentials," Psychophysiology **36**, 22–34.

Tansley, B.W., and Regan, D. (**1979**). "Separate auditory channels for unidirectional frequency modulation and unidirectional amplitude modulation," Sens Processes **3**, 132–140.

Terhardt, E. (**1970**). "Oktavspreizung und Tonhöhenverschiebung bei Sinustönen," Acustica **22**, 345–351.

van Noorden, L.P.A.S. (**1975**). "Temporal coherence in the perception of tone sequences," Ph.D. dissertation (Technische Hogeschool, Eindhoven, The Netherlands).

Wakefield, G.H., and Viemeister, N.F. (**1984**). ''Selective adaptation to linear frequency-modulated sweeps: Evidence for direction-specific FM channels?,'' J. Acoust. Soc. Am. **75**, 1588–1592.

Ward, W.D., and Burns, E.M. (**1982**). ''Absolute pitch,'' in *The Psychology of Music*, edited by D. Deutsch (Academic, New York), pp. 431–451.

Warren, J.D., and Griffiths, T.D. (**2003**). ''Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain,'' J. Neurosci. **23**, 5799–5804.

Warren, J.D., Uppenkamp, S., Patterson, R.D., and Griffiths, T.D. (**2003**). ''Separating pitch chroma and pitch height in the human brain,'' Proc. Natl. Acad. Sci. U.S.A. **100**, 10038–10042.

Weinberger, N.M., and McKenna, T.M. (**1988**). ''Sensitivity of single neurons in auditory cortex to contour: Toward a neurophysiology of music perception,'' Music Percept. **5**, 355–390.

Whitfield, I.C., and Evans, E.F. (**1965**). ''Responses of auditory cortical neurons to stimuli of changing frequency,'' J. Neurophysiol. **28**, 655–672.

Wilson, J.P., Frost, D., and Wilkes, G.D. (**1994**). ''Adaptation to unidirectional sweeping spectral peaks,'' in *Advances in Hearing Research*, edited by G.A. Manley, G.M. Klump, C. Koppl, H. Fastl, and H. Oeckinghaus (World Scientific, Singapore), pp. 452–461.

Winkler, I., Tervaniemi, M., Huotilainen, M., Ilmoniemi, R., Ahonen, A., Salonen, O., Standertskjöld-Nordenstam, C.G., and Näätänen, R. (**1995**). ''From objective to subjective: Pitch representation in the human auditory cortex,'' NeuroReport **6**, 2317–2320.

Zhang, L.I., Tan, A.Y.Y., Schreiner, C.E., and Merzenich, M.M. (**2003**). ''Topography and synaptic shaping of direction selectivity in primary auditory cortex,'' Nature (London) **424**, 201–205.