What makes a melody: The perceptual singularity of pitch sequences

Marion Cousineau[a)]
Laboratoire de Psychologie de la Perception (UMR CNRS 8158), Université Paris-Descartes and Département d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, F-75230 Paris Cedex 05, France

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

Daniel Pressnitzer
Laboratoire de Psychologie de la Perception (UMR CNRS 8158), Université Paris-Descartes and Département d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, F-75230 Paris Cedex 05, France

(Received 5 January 2009; revised 25 September 2009; accepted 28 September 2009)

This study investigated the ability of normal-hearing listeners to process random sequences of tones varying in either pitch or loudness. Same/different judgments were collected for pairs of sequences with a variable length (up to eight elements) and built from only two different elements, which were 200-ms harmonic complex tones. The two possible elements of all sequences had a fixed level of discriminability, corresponding to a d' value of about 2, irrespective of the auditory dimension (pitch or loudness) along which they differed. This made it possible to assess sequence processing per se, independent of the accuracy of sound encoding. Pitch sequences were found to be processed more effectively than loudness sequences. However, that was the case only when the sequence elements included low-rank harmonics, which could be at least partially resolved in the auditory periphery. The effect of roving and transposition was also investigated. These manipulations reduced overall performance, especially transposition, but an advantage for pitch sequences was still observed. These results suggest that automatic frequency-shift detectors, available for pitch sequences but not loudness sequences, participate in the effective encoding of melodies.
 [DOI: 10.1121/1.3257206]

PACS number(s): 43.66.Mk, 43.66.Hg [BCM] Pages: 3179–3187

I. INTRODUCTION

To understand speech or appreciate music, one has to process sequences of sounds. Among the diverse auditory features that may vary between sounds of a sequence, pitch seems to have a peculiar importance for human listeners. Pitch sequences constitute melodies and play a major role in music (Dowling and Harwood, 1986). Melodies have also been the example of choice used by Gestalt psychologists to illustrate the emergence of form out of the combination of discrete elements (Wertheimer, 1924).

Do sequences of pitch truly have special perceptual properties, compared to sequences of other auditory features such as loudness or timbre? Sequences of pitch have the ability to convey contour and relative interval information (Dowling and Fujitani, 1971), a property extensively used in music. Moore and Rosen (1979) suggested that familiar melodies cannot be recognized when the pitch changes are replaced by equivalent loudness changes. However, this suggestion is at odds with the results of a more recent study by McDermott *et al.* (2008). McDermott *et al.* (2008) argued that contour perception is not specific to pitch. They found that contours in the dimension of either pitch, loudness, or timbre can be recognized cross dimensionally. They contended that pitch is particularly useful to construct melodies merely because this perceptual feature is encoded very accurately over a wide range.

McFarland and Cacace (1992) reached an opposite conclusion. They employed pure-tone sequences built from only two possible values of a given feature, separated by a fixed number of just-noticeable differences. This was intended to equate the discriminability of individual elements of the sequences, in order to reveal mechanisms of sequence processing that otherwise might have been obscured by the peculiarities of feature encoding for single sounds. An advantage for pitch sequences was found: listeners could recognize longer pitch sequences than loudness or duration sequences. There are, however, potential concerns about the experimental procedure. First, it was assumed that a fixed number of just-noticeable differences would produce equal discriminability. This hypothesis is supported by introspective data (Terhardt, 1968), but has not been confirmed by performance data. Second, a large number of elements had to be memorized on each trial, so that the task presumably recruited high-level cognitive strategies (Cowan, 2001). The important question of a possible sensory advantage for pitch-sequence processing thus remains unresolved.

In the present study, we reassessed the processing of pitch sequences and loudness sequences. Binary sequences

a)Author to whom correspondence should be addressed. Electronic mail: marion.cousineau@ens.fr

of tones were constructed and the discriminability between the two possible component tones was equated in terms of the d' sensitivity index of signal detection theory (Green and Swets, 1966). Each sequence contained at most eight elements, which were complex tones. The tones consisted of harmonics with variable ranks, so that in some sequences the spectral components of the tones were completely resolvable in the auditory periphery, whereas in other sequences the spectral components were completely unresolvable. These two kinds of complex tones convey musical pitch (Moore and Rosen, 1979), but with a different accuracy (Kaernbach and Bering, 2001) and perhaps by means of different mechanisms (de Cheveigné, 2005). We wished to determine whether the advantage of pitch over loudness suggested by McFarland and Cacace (1992) for pure tones would generalize to complex tones of either type when discriminability is equalized.

II. EXPERIMENT 1: SEQUENCES OF PITCH AND LOUDNESS

In this experiment, sequence processing was evaluated using tones differing in (1) loudness (condition L), (2) pitch produced by resolved harmonics (condition P-R), and (3) pitch produced by unresolved harmonics (condition P-U). In order to dissociate single-sound encoding from sequence processing per se, we equalized the discriminability index d' between elements of the sequences in these three conditions.

A. Method

The sequence elements (see Fig. 1) were complex tones with a fundamental frequency (F0) close to 125 Hz. The tones were bandpass-filtered click trains (eighth order Butterworth filters). The pass-band of the filter was set between 125 and 625 Hz in the P-R and L conditions, and between 3900 and 5400 Hz in the P-U condition. Identical tones had been used by Shackleton and Carlyon (1994). Based on frequency difference limens and phase sensitivity measures, Shackleton and Carlyon (1994) argued that spectral components would be resolved in the P-R condition and unresolved in the P-U condition. In order to mask distortion products that could affect the internal spectrum of the tones in the P-U condition (Pressnitzer and Patterson, 2001), the tones were mixed with pink noise in all conditions. The pink noise was generated in the spectral domain with components between 62.5Hz and half the sampling rate. For each stimulus, the overall level of the noise was set at 6 dB below the overall level of the tone; then, the sound pressure level (SPL) of the tone plus noise compound was set close to 65 dB. Each stimulus had a duration of 200 ms and was gated on and off with 25-ms raised-cosine ramps.

In a preliminary adjustment phase, listeners had to perform a same/different task on two tones with variable differences in F0 or SPL. The reference tone had an F0 of 125 Hz and a SPL of 65 dB, while the other tone was higher in either F0 or SPL. In a given block of trials, only one tone was used in addition to the reference tone; each of the two tones to be compared on a given trial was randomly chosen to be the reference or the other tone. For each condition (P-R, P-U, or L) and listener, several adjustment blocks were run to estimate the stimulus difference Δ yielding a d' value of 2. The Δ values tested as well as the number of adjustment blocks were determined heuristically by the experimenter. Note, however, that this adjustment was independently verified in the main part of the experiment.

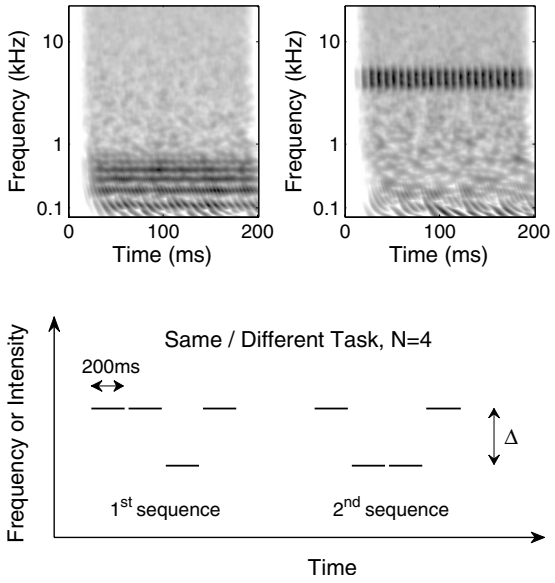

FIG. 1. Top panels: Stimuli of Experiment 1. A simulation of peripheral auditory processing (Patterson *et al.*, 1995) illustrates the predicted activation of different frequency channels after cochlear encoding. (Top-left) When the tones were filtered in a low frequency region (125–625 Hz), their spectral components were processed in independent channels and thus "resolved" (conditions P-R and L). (Top-right) When the tones were filtered in a high frequency region (3900–5400 Hz), their spectral components were unresolved (condition P-U). (Bottom panel) Experimental task. Listeners made same/different judgments on binary sequences. The elements of the sequences could take only two values of either F0 (conditions P-R and P-U) or SPL (condition L). These two values differed by Δ. The number of elements, N, could be equal to 1, 2, 4, or 8. On a given trial, only one element (chosen at random) could change. In this example of a "different" trial, N is equal to 4 and the second element changes.

In the subsequent and main part of the experiment, the previously determined values of Δ were used in binary sequences of $N=1$, 2, 4, or 8 tones. On each trial, two sequences separated by a 400-ms silence were presented (see Fig. 1, bottom panel). In the first sequence, each tone was, at random, either a reference stimulus, A, with a 125-Hz F0 and a 65-dB level, or another stimulus, B, differing positively from A by Δ. The second sequence was equiprobably identical to the first sequence or different from it. In the latter case, a single, randomly chosen tone was changed from A to B or vice versa. The listener had to make a same/different judgment. For each listener, condition, and N value, four blocks of 50 trials were run. The ordering of conditions and N values was randomized. No training was provided apart from the series of adjustment blocks, which corresponds to sequences with $N=1$.

Thirteen listeners with no self-reported hearing disorder participated in the experiment (mean age=24.0, SD=6.1, six female). A questionnaire was used to evaluate musical training. The measure used was the number of years participants had practiced a musical instrument (mean=6.9, SD=7.2). For displaying individual data, participants were sorted in

TABLE I. Means and standard deviations of the Δ values used in Experiment 1. These values are expressed as percentages of the reference frequency or sound pressure. The values in parentheses are conversions to semitones (P-R or PU conditions) or dB (L condition).

Experiment 1			
Condition	P-R	P-U	L
Reference	125 Hz	125 Hz	65 dB
Mean Δ	1.9 (0.3)	38.8 (5.7)	49.6 (3.5)
s.d.	0.9 (0.2)	29.6 (4.5)	37.3 (2.8)

three groups: no musical training (4), less than 10 years (5), and more than 10 years (4). Stimuli were generated with an RME Fireface audio sound card and digital to analog converter, with 16-bit coding accuracy and 44.1 kHz sampling rate. They were delivered diotically by means of Sennheiser HD 250 linear II headphones. Listeners were seated in a double-walled sound insulated booth (IAC). Responses were given by means of button press. No feedback was provided.

B. Results

Table I displays the mean results of the adjustment phase and shows that Δ was much higher in the P-U condition than in the P-R condition. This is consistent with many previous data showing that the pitch percepts evoked by unresolved harmonics are less precise than those evoked by resolved harmonics (see, e.g., Shackleton and Carlyon, 1994; Houtsma and Smurzynski, 1990). In the L condition, the average Δ value was 3.5 dB. If relative thresholds are compared across attributes, in terms of percent of the reference, Δ values for loudness were larger than Δ values in the two pitch conditions. A significant correlation was observed between musical expertise (years of musical training) and Δ values for the P-R condition ($r=-0.80$, $t=-4.44$, $P=0.001$). Consistent with previous reports (Micheyl *et al.*, 2006), musicians displayed smaller thresholds. Similar correlations were observed between musical expertise and Δ values in the P-U condition ($r=-0.70$, $t=-3.29$, $P=0.007$) and Δ values in the L condition ($r=-0.66$, $t=-2.90$, $P=0.014$).

The top-left panel of Fig. 2 shows the effect of N and condition on d'. For $N=1$, d' was similar in all three conditions, indicating that the preliminary adjustment phase had been successful. Performance was higher than 2, suggesting equivalent improvement between adjustment and test phase in all conditions. When N was greater, however, a large difference was observed between conditions. For conditions L and P-U, almost identical trends were obtained: performance steadily decreased as soon as N exceeded 1. For P-R, in contrast, performance was approximately constant up to $N=4$ and decreased only when N reached its highest value, 8.

A repeated-measures analysis of variance (ANOVA) ($N\times$condition) confirmed the existence of a significant interaction between the two experimental factors [$F(6,72)=4.95$, $P=0.0002$], in addition to main effects of N [$F(3,36)=12.96$, $P<0.0001$] and condition [$F(2,24)=6.45$, $P=0.0057$]. *Post-hoc* tests (Tukey's HSD) confirmed that, for P-R, performance decreased only when N varied from 4 to 8.

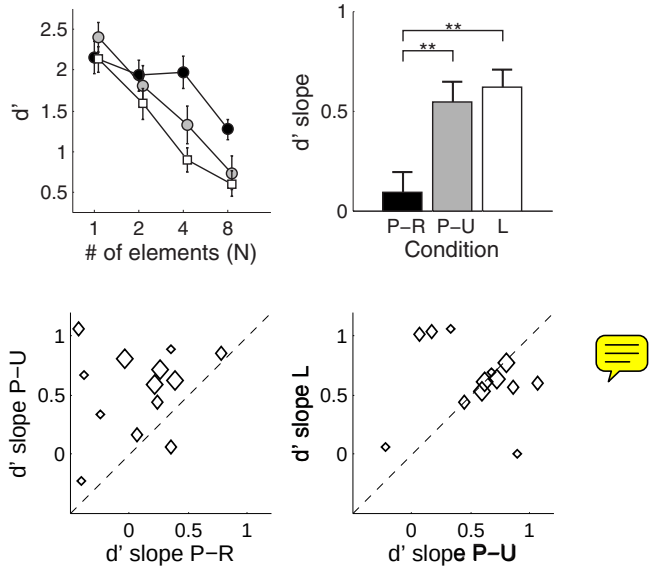

FIG. 2. Results of Experiment 1. (Top-left) Mean value of d' as a function of N in the three experimental conditions (P-R: black circles; P-U: gray circles; L: white squares). Error bars are standard errors of the means. (Top-right) Mean and standard error of the d' slope, summarizing the decrease in performance from $N=1$ to $N=4$. (Bottom-left) Individual data for the d' slope in the P-R and P-U conditions. Each diamond represents one listener. The size of the diamonds represents the musical experience of listeners: small for no experience, medium for less than 10 years, and large for more than 10 years. (Bottom-right) Same, but for the P-U and L conditions.

Figure 2 and the *post-hoc* tests indicated that, between $N=1$ and $N=4$, d' remained constant in the P-R condition but decreased regularly in the P-U and L conditions. We performed an additional analysis to summarize these patterns: Straight lines were fitted to the individual data obtained for $N=1$, 2 and 4, using a log-scale for N. The slope of the fitted lines characterizes the initial effect of N on performance, while normalizing for possible individual differences in performance at $N=1$. The top-right panel of Fig. 2 displays the results of this analysis. The slope found for P-R was small and differed significantly from those found for both P-U and L ($P<0.01$), whereas slopes for P-U and L did not differ reliably from each other ($P>0.10$). The bottom panels of Fig. 2 display individual results for the d' slope. All but one listener displayed a shallower slope for P-R than P-U, that is, an advantage for P-R sequences. Some listeners even displayed a negative slope for P-R, indicating that their performance actually increased between sequences of $N=1$ and $N=4$ elements. In contrast, results seem evenly distributed when P-U and P-L are compared.

There was no obvious effect of musical training on the magnitude of the pitch advantage, as indicated by the individual data plots. The correlation between musical expertise and the magnitude of the pitch advantage (d' slope for loudness minus d' slope for pitch) was not significant ($r=-0.20$, $t=-0.69$, $P=0.51$).

In summary, Experiment 1 demonstrated superior processing of sequences for pitch compared to loudness, but only when resolved harmonics were present.

III. IDEAL OBSERVER SIMULATION

Within the framework of signal detection theory, it is possible to simulate an ideal observer required to make

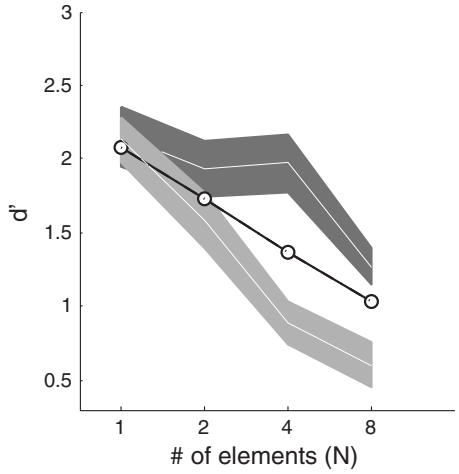

FIG. 3. Ideal observer simulation. The solid curve shows the prediction derived from an ideal observer simulation based on independent processing of single elements. The experimental data for conditions P-R (dark gray) and L (light gray), Experiment 1, are replotted as shaded areas encompassing two standard errors of the mean.

same/different judgments on two elements (Green and Swets, 1966; Dai *et al.*, 1996). Observation variables are associated with each element, taking into account internal noise. If the two observations are independent, the optimal strategy is to compare each of them with a fixed criterion to determine which distribution they belong to and then take the same/different decision (Dai *et al.*, 1996).

Our experimental task was not a comparison between two elements, but rather a comparison between sequences comprising several elements. If no specific sequence-processing mechanism is assumed, optimal performance on this task may be simulated by repeating the ideal observer's strategy on corresponding elements across sequences, under the assumption that the elements are encoded independently of each other and with perfect memory.

We implemented such a model to quantify performance for N independent comparisons. Binary sequences of A and B elements were generated as in Experiment 1. Each element was transformed into its observation variable by combining its true value, 0 or 1, with a Gaussian noise of zero mean and standard deviation σ. No bias was assumed, so the criterion was set to 0.5 (note that changing the criterion would not affect the obtained d'). For each element, the observation variable was compared to the criterion and an A or B decision was taken. All corresponding pairs of observed elements, matched across the two sequences, were then compared and a "same" decision was taken if all pairs were identical. The magnitude of the noise, σ, was the only free parameter of the model. It was fitted to the experimental data for $N=1$.

For each value of N, we computed d' from 100 000 simulated trials. The simulation results are displayed in Fig. 3, together with the data of Experiment 1 (for simplicity, we omit condition P-U). As N increased, the predicted d' decreased less rapidly than in condition L but more rapidly than in condition P-R. Further simulations showed that changing σ would shift overall performance but would not change the predicted slope of the function relating d' to N. We compared, by means of student t tests, the d' slopes of the experimental data (Fig. 2) to the predicted slope. It appeared that the model significantly outperformed listeners for condition L but was significantly outperformed by listeners for condition P-R ($P<0.025$, one-tailed tests).

That the ideal observer model outperformed listeners in condition L can be easily accounted for, by assuming for instance that listeners were unable to achieve optimal performance because their memory was not perfect. In condition P-R, however, listeners were more efficient than the ideal observer. This seemingly paradoxical finding shows that the most important assumption of the model does not hold: the high performance of listeners in the P-R condition cannot be based on independent processing of the elements of the sequences. Instead, listeners had to use sequence-specific mechanisms in this condition.

TABLE II. Means and standard deviations of the Δ values used in Experiment 2. These values are expressed as percentages of the reference frequency. The values in parentheses are conversions to semitones.

Experiment 2			
Condition	P-F0high	P-F0mid	P-F0low
Reference	250 Hz	125 Hz	62.5 Hz
Mean Δ	5.8 (1.0)	9.6 (1.6)	32.3 (4.9)
s.d.	9.2 (1.5)	9.6 (1.6)	22 (3.4)

IV. EXPERIMENT 2: EFFECT OF FREQUENCY REGION AND HARMONIC RESOLVABILITY

In Experiment 1, complex tones were filtered in different frequency regions to produce resolved or unresolved harmonics. Resolvability was therefore confounded with frequency region, which could have influenced pitch processing by itself (Meddis and O'Mard, 1997; Pressnitzer *et al.*, 2001). Experiment 2 controlled for this possibility by manipulating resolvability within a fixed frequency region.

A. Method

As in Experiment 1, band-pass filtered click trains were used. The pass-band was fixed here between 1375 and 1875 Hz. There were three pitch conditions, with reference F0s of 250, 125, and 62.5 Hz. They were, respectively, termed P-F0high, P-F0mid, and P-F0low. Again, our choice of the stimulus parameters was motivated by observations of Shackleton and Carlyon (1994). They found low frequency difference limens for P-F0high, suggesting that spectral components would be resolved in this case. In contrast, difference limens were poor for P-F0low, suggesting that spectral components would be unresolved in this case. Results for P-F0mid were intermediate, with some variability across subjects.

The tones were mixed with pink noise and all other experimental details were as in Experiment 1, except that here only two values of N were used: 1 and 4. The Δ values are displayed in Table II. Five listeners with no self-reported hearing disorder participated (mean age=24.4, SD=3.4, three female). Musical training was evaluated as in Experiment 1 and quantified by the number of years of musical

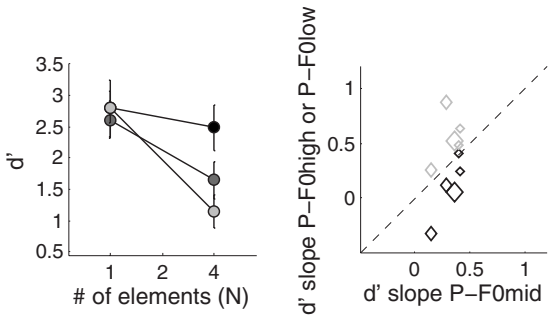

FIG. 4. Results of Experiment 2. (Left) Mean and standard error of d' for each F0 condition as a function of N (P-F0high: black circles; P-F0mid: dark gray circles; P-F0low: light gray circles). (Right) Individual data for the d' slope in the P-F0mid condition, plotted against d' slope for P-F0high (black diamonds) or d' slope for P-F0low (light gray diamonds). The size of the markers indicates musical experience as in Fig. 2.

practice (mean=4.8, SD=6.2; two participants with no musical training, two with less than 10 years, and one with more than 10 years).

B. Results and discussion

The left panel of Fig. 4 displays the mean results. A repeated-measures ANOVA revealed a significant interaction between N and F0 [$F(2,8)=10.77$, $P=0.0053$], in addition to main effects of N [$F(1,4)=17.84$, $P=0.013$] and F0 [$F(2,8)=6.19$, $P=0.023$]. The hypothesis tested by the experiment was that performance would be degraded in the case of unresolved harmonics. To test for this, *post-hoc* analyses (Fisher LSD) were performed. For $N=4$, performance was significantly better for P-F0high than for P-F0low ($P=0.00018$); performance for P-F0mid was intermediate, significantly poorer than for P-F0high ($P=0.0035$) and significantly better than for P-F0low ($P=0.041$). No significant differences were found across conditions for $N=1$. It can also be seen in Fig. 4 that when N varied from 1 to 4, performance was essentially unchanged for P-F0high but decayed in the other two conditions.

Individual d' slopes for this experiment are shown in the right panel of Fig. 4. For most listeners, the d' slopes were larger for P-F0mid than for P-F0high, and smaller for P-F0mid than for P-F0low. There was no obvious effect of musicianship on the pattern of results.

In Experiments 1 and 2, performance with pitch sequences appears to be determined by the rank of the lowest harmonic present in the pass-band of the stimulus (R_{lh}) rather than by the absolute frequency region. Poor performance was observed in Experiment 1 for a high frequency region (lower limit: 3.9 kHz) and a R_{lh} of 32. A similarly poor performance was observed in Experiment 2 for a medium frequency range (lower limit: 1.375 kHz) but a R_{lh} which was again very high (22). In Experiment 2, performance was good even when relatively high harmonics were used (P-F0high condition, $R_{lh}=6$). Estimates of resolution based on hearing out individual harmonics would fail with such stimuli (Plomp, 1964; Moore, 1973). However, a large number of studies based on frequency difference limens find a transition region from good to poor performance between the 10th and 13th harmonics (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003; de Cheveigné and Pressnitzer, 2006). It is therefore likely that at least some of the harmonics of the P-F0high condition were resolved. For the P-F0mid condition, R_{lh} was 11, which is within the transition region for estimates of resolvability based on difference limens (Shackleton and Carlyon, 1994). The small but significant pitch advantage that we observed in this condition could be due to partial resolvability of the lower harmonics of the stimuli.

Moore and Rosen (1979) showed that listeners are able to recognize familiar melodies produced by varying the pitch of unresolved complexes. The condition in which they observed best performance (2 kHz highpass filtering and F0 between 100 and 200 Hz) resembles the P-F0mid condition of the present experiment, for which we also found evidence of an advantage for pitch sequences. In another condition which resembles the P-U condition of Experiment 1 (4 kHz highpass filtering), Moore and Rosen (1979) found that recognition performance degraded significantly. The two sets of results are thus consistent if one considers the different possible definitions of resolvability: pitch sequences may be processed efficiently when composed of tones with harmonics that cannot be heard out individually (Plomp, 1964) but for which there is evidence of at least partial peripheral resolvability (Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003).

In summary, here and in Experiment 1, performance with pitch sequences was better when at least some of the harmonics of the stimuli were partially resolved than when all harmonics were clearly unresolved.

V. EXPERIMENT 3: EFFECT OF FREQUENCY AND AMPLITUDE ROVE

The previous experiments used the same two elements for all sequences, which were presented in blocks of 50 contiguous trials. It is conceivable that listeners could have memorized the absolute values of each attribute within the block, which could in turn interfere with sequence processing. We tested for this possibility by applying a random rove on the dimension tested within each block. From trial to trial, the reference value of F0 for pitch sequences and SPL for loudness sequences was randomly varied. The two sequences within a trial still had identical reference tones, as in Experiment 1.

A. Method

The sequences were similar to those of Experiment 1. However, only conditions P-R and L were used and the maximum value of N was 4. We chose to adjust the Δ value for each condition and listener without any rove (the Δ values are displayed in Table III). Then, a first series of four blocks of 50 trials was run with the Δ value selected and $N=1$, for each listener and condition, in order to check for equal discriminability without rove. In subsequent blocks, roving was applied. For pitch sequences, the reference F0 on a given trial was randomly chosen between 125 Hz and 125 Hz+2Δ. For loudness sequences, the reference SPL was randomly chosen between 65 dB and 65 dB+2Δ. The

TABLE III. Means and standard deviations of the Δ values used in Experiments 3 and 4. These values are expressed as percentages of the reference frequency or sound pressure. The values in parentheses are conversions to semitones (P-R condition) or dB (L condition).

Experiments 3 and 4		
Condition	P-R	L
Reference	125 Hz	65 dB
Mean Δ	1.27 (0.2)	28.16 (2.2)
s.d.	0.38 (0.1)	6.74 (0.6)

amount of roving was thus aimed to be comparable across attributes and listeners. We measured discriminability for roved sequences with $N=1$, 2, and 4 elements in interleaved blocks of 50 trials (four blocks per condition and listener). Six listeners with no self-reported hearing impairment participated (mean age=24.0, SD=2.2, four female). Musical training was evaluated as in Experiment 1 and quantified by the number of years of musical practice (mean=9.7, SD =8.5; one participant with no musical training, three with less than 10 years, and two with more than 10 years). All listeners also participated in Experiment 4 (described below). Half of the subjects started by Experiment 3 and the other half by Experiment 4.

B. Results and discussion

The left panel of Fig. 5 displays the mean results. The preliminary adjustment phase was satisfactory: a paired sample t-test revealed no significant difference between P-R and L for $N=1$, no rove ($P=0.5$). For the data with rove, performance was globally lower than in Experiment 1, where no rove was applied. The pattern of the results, however, was highly similar, with an advantage of P-R over L as sequences increased in length. A repeated-measures ANOVA revealed a significant effect of N [$F(2,10)=6.86$, $P=0.003$] and a significant interaction between N and condition [$F(2,10)=7.29$, $P=0.01$]. The right panel of Fig. 5 displays the individual d' slopes. All but one listener displayed a pitch advantage. The largest pitch advantage was observed for a participant with no musical training, but large effects were also observed for musicians.

It is not surprising that roving should degrade overall performance. Discrimination tasks involving pitch (Harris, 1948; Demany and Semal, 2005; Ahissar *et al.*, 2006) or loudness (Oxenham and Buus, 2000) always result in poorer performance with roving than without. Many factors may explain this finding, including perceptual learning (Demany and Semal, 2002) or stimulus uncertainty (Watson *et al.*, 1976). Importantly, in our case, performance was equally degraded by roving for pitch and loudness. As a result, roving did not change the main features of the results and the advantage for pitch sequences was replicated. In Experiment 1, therefore, the possibility of memorizing the sequence elements across trials was not the cause of the observed advantage of pitch sequences over loudness sequences.

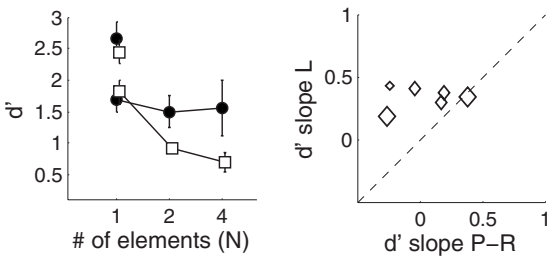

FIG. 5. Results of Experiment 3. (Left) Mean and standard error of d' in conditions P-R (black circles) and L (white squares) as a function of N, with rove. The disconnected symbols show mean and standard error of d' for N=1, with the same Δ values but without rove. (Right) Individual data for the d' slope in the P-R condition, plotted against d' slope for L. The size of the markers indicates musical experience as in Fig. 2.

VI. EXPERIMENT 4: EFFECT OF TRANSPOSITION

In all experiments reported so far, the two sequences presented on a given trial were built from the same pair of tone elements. Absolute cues for each attribute were thus potentially available within a trial. In musical situations, by contrast, melodies may be transposed; that is, the same sequence of pitch intervals may be presented with different starting frequencies (Dowling and Harwood, 1986). Listeners are able to recognize transposed melodies, but the task becomes increasingly difficult if the melody is unfamiliar and if contour cues are not affected (Dowling and Fujitani, 1971; Kidd and Watson, 1996). Recently, it has also been shown that listeners are able to recognize transposed loudness sequences (McDermott *et al.*, 2008). In this final experiment, we introduce transpositions in the pitch or the loudness domain, in order to investigate whether the pitch-sequence advantage remains with transposed material.

There is, however, a fundamental limitation with introducing transposition in our task: a same/different task with only one element and transposition is not possible. It is therefore impossible to control for performance at $N=1$, which was a strong prerequisite for each of the experiments reported above.

A. Method

The stimuli were similar to those of Experiments 1 and 3, but N had only two possible values: 2 and 4. Only conditions P-R and L were tested. As mentioned above, the subjects were the same as in Experiment 3. Δ also had the same values as in Experiment 3 (displayed in Table III). In the first sequence presented on each trial, the reference tone had an F0 of 125 Hz and a SPL of 65 dB. In the second sequence, the reference F0 was changed to 125 Hz+2Δ for condition P-R and the reference SPL was changed to 65 dB+2Δ for condition L. Again, this was aimed at introducing comparable amounts of transposition across listeners and conditions.

B. Results and discussion

Mean results are shown in the left panel of Fig. 6. Transposition clearly had a deleterious effect since performance was generally quite poor: d' was always close to 0.5, except when N was equal to 4 in condition P-R. A repeated-measures ANOVA revealed that the main effects of condition

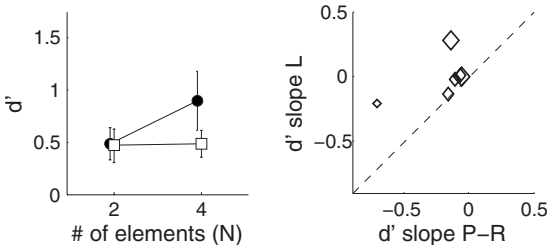

FIG. 6. Results of Experiment 4. (Left) Mean and standard error of d' in conditions P-R (black circles) and L (white squares) as a function of N. (Right) Individual data for the d' slope in the P-R condition, plotted against d' slope for L. The size of the markers indicates musical experience as in Fig. 2.

and N were not significant, but that the interaction between the two factors was marginally significant [$F(1,5)=5.5$, $P=0.06$]. *Post-hoc* tests (Fisher LSD) indicated that performance was significantly better in condition P-R than in condition L for $N=4$ ($P=0.017$), whereas there was no significant difference between the two conditions for $N=2$ ($P=0.87$).

Individual data are shown on the right panel of Fig. 6. Only two listeners out of six, a musician and a non-musician, showed a strong pitch advantage. For the other four listeners, the d' slopes were generally small. However, all six listeners consistently showed negative slopes in the pitch condition, which was not the case for loudness.

The fact that transposition made the comparison between pitch sequences more difficult is in line with previous data obtained with unfamiliar melodies (Dowling and Fujitani, 1971). Moreover, Kidd and Watson (1996) reported that the adverse effect of transposition was already substantial when the amount of transposition was modest and the transposition interval was constant from trial to trial, as in the present experiment. We also observed that longer sequences ($N=4$) showed improved performance compared to short sequences ($N=2$), corresponding to negative d' slopes. This is again consistent with results of Kidd and Watson (1996), who found better performance for changes embedded in sequences of five elements than for changes embedded in sequences of two elements. They interpreted this finding by remarking that a change in the pitch of a single tone modifies two consecutive pitch intervals if and only if the tone that changes is surrounded by two other tones. When only relative information is relevant (which is the case in the presence of a transposition), this predicts an increase in performance for sequences with more than two elements.

In our L condition, performance was close to the chance level for each value of N. This suggests that contour cues are not available for the discrimination of loudness sequences. McDermott *et al.* (2008) suggested the opposite, but in their study the loudness changes taking place from tone to tone were larger than in our study. It seems likely that human listeners are sensitive to *relative* loudness only for rather large loudness changes.

VII. GENERAL DISCUSSION

The following conclusions can be drawn from our data: (1) pitch sequences are processed more efficiently than loudness sequences, (2) the discriminability of pitch sequences is better than predicted from independent processing of their individual elements, (3) the perceptual advantage of pitch sequences over loudness sequences requires that resolved harmonics be available, (4) this advantage does not depend on the familiarity of the sequence elements, and (5) transposed sequences are less efficiently compared than non-transposed sequences, but transposition does not abolish the advantage of pitch sequences for sequences of more than two elements.

A first possible explanation for the advantage of pitch sequences would be that pitch has access to a specific and efficient short-term memory store. However, one key aspect of our data is not consistent with this interpretation. Here, resolved harmonics were required to support high performance, whereas short-term pitch memory seems to have the same temporal characteristics for resolved and unresolved harmonics when differences are adjusted in discriminability (Clément *et al.*, 1999). Also, listeners outperformed the ideal observer model which had perfect memory.

Another explanation is that superior performance for the pitch of resolved harmonics could be caused by a greater familiarity with such sequences compared to sequences of unresolved harmonics or loudness. In this case, musically experienced listeners who have received extensive training with pitch sequences should show a larger advantage. An effect of musicianship was observed in the initial discriminability adjustment step, but it was present for all dimensions tested. In the main experiment, when the discriminability of individual elements was factored out, musicians did not display any increased advantage for pitch-sequence processing.

A more likely explanation is that the advantage is based on the recruitment of an additional mechanism to encode pitch sequences. There are different strands of evidence for the existence of such a mechanism. Brain imaging studies showed that secondary auditory regions in the right hemisphere respond more strongly to melodies with pitch changes than to sequences of tones with a fixed pitch (Patterson *et al.*, 2002; Hyde *et al.*, 2008). Neuropsychological studies have shown that lesions lateralized in the right auditory cortex can impair the sensitivity to frequency-shift direction without impairing absolute frequency discrimination (Johnsrude *et al.*, 2000). Behavioral data indicate that listeners can consciously perceive an upward or downward pitch shift between two consecutive pure tones even when they did not hear out the first tone because it was fused in a complex chord (Demany and Ramos, 2005; Demany *et al.*, 2008). The latter result has been interpreted as evidence for automatic frequency-shift detectors (FSDs).

A mechanism based on FSDs can account for most of our findings. First, since FSDs detect shifts in frequency, they should not be activated by amplitude shifts, consistent with the results of Experiment 1. Second, the fact that FSDs identify by definition a *relation* between tonal elements is consistent with the refutation, by our ideal observer simulation, of the element-independence assumption for pitch sequences consisting of resolved harmonics. Third, behavioral evidence for the FSDs only exists for *spectral* shifts, that is,

shifts of resolved spectral components. The poorer performance we observed with stimuli consisting of unresolved components suggests that there are no equivalent periodicity-shift detectors (see also Demany and Semal, 2008).

The FSD hypothesis may provide a low-level basis for the long-standing observation that contour is an essential cue to melody recognition (White, 1960; Dowling and Fujitani, 1971). There are differences between our task and realistic musical situations, however. Because we aimed at equating strictly all aspects of the task that were not directly related to sequence processing, a single interval between elements of the sequences was used in all conditions and this interval was close to threshold. Within these constraints, we found evidence for a contour-extraction mechanism in pitch sequences only. Consistent with previous findings (Kidd and Watson, 1996), we also observed that even a modest amount of transposition had a strong deleterious effect on sequence discrimination. This may seem contradictory with a relative, contour-extraction mechanism, as well as with recent observations that pitch and loudness sequences can be recognized after transposition (McDermott *et al.*, 2008). McDermott *et al.* (2008) used acoustic changes not matched in discriminability and much larger than the ones used in our experiments. Using smaller changes, Moore and Rosen (1979) failed to find any evidence for contour recognition with loudness sequences. All of these observations may be reconciled if one assumes two distinct steps in any contour-matching task: (1) a sensory encoding stage, where absolute and relative cues may be pooled, followed by (2) a decision stage. If transposition introduces a fixed amount of noise in the decision stage, consistent with the observation of Kidd and Watson (1996) that transposition produces essentially the same impairment over a wide range (from 2 to 12 semitones), then this noise will swamp small differences at the sensory encoding stage, as in our experiment, but it will be overcome by large differences, as investigated by McDermott *et al.* (2008) and as used in realistic musical melodies. To test for this hypothesis, further experiments are required where the steps on each dimension are larger than in the present series of experiments, but still controlled for equal discriminability across dimensions.

Sensory encoding of pitch contour by means of FSDs may be one of the several mechanisms that "make" a melody, especially as it may occur without attention (Demany and Ramos, 2005). This hypothesis has important implications for hearing-impaired listeners. Cochlear implant users, for instance, hear through a device that directly stimulates the auditory nerve, but with a limited number of frequency channels that, to date, cannot transmit resolved harmonics. While speech intelligibility in quiet can be high, music listening is more challenging. In particular, melody recognition is poor (Kong *et al.*, 2004; Pressnitzer *et al.*, 2005; Cooper *et al.*, 2008). Providing the necessary cues for automatic contour encoding may help to improve music perception with such devices.

ACKNOWLEDGMENTS

We wish to thank Josh H. McDermott and Brian C. J. Moore for suggestions that led to the design of Experiments 3 and 4. Portions of these results were presented at the 155th Meeting of the Acoustical Society of America.

Ahissar, M., Lubin, Y., Putter-Katz, H., and Banai, K. (**2006**). "Dyslexia and the failure to form a perceptual anchor," Nat. Neurosci. **9**, 1558–1564.
Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?" J. Acoust. Soc. Am. **113**, 3323–3334.
Clément, S., Demany, L., and Semal, C. (**1999**). "Memory for pitch versus memory for loudness," J. Acoust. Soc. Am. **106**, 2805–2811.
Cooper, W. B., Tobey, E., and Loizou, P. C. (**2008**). "Music perception by cochlear implant and normal hearing listeners as measured by the Montreal battery for evaluation of amusia," Ear Hear. **29**, 618–628.
Cowan, N. (**2001**). "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," Behav. Brain Sci. **24**, 87–114.
Dai, H., Versfeld, N. J., and Green, D. M. (**1996**). "The optimum decision rules in the same-different paradigm," Percept. Psychophys. **58**, 1–9.
de Cheveigné, A. (**2005**). "Pitch perception models," in *Pitch, Neural Coding and Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York), pp. 169–233.
de Cheveigné, A., and Pressnitzer, D. (**2006**). "The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction," J. Acoust. Soc. Am. **119**, 3908–3918.
Demany, L., and Ramos, C. (**2005**). "On the binding of successive sounds: perceiving shifts in nonperceived pitches," J. Acoust. Soc. Am. **117**, 833–841.
Demany, L., and Semal, C. (**2002**). "Learning to perceive pitch differences," J. Acoust. Soc. Am. **111**, 1377–1388.
Demany L. and Semal C. (**2005**). "The slow formation of a pitch percept beyond the ending time of a short tone burst," Percept. Psychophys. **67**, 1376–1383.
Demany, L.. and Semal, C. (**2008**). "The role of memory in auditory perception," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 77–113.
Demany, L., Trost, W., Serman, M., and Semal, C. (**2008**). "Auditory change detection—Simple sounds are not memorized better than complex sounds," Psych. Sci. **19**, 85–91.
Dowling, W. J., and Fujitani, D. S. (**1971**). "Contour, interval, and pitch recognition in memory for melodies," J. Acoust. Soc. Am. **49**, 524–531.
Dowling, W. J., and Harwood, D. L. (**1986**). *Music Cognition* (Academic, Orlando, CA).
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Harris, J. (**1948**). "Discrimination of pitch: Suggestions toward method and procedure," Am. J. Psychol. **61**, 309–322.
Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.
Hyde, K. L., Peretz, I., and Zatorre, R. J. (**2008**). "Evidence for the role of the right auditory cortex in fine pitch resolution," Neuropsychologia **46**, 632–639.
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.
Kaernbach, C., and Bering, C. (**2001**). "Exploring the temporal mechanism involved in the pitch of unresolved harmonics," J. Acoust. Soc. Am. **110**, 1039–1048.
Kidd, G. R., and Watson, C. S. (**1996**). "Detection of frequency changes in transposed sequences of tones," J. Acoust. Soc. Am. **99**, 553–566.
Kong, Y. Y., Cruz, R., Jones, J. A., and Zeng, F. G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear Hear. **25**, 173–185.
McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (**2008**). "Is relative pitch specific to pitch?" Psych. Sci. **19**, 1263–1271.
McFarland, D. J., and Cacace, A. T. (**1992**). "Aspects of short-term acoustic recognition memory: Modality and serial position effects," Audiology **31**, 342–352.
Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.
Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (**2006**). "Influence of musical and psychoacoustical training of pitch discrimination," Hear. Res. **219**, 36–47.
Moore, B. C. J. (**1973**). "Some experiments relating to the perception of complex tones," Q. J. Exp. Psychol. **25**, 451–475.

Moore, B. C. J., and Rosen, S. M. (**1979**). "Tune recognition with reduced pitch and interval information," Q. J. Exp. Psychol. **31**, 229–240.
Oxenham, A. J., and Buus, S. (**2000**). "Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions," J. Acoust. Soc. Am. **107**, 1605–1614.
Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.
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 analyser," J. Acoust. Soc. Am. **36**, 1628–1636.
Pressnitzer, D., Bestel, J., and Fraysse, B. (**2005**). "Music to electric ears: Pitch and timbre perception by cochlear implant patients," Ann. N. Y. Acad. Sci. **1060**, 343–345.
Pressnitzer, D., and Patterson, R. D. (**2001**). "Distortion products and the pitch of harmonic complex tones," in *Physiological and Psychophysical Bases of Auditory Function*, edited by A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht, The Netherlands), pp. 97–104.
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.
Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.
Terhardt, E. (**1968**). "Über ein äquivalenzgesetz für intervalle akustischer empfindungsgrössen [On an equivalence rule for intervals between the magnitudes of acoustic sensations]," Kybernetik **5**, 127–133.
Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am. **60**, 1176–1186.
Wertheimer, M. (**1924**). *Gestalt Theory* (Hayes Barton, Raleigh, NC).
White, B. W. (**1960**). "Recognition of distorted melodies," Am. J. Psychol. **73**, 100–107.