Perception & Psychophysics
2003, 65 (7), 1019-1028

Time course of melody recognition: A gating paradigm study

SIMONE DALLA BELLA, ISABELLE PERETZ, and NEIL ARONOFF
University of Montreal, Montreal, Quebec, Canada

Recognizing a well-known melody (e.g., one's national anthem) is not an all-or-none process. Instead, recognition develops progressively while the melody unfolds over time. To examine which factors govern the time course of this recognition process, the gating paradigm, initially designed to study auditory word recognition, was adapted to music. Musicians and nonmusicians were presented with segments of increasing duration of familiar and unfamiliar melodies (i.e., the first note, then the first two notes, then the first three notes, and so forth). Recognition was assessed after each segment either by requiring participants to provide a familiarity judgment (Experiment 1) or by asking them to sing the melody that they thought had been presented (Experiment 2). In general, the more familiar the melody, the fewer the notes required for recognition. Musicians judged music's familiarity within fewer notes than did nonmusicians, whereas the reverse situation (i.e., musicians were slower than nonmusicians) occurred when a sung response was requested. However, both musicians and nonmusicians appeared to segment melodies into the same perceptual units (i.e., motives) in order to access the correct representation in memory. These results are interpreted in light of the cohort model (Marslen-Wilson, 1987), as applied to the music domain.

In everyday life, listeners rapidly switch from one radio station to the other, assessing within a few seconds whether or not they are hearing their preferred song. Similarly, in the radio game "Name-That-Tune," when asked to recognize popular musical excerpts, contestants can often provide a correct guess within a few notes. From these examples, recognizing well-known melodies may appear to be a fairly automatic activity that can be performed with little effort and with a minimal amount of information. However, despite its apparent simplicity, music recognition involves several cognitive operations. In order for a melody to be recognized, a structural representation of the presented musical stimulus has to be computed on the basis of its perceptual features. Then this representation has to be matched to a representation stored in memory. Recognition occurs when there is a correspondence between the perceptual and the stored representations. The study of the musical features involved in this matching process has received particular attention. For instance, researchers have shown that changes in melodic contour (i.e., the sequence of ups and downs in a melody, regardless of interval size; Dowling & Fujitani, 1971; Dowling & Hollombe, 1977; Idson & Massaro, 1978; Kallman & Massaro, 1979) and in tone chroma (i.e., the position of the notes relative to the musical octave; Idson & Massaro, 1978; Kallman & Massaro, 1979; for a different view, see Deutsch, 1972) are both detrimental to recognition. The temporal dimension is less important for recognition than the pitch dimension is (Hébert & Peretz, 1997; White, 1960).

Surprisingly, none of these studies has attempted to explore the dynamic *process* of melody recognition. Therefore, there is currently no model that explains which properties contribute to the recognition process while the melody unfolds over time. One model that has proved useful in auditory word recognition is the cohort model (Marslen-Wilson, 1987; Marlsen-Wilson & Welsh, 1978; for reviews, see Frauenfelder, 1996; Protopapas, 1999). This could serve as a good departure point for modeling the time course of melody recognition. In the most recent version of the cohort model (Marslen-Wilson, 1987), the overall process of word recognition is divided into three stages: access, selection, and integration. In the access stage, the first 100–150 msec of the incoming acoustic signal (i.e., the beginning of the word) activate a series of potential lexical candidates by multiple access. This set of candidates is referred to as the *word-initial cohort* and is determined solely on the basis of bottom-up information deriving from the sensory input. In the selection stage, the cohort is progressively reduced with the increase of available acoustic information about the word. During this discrimination phase, the level of activation of the candidates is raised or lowered depending on their compatibility with the sensory input, until the best-fitting match is selected. Recognition is achieved when the most activated candidate surpasses a threshold value, as compared with its nearest lexical competitors. Finally, the integration stage im-

The research was supported by a doctoral scholarship from the Groupe de Recherche en Neuropsychologie Expérimentale (University of Montreal) to S.D.B., a fellowship from the Fonds Concertés de l'Aide à la Recherche to N.A., and a grant from the Natural Science and Engineering Research Council of Canada to I.P. We thank Glenn Schellenberg, Bruno Repp, and one anonymous reviewer for helpful comments on an earlier draft of this article. Correspondence concerning this article should be addressed to S. Dalla Bella, Department of Psychology, Ohio State University, 1885 Neil Ave., Columbus, OH 43210 (e-mail: dalla-bella.2@osu.edu).

 Copyright 2003 Psychonomic Society, Inc.

plies that the syntactic and semantic information of the recognized word is integrated within the higher level representation of the utterance.

Based on the cohort model, the gating paradigm introduced by Grosjean (1980; for a review, see Grosjean, 1996) has represented one of the major procedures used to study the time course of spoken word recognition (see Marslen-Wilson, 1987). Participants are presented with segments of a word of increasing duration (for instance, the first 30 msec, then the first 60 msec, and so forth). At each presentation, the length of the acoustic segment is increased by a constant (e.g., 30 msec). Then, after each presentation, the participants have to guess the identity of the word, and they are requested to indicate how confident they are about their guess (e.g., on a scale from *very sure* to *very unsure*). The point at which a word is identified is estimated in two ways. The isolation point (IP, measured in milliseconds or percentage of the word) is the point at which the word is correctly identified, without any change in response thereafter. The term *isolation* refers to the selection of a unique element within a cohort. Isolation does not imply that the participants are totally confident about their guess. Indeed, maximum confidence is reached at the recognition point (RP, also called the *total acceptance point*, measured in milliseconds or percentage of the word), which is the point at which the participants are certain of having identified the word. Hence, IP does not necessarily coincide with RP.

The usefulness of the gating paradigm for uncovering the time course of spoken word recognition has been largely validated by replicating well-established effects formerly obtained with other psycholinguistic paradigms (for a review, see Grosjean, 1996). A robust phenomenon that has been successfully and consistently obtained with the gating paradigm is the word frequency effect (Grosjean, 1980; Tyler, 1984; Walley, Michela, & Wood, 1995). High-frequency words are recognized more rapidly than low-frequency words, in that their IP and RP occur earlier. This phenomenon is of particular interest to us, insofar as an analogous effect is likely to occur when recognition of highly familiar and moderately familiar music is compared, as will be clarified below.

Despite the successful application of the gating paradigm in psycholinguistic research, it has been argued that the unnatural sequential format of presentation and the fact that participants have unlimited response time may induce a task-specific strategic behavior. Hence, the gating paradigm might provide an unrealistic image of the word recognition process. However, both objections have been discarded. The results of the original study (Grosjean, 1980) were replicated when each participant was presented with only one fragment of a word (Cotton & Grosjean, 1984; Walley et al., 1995) and when participants were encouraged to respond as rapidly as possible (Tyler & Wessels, 1985). Thus, neither the presentation format nor the unlimited response time appears to detract from the validity of the gating paradigm.

Our main purpose in the present study was to explore the time course of the melody recognition process through the gating paradigm. This procedure appears to be well suited for the study of music recognition. Gated presentation of music has been previously adopted in music cognition research (e.g., Bigand & Parncutt, 1999; Palmer & Krumhansl, 1987a, 1987b; Peretz, Gagnon, & Bouchard, 1998). Moreover, gated presentation has been successfully applied to the recognition of acoustical properties of musical recordings (Schellenberg, Iverson, & McKinnon, 1999). Therefore, on the basis of these previous applications to the study of musical phenomena, the gating paradigm appears to be a good choice for studying the process by which well-known tunes are recognized.

A second aim of our study was to examine whether the time course of music recognition is affected by musical training and by music familiarity. Training and familiarity are major factors that typically influence music processing (for reviews, see Dowling & Harwood, 1986; Smith, 1997). For example, musicians perform better than nonmusicians on many musical tasks, such as recognizing the exact transposition of a melody (e.g., Bartlett & Dowling, 1980; Halpern, Bartlett, & Dowling, 1995), detecting whether two melodies are the same or different (e.g., Dowling, 1978, 1984), or rating to what degree a certain tone fits with an established musical key (e.g., Krumhansl & Shepard, 1979; for a review, see Krumhansl, 1990). Similarly, performance in musical tasks is usually better with familiar than with unfamiliar material. Familiarity improves nonmusicians' performance, for instance, when they are asked to transpose a melody (Attneave & Olson, 1971), recognize the exact transposition of a melody (Bartlett & Dowling, 1980), or categorize musical intervals (Smith, Nelson, Grohskopf, & Appleton, 1994). In sum, the prominence of the effects of training and familiarity on music processing indicates that these factors may influence the time course of music recognition as well.

A further objective was to explore the possibility that structurally salient units inherent to the melody itself govern the recognition process. A melody is unlikely to be perceived in a note-by-note fashion; instead, grouping processes are thought to occur (Deliège, 1987; Lerdahl & Jackendoff, 1983; Peretz, 1989; for a review, see Handel, 1989). Listeners segment a musical sequence according to its surface properties (e.g., pitch, intensity, duration, and so forth), following the bottom-up principles laid down by Gestalt psychologists. A change in any of these parameters leads to the perception of a break in the sequence and to the creation of groups separated by the changes (i.e., at the group boundaries). These groups are typically referred to as *motives*. Clustering notes into motives allows the storing of more information in working memory, since the events can be grouped into perceptual units that are larger than the individual notes. A typical motive is about three to four notes long and is considered to be the basic segmentation unit of a musical sequence (e.g., Lerdahl & Jackendoff, 1983). Since motivic structure has been found to affect the way musicians and nonmusicians segment melodies (Deliège, 1987; Peretz, 1989), it might also play a relevant role in memory access during the recognition process.

For these purposes, unfamiliar and familiar synthesized tunes taken from the traditional song repertoire were played with gated presentation to musicians and nonmusicians. Familiar tunes were divided into highly familiar and moderately familiar in order to parallel the opposition between frequent and less frequent words in psycholinguistics research. Instead of using fixed duration gates, as adopted in Grosjean's (1980) original study, we employed a variable size gate corresponding to the duration from onset to offset of each note in the melody. Indeed, the variability in note duration and tempi from one melody to the next makes the use of a fixed gate inadequate. For instance, choosing the shortest note gate would result in tediously long experimental sessions, whereas larger gates would lack precision. Moreover, it is likely that the number of notes is more relevant to melody recognition than is the duration of the segment. In Experiment 1, participants were requested to judge whether the tunes were familiar or unfamiliar. This judgment reflects the general feeling of knowing evoked by the presented tune without necessarily identifying it—for instance, by providing its title. The feeling-of-knowing judgment is treated as an estimate of the retrievability of the melody from memory (see, e.g., Koriat, 1993; Koriat & Levy-Sadot, 2001) and has been found to be a good predictor of melody recognition (Peynircioğlu, Tekcan, Wagner, Baxter, & Shaffer, 1998). In Experiment 2, recognition was assessed by asking participants to sing the tune that they thought had been presented. Sung responses allowed a more reliable verification of recognition than did the usual naming or title verification tasks. Indeed, a tune can be well recognized even though its title cannot be retrieved from memory.

The following predictions were made. Because the feeling of knowing can emerge in the absence of recognition (e.g., in the *tip-of-tongue* effect; see Brown, 1991), it was expected that fewer notes would be necessary to judge the familiarity of a musical excerpt than to sing its continuation. Moreover, by analogy with the language domain (i.e., the word frequency effect), it was predicted that highly familiar tunes would lead to a feeling of knowing and would be identified earlier than moderately familiar tunes. Familiar tunes would, in turn, require fewer notes to be correctly judged than unfamiliar tunes, since to provide a *nonfamiliarity* judgment, an exhaustive search in memory is more likely to be performed. In addition, since musicians typically perform better than nonmusicians on musical tasks, we predicted that the former would be more confident in their responses, thus showing earlier recognition than the latter. Finally, if motives contribute to recognition, it was expected that the emergence of a feeling of knowing or the isolation of a melody would occur at motives' boundaries.

EXPERIMENT 1

Method

Participants. Twenty-three students from the University of Montreal volunteered to participate in the experiment. The participants were assigned to one of two groups defined by musical training. Eleven (6 males and 5 females) were referred to as *musicians*. They had at least 4 years of formal musical training ($M = 7.7$ years). Their ages ranged from 20 to 27 years ($M = 22.3$ years). The other 12 (6 males and 6 females) were classified as *nonmusicians*, since they had no formal education in music. Their ages ranged from 19 to 27 years ($M = 20.8$ years). All the participants were francophone and had lived in Quebec since the age of 2 years or earlier. These criteria served to promote a homogeneous musical culture among the participants (Peretz, Babaï, Lussier, Hébert, & Gagnon, 1995). All the participants were remunerated.

Material. Eighty songs (40 familiar and 40 unfamiliar) were selected to be used in the experiment. For each song, solely the melody, without lyrics and accompaniment, was utilized as a stimulus. All the melodies were taken from a repertoire of French traditional songs (Berthier, 1979) for which norms of familiarity had previously been established (see Peretz et al., 1995). Seven amateur musicians were asked to segment each of the 88 tunes considered in our norms into their natural parts by placing slashes on scores. The first and second motive boundaries were determined by compiling their segmentation data. Only those 40 melodies whose boundaries were marked by the majority were considered appropriate for the study. Half (20) were the most familiar melodies (mean familiarity rating = 4.9, range = 4.7–5.0, with 1 meaning *unfamiliar* and 5 meaning *very familiar*; see Peretz et al., 1995), and the other half (20) were the least familiar melodies (mean familiarity rating = 4.1, range = 3.2–4.6). Throughout the article, these two subsets of familiar melodies will be referred to as *high-familiarity tunes* and *moderate-familiarity tunes*, respectively. The number of notes in the high-familiarity tunes ranged from 10 to 30 ($M = 16.1$) and, in the moderate-familiarity tunes, from 11 to 31 ($M = 20.9$). Duration of tune excerpts varied from 4.0 to 17.4 sec. The length of each melody was the same as that used in the normative study (Peretz et al., 1995) and provided ample opportunities for identification. Forty unfamiliar tunes, taken from the same repertoire of traditional songs (Berthier, 1979), were also used in the experiment. These were considered unfamiliar by more than 200 students tested in a previous study (Peretz, Gaudreau, & Bonnel, 1998). The number of notes in the unfamiliar tunes ranged from 11 to 33 ($M = 18.8$), and they lasted between 4.5 and 13.2 sec.

All the melodies were played with a piano timbre through a Yamaha TX-81Z synthesizer under the control of an IBM-compatible computer. The analog ouput was sampled onto a Macintosh PowerPC computer, using SoundEdit II software. The beginning and the end of each note in a melody were determined manually through visual inspection of the waveform. Boundaries were carefully determined in order to minimize spectral splatter at the offset of each note. When rests occurred in a tune, they were treated as part of the previous note. The information on notes' boundaries was used by the Psyscope software (Cohen, MacWhinney, Flatt, & Provost, 1993) to play each tune with gated presentation, as will be described below.

Procedure. Musicians and nonmusicians were tested individually, using the same set of 80 tunes, presented in one of two random orders. Each tune was played with gated presentation. An example of gated presentation is illustrated in Figure 1. For the first presentation, only the first note of the melody (i.e., the first gate) was played. On the following presentation, the first two notes of the melody (the first two gates) were played, and so forth. After each presentation, the participants were asked to say whether the tune was familiar or unfamiliar and to rate their level of confidence on a 7-point scale (with 1 = *very unsure* and 7 = *very sure*). The participants were asked to provide a response even though, at first, they were unsure. Presentation of each melody lasted until the participant gave three successive consistent judgments with a confidence rating of 7. The responses were provided verbally and were noted by the experimenter.

The participants were informed that they would be hearing a set of familiar and unfamiliar songs, and the nature of the presentation format was explained. Moreover, a practice trial was given in order

to familiarize the participant with the task. The participants were tested in a quiet room, and the stimuli were presented through Labtec (Spatializer) loudspeakers in free field, at a comfortable intensity level. The test was divided into two sessions of about 1.5 h each. The experiment was run on PsyScope, using a Macintosh computer PowerPC.

Results

One familiar melody and one unfamiliar melody were judged to be unfamiliar and familiar, respectively, by more than 50% of the participants. These were discarded from the analyses. Similarly, when a participant did not provide three successive maximum ratings before the end of the melody, his or her responses were not considered in the analyses. The overall rate of rejection was 14.7% of the responses (corresponding to 270 out of 1,840 judgments).

For each of the 78 remaining tunes, two measures were taken to express the number of notes needed to develop a feeling of knowing: The familiarity emergence point[1] (FEP), which was the point at which the participant correctly began to consider that a melody was familiar, corresponded to the note (or gate) number at which the correct response *familiar* was given for the first time and was never changed thereafter. The familiarity point (FP) was the point at which the participant was totally confident that the melody was familiar or unfamiliar. FP corresponded to the note (or gate) number at which the participant provided a correct answer and at which he or she reached maximum confidence for the first of three consecutive presentations. Note that FP is used with both familiar and unfamiliar melodies, even though, in the latter case, the term *unfamiliarity point* would be more adequate. Nevertheless, for sake of simplicity, only FP will be used throughout the article. In addition, it is worth mentioning that although FEP applies well to familiar melodies, this measure seems unsound with unfamiliar melodies. Typically, a melody is considered unfamiliar when its first notes are presented, so it is impossible to determine when a *feeling of not knowing* does emerge. Thereby, only FEPs obtained with familiar melodies were submitted to the following analyses.

In order to assess whether FEP and FP varied as a function of the melodies' familiarity and whether these measures were affected by musical training, separate 2 (group) × 2 (familiarity) repeated measures analyses of variance (ANOVAs) were run by taking either subject (F_1) or item (F_2) as the random variable. When subject was the random variable, group (musicians or nonmusicians) was the between-subjects factor, and familiarity (familiar or unfamiliar) was the within-subjects factor. When item was the random variable, group (musicians or nonmusicians) was considered as the within-items factor, and familiarity (familiar or unfamiliar) as the between-items factor. Further analyses were performed in order to compare the results obtained with high-familiarity and moderate-familiarity tunes. In this case, the degree of familiarity (high familiarity and moderate familiarity) was used, instead of familiarity, as a factor.

Unfamiliar versus familiar melodies. Throughout the study, only the number of notes (e.g., at FP) were submitted to the analyses. The mean stimulus duration will also be reported, but for descriptive purposes. The mean FPs obtained with unfamiliar and familiar melodies for the musicians and the nonmusicians are presented in Table 1. It was found that FP occurred earlier with familiar than with unfamiliar melodies. This difference was affected by musical training, as was revealed by a significant group × familiarity interaction [$F_1(1,21) = 5.66, p < .05; F_2(1,76) = 14.15, p < .001$]. Post hoc pairwise comparisons (by

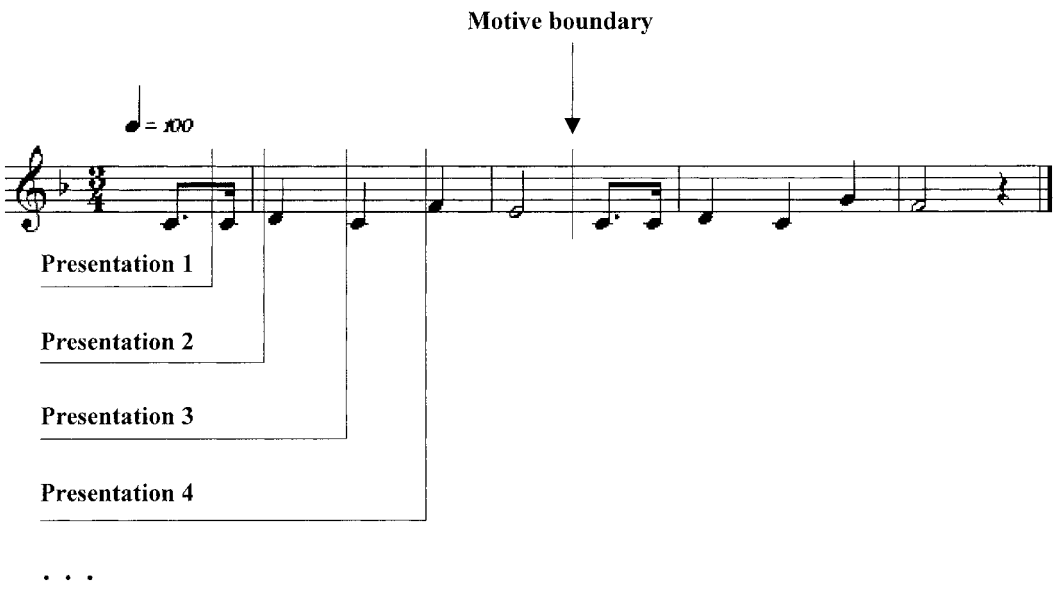

Figure 1. Example of gated presentation as applied to the melody of "Happy Birthday," showing Presentations 1–4. The motive boundary of the melody is also displayed.

Table 1
Mean Familiarity Points in Experiment 1 With Familiar and Unfamiliar Melodies for Musicians and Nonmusicians, With the Corresponding Mean Duration

	Notes		
Group	No.	*SE*	Duration (sec)
Musicians			
Familiar tunes	6.3	0.2	3.3
Unfamiliar tunes	8.5	0.4	3.4
Nonmusicians			
Familiar tunes	6.9	0.3	3.6
Unfamiliar tunes	10.2	0.5	4.1

Tukey HSD, corrected for unequal samples) revealed that musicians attained FP earlier than did nonmusicians. However, this effect was larger with unfamiliar melodies ($p < .001$, by subject and by item) than with familiar melodies ($p < .001$, by item, but nonsignificant in the analysis by subject). Nevertheless, for both groups, FP occurred later with unfamiliar melodies than with familiar melodies (with $p < .001$ in the analysis by subject and in the analysis by item).

High-familiarity versus moderate-familiarity melodies. The mean FEPs and FPs obtained with high-familiarity and moderate-familiarity tunes for the musicians and the nonmusicians are shown in Table 2. Both FEP and FP were found to occur earlier with high-familiarity tunes than with moderate-familiarity tunes, as was attested by a significant main effect of degree of familiarity [at FEP, $F_1(1,21) = 71.17$, $p < .001$, and $F_2(1,37) = 11.85$, $p < .01$; at FP, $F_1(1,21) = 117.38$, $p < .001$, and $F_2(1,37) = 21.23$, $p < .001$]. Moreover, the musicians attained FEP and FP earlier than did the nonmusicians; yet this main effect of group was significant solely in the analysis by item [at FEP, $F_2(1,37) = 21.17$, $p < .001$; at FP, $F_2(1,37) = 18.36$, $p < .001$]. Finally, the group × degree of familiarity interaction did not reach significance. In order to examine whether the musicians were more confident than the nonmusicians about their responses at the FEP, a similar 2 (group) × 2 (degree of familiarity) repeated measures ANOVA was performed in which the mean confidence rating provided at the FEP was considered the dependent variable (for the mean confidence values, see Table 2). The musicians appeared to be more confident than the nonmusicians; however, the group effect was again limited to the analysis by item [$F_1 < 1$; $F_2(1,37) = 8.38$, $p < .01$]. Note that confidence was higher for moderate-familiarity tunes than for high-familiarity tunes, but this main effect of degree of familiarity reached significance only with the analysis by subject [$F_1(1,21) = 7.81$, $p < .05$]. The group × degree of familiarity interaction did not reach significance.

One of the purposes of the present study was to assess whether the emergence of a feeling of knowing (i.e., at FEP) coincides with the boundaries of motives. To do so, the mode of the FEPs and of the FPs (i.e., the gate number at the FEP or the FP occurring most frequently) was taken for each melody. Then the percentage of melodies in which the FEP or the FP coincided with the boundary of the first or the subsequent motive was computed. These percentages were compared with a chance level. If the FEP or the FP occurred on the motives' boundaries, it would be expected that the FEP or the FP would fall significantly more often on this position than on other notes of the melody preceding or following the FEP or the RP. For our purposes, we considered the four notes around the mode (i.e., two preceding and two following the mode), so that the chance probability was 20%. This comparison revealed that the percentage of melodies at which FEP corresponded to a motive boundary was significantly above chance for the nonmusicians (42.6%, $p < .05$, by a binomial test; 17.9% for the musicians). Conversely, the FPs significantly coincided with motive boundaries for the musicians (35.9%, $p < .05$, by a binomial test; 28.9% for the nonmusicians).

Discussion

The results show that 3 to 6 notes (i.e., about 2 sec) of a familiar melody were sufficient to evoke a feeling-of-knowing judgment (at the FEP). However, to gain full confidence about this first estimate, approximately 2 additional notes were needed (FP). As was predicted, decisions were made later with unfamiliar melodies than with familiar melodies: 8 to 10 notes were needed to decide that melodies are unfamiliar. Moreover, a clear effect of familiarity was shown when high-familiarity tunes were

Table 2
Mean Familiarity Emergence Points (FEPs) and Mean Familiarity Points (FPs) in Experiment 1 With High-Familiarity (HF) and Moderate-Familiarity (MF) Tunes for Musicians and Nonmusicians, With Mean Durations and Mean Confidence Ratings

	FEP				FP		
	Notes		Duration		Notes		Duration
Group	No.	*SE*	(sec)	Confidence	No.	*SE*	(sec)
Musicians							
HF tunes	3.2	0.3	1.8	3.4	5.4	0.2	2.9
MF tunes	4.8	0.4	2.6	4.0	7.3	0.3	3.8
Nonmusicians							
HF tunes	4.0	0.4	2.3	3.2	5.9	0.2	3.1
MF tunes	5.7	0.5	3.0	3.4	8.2	0.4	4.2

compared with moderate-familiarity tunes. Fewer notes were needed to judge high-familiarity tunes as familiar, as compared with moderate-familiarity tunes. This finding is unlikely to have stemmed from the participants' being more confident with high-familiarity than with moderate-familiarity tunes. Indeed, the opposite pattern was found (i.e., the participants were more confident when judging moderate-familiarity melodies than with high-familiarity melodies). Overall, these results reveal a robust effect of familiarity on the music recognition process.

As was expected, musical training was found to affect the familiarity judgments, since the musicians reached the FEP and the FP earlier than did the nonmusicians. However, the musicians' advantage was less pronounced for familiar than for unfamiliar melodies. This finding is consistent with prior evidence showing that the use of familiar music tends to reduce musicians' advantage, as compared with nonmusicians (Smith et al., 1994). In addition, the musicians were more confident in their judgments than were the nonmusicians at the FEP. This result may explain why the two groups obtained different FPs, insofar as the musicians' higher level of confidence at the FEP might have led them to reach maximum confidence within fewer notes (i.e., at the FP). Further support for this interpretation can be found in the observation that motive structure had a larger and earlier impact on the judgments of the nonmusicians than on those of the musicians. In conclusion, these findings indicate that musical training affects the efficiency of the recognition process.

EXPERIMENT 2

In the second experiment, we sought to explore the time course of melody recognition by asking the participants to sing melodies, instead of providing familiarity judgments. For this purpose, the participants were presented with gated melodies and had to sing the tune that they thought had been played.

Method

Participants. Twenty-four participants who did not participate in Experiment 1 were recruited, mostly from the university community, to participate in the experiment. Half (6 males and 6 females) formed the musicians' group. They had at least 4 years of formal musical training (M = 9.4 years). Ten had degrees in music, and all had specific training or experience in singing. Their ages ranged from 21 to 45 years (M = 27.0 years). The other half (3 males and 9 females) formed the nonmusicians' group, since they had no formal musical training. Despite their lack of musical education, however, the nonmusicians were willing to sing. Their ages ranged from 20 to 42 years (M = 28.3 years). As in Experiment 1, all the participants were francophone and had lived in Quebec since the age of 2 years or earlier. All the participants were remunerated.

Materials and Procedure.The materials and the presentation format were the same as those in Experiment 1, except that only the 40 familiar songs (20 high-familiarit y tunes and 20 moderate-familiarity tunes) were considered here. The participants were instructed to listen to each presentation of a melody, to sing the melody that they thought had been presented, and to indicate their level of confidence on a 7-point scale (with 1 = *very unsure* and 7 = *very sure*). They were encouraged to sing whatever came to mind and, in addition, to provide a title when one was available. The trial lasted until the participant was able to sing the correct melody and had provided three consecutive confidence ratings of 7. A melody was considered correctly sung when the participant produced at least three recognizable notes beyond the end of the gated stimulus, as judged by the experimenter, a professional singer. At the end of each trial, feedback (including the title) was given. The sung responses were tape-recorded, and the confidence ratings were noted by the experimenter.

The participants were informed that they would be hearing a set of familiar songs, and the nature of the presentation format was explained. Moreover, there was a practice trial. The experiment lasted approximately 2 h. The experiment was run on the Experimenter software (Altmann, Wathanasin, Birkett, & Russell, 1992), using a Macintosh computer IIfx.

Results

The data were scored according to the same criteria as those used in Experiment 1. Two moderate-familiarity tunes were dropped from the data set, since they were not recognized by more than 50% of the participants. In all, 129 judgments out of 960 (13.4% of the entire set) were discarded.[2] The number of notes needed to recognize a melody was expressed by the IP and the RP. The IP was the point at which the participant demonstrated a correct insight into the identity of the melody, by singing it for the first time. The IP corresponded to the note (or gate) number at which the participant first successfully sang three notes of the target melody beyond the presented melodic segment,

Table 3
Mean Isolation Points (IPs) and Mean Recognition Points (RPs) in Experiment 2 With High-Familiarity (HF) and Moderate-Familiarity (MF) Tunes for Musicians and Nonmusicians, With Mean Durations and Mean Confidence Ratings

	IP				RP		
	Notes		Duration		Notes		Duration
Group	No.	*SE*	(sec)	Confidence	No.	*SE*	(sec)
Musicians							
HF tunes	5.2	0.2	2.5	5.8	6.0	0.2	3.0
MF tunes	6.9	0.3	3.6	5.8	8.0	0.2	4.1
Nonmusicians							
HF tunes	5.0	0.1	2.5	4.5	6.5	0.3	3.2
MF tunes	6.6	0.3	3.7	4.6	8.4	0.4	4.4

without any change in response thereafter. The RP was the point at which the participant was certain of having sung the corresponding melody. The RP corresponded to the note (or gate) number at which the participant correctly sang the melody and at which his or her confidence rating reached 7 for the first of three consecutive presentations.

The mean IPs and RPs obtained with high-familiarity tunes and moderate-familiarity tunes for the musicians and the nonmusicians are shown in Table 3. The average IP ranged from about five to seven notes, and the RP ranged from six to eight notes. IPs and RPs were submitted to 2 (group) × 2 (degree of familiarity) repeated measures ANOVAs by considering either subject (F_1) or item (F_2) as the random variable. When subject was the random variable, group (musicians or nonmusicians) was the between-subjects factor, and degree of familiarity (high familiarity or moderate familiarity) was the within-subjects factor. When item was the random variable, group (musicians or nonmusicians) was considered the within-items factor, and degree of familiarity (high familiarity or moderate familiarity) was considered the between-items factor. The analyses revealed that both the IP and the RP occurred later for moderate-familiarity tunes than for high-familiarity tunes, as was attested by significant main effects of degree of familiarity [at IP, $F_1(1,22) = 110.69$, $p < .001$, and $F_2(1,36) = 9.60$, $p < .01$; at RP, $F_1(1,22) = 147.20$, $p < .001$, and $F_2(1,36) = 11.78$, $p < .01$]. Moreover, the musicians isolated melodies earlier than did the nonmusicians, but this effect was reversed at the RP. These effects of musical expertise were statistically significant by item analysis only [at IP, $F_1 < 1$, and $F_2(1,36) = 6.84$, $p < .05$; at RP, $F_1(1,22) = 1.64$, n.s., and $F_2(1,36) = 4.62$, $p < .05$]. No group × degree of familiarity interactions reached significance. In order to examine whether the musicians were more confident than the nonmusicians about their responses at the IP, a similar 2 (group) × 2 (degree of familiarity) repeated measures ANOVA was performed considering the mean confidence rating provided at the IP as a dependent variable (for the mean confidence values, see Table 3). The musicians were more confident than the nonmusicians, as was shown by a significant main effect of group [$F_1(1,22) = 7.40$, $p < .05$; $F_2(1,36) = 105.86$, $p < .001$]. No main effect of degree of familiarity and no group × degree of familiarity interaction reached significance.

To examine the contribution of motives to isolating melodies, the percentage of melodies whose IP coincided with the boundaries of the first or the subsequent motive was compared with chance level, computed following the same formula as that in Experiment 1. For both the musicians and the nonmusicians, the IP coincided with the motives' boundaries above chance (52.6% for the musicians, $p < .01$; 67.6% for the nonmusicians, $p < .01$, by a binomial test). The RP corresponded to the motives' boundaries above chance only for the musicians (47.4%, $p < .01$, by a binomial test; 15.8% for the nonmusicians).

In order to assess the effects of task demands on the melody recognition process, the results obtained in this experiment were compared with those obtained in Experiment 1. The results showed that fewer notes were necessary to *judge* the familiarity of a musical excerpt than to *sing* its continuation, for both the musicians and the nonmusicians [IP > FEP, with $t_{mus}(37) = 9.89$, $t_{nonmus}(37) = 5.55$, $ps < .001$; RP > FP, with $t_{mus}(37) = 4.34$, $t_{nonmus}(37) = 2.07$, $ps < .001$].

Discussion

Within five to seven notes (i.e., about 3 sec), listeners are able to isolate the correct melody. To reach full confidence in their guess, however, they need approximately one additional note. As in Experiment 1, listeners isolated and recognized high-familiarity tunes earlier than moderate-familiarity tunes. The effect of familiarity on recognition was consistent and robust.

The results were less consistent when the effect of musical training were considered. The musicians' RP occurred earlier than the nonmusicians' RP, hence replicating the results of Experiment 1. This difference, again, might have resulted from the musicians being more confident than the nonmusicians about their judgments. However, when IP was considered, the reverse situation occurred. The nonmusicians isolated the correct melody earlier than did the musicians. Although the IPs in the two groups were not far apart, the difference was statistically significant. This finding cannot be accounted for by differences in confidence, since the musicians were more confident in their production. This apparent paradox is best understood in the framework of the cohort model, to be discussed hereafter.

GENERAL DISCUSSION

Time Course of the Music Recognition Process

As is shown in Figure 2, summarizing the results of the two experiments, both musicians and nonmusicians seem to proceed in the same manner when recognizing a tune. At first, the familiar music evokes a general feeling of knowing (at FEP); then, after a few notes, a match with a memory representation is found, and a tune is selected (at IP). Only later do the listeners gain full confidence in their judgments (at FP followed by RP). These results confirm that recognition occurs later than the feeling of knowing. Much more information is needed to judge a tune as unfamiliar. In general, nonmusicians seem to require more notes than do musicians to achieve recognition. However, nonmusicians isolate a melody more rapidly than do musicians.

These findings can be easily interpreted using a cohort model (Marlsen-Wilson, 1987). By analogy to word recognition, the music recognition process can be divided into two stages: access and selection. In the access stage, an initial cohort of melodies stored in memory is activated on the basis of the first notes of the presented melody. During the selection stage, the members of the initial cohort that do not match the incoming musical information are dropped from the cohort. At the end of this stage, only one

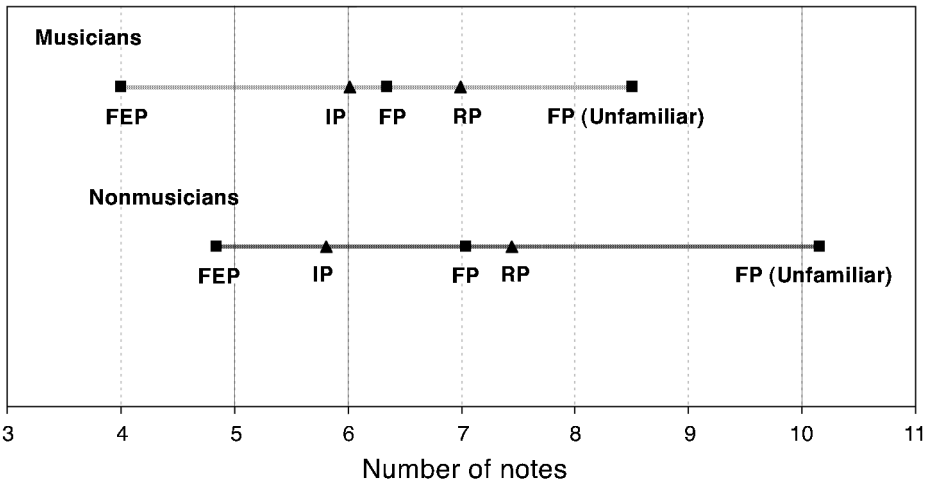

Figure 2. Summary of the results obtained in Experiments 1 (square symbol) and 2 (triangle symbol) for musicians and nonmusicians. The mean values of each point (FEP, familiarity emergence point; IP, isolation point; FP, familiarity point; and RP, recognition point) are ordered as a function of the number of notes. The first four points all refer to judgments provided with familiar melodies, whereas the last point, as indicated, has been obtained with unfamiliar melodies.

candidate will reach a threshold level, thus leading to isolation. The model accounts for the fact that more notes are needed to judge melodies as unfamiliar than to judge them as familiar. To judge a tune as unfamiliar, the listener has to acknowledge that the melodies stored in memory do not correspond to the presented melody. To this end, all the members of the initial cohort must be examined. Since none of the melodies reaches the threshold needed for isolation, listeners have to infer, by default, that the melody is unfamiliar. It is likely that both the search for a candidate within the cohort and the need for an additional inference process concur to delay the judgment for unfamiliar music. Furthermore, the model accounts for the difference between high-familiarity and moderate-familiarity tunes by referring to the concept of level of activation (as in Marslen-Wilson, 1987). In the model of spoken word recognition, it is assumed that the initial level of activation varies as a function of frequency (i.e., the more frequent a word is, the higher its activation). Analogously, high-familiarity music may present a higher initial level of activation than does moderate-familiarity music. It follows that less activation is needed to reach the threshold for isolation, which is assumed to be the same for all melodies, thus accounting for the earlier recognition of high-familiarity tunes.

Finally, the cohort model can account for the differences due to musical training. Musicians have probably memorized a larger number of melodies than have nonmusicians, following extensive musical exposure. Accordingly, musicians would access a larger initial cohort of melodies than would nonmusicians during the access stage of the recognition process. Since the feeling of knowing is based partly on the total amount of accessed information (see Koriat, 1993; Koriat & Levy-Sabot, 2001), a larger initial cohort size should entail a higher feeling of knowing, thus leading musicians to reach FEP earlier than do nonmusicians. Conversely, the larger size of the initial cohort may hinder the isolation of the correct melody during the selection stage, insofar as a larger number of nonmatching candidates has to be examined. This may, in turn, account for the more precocious IP in nonmusicians. The other differences occurring between musicians and nonmusicians are best explained by differences in confidence. Both the FP and the RP require maximal confidence. Since musicians are generally more confident than nonmusicians when providing a judgment about music, it is not surprising that they will attain FP and RP earlier than will nonmusicians.

Variables Mediating the Recognition Process

It is likely that musical motives affect the time course of the recognition process. Indeed, musicians and nonmusicians are found to isolate the correct melody in accordance with motive boundaries. Moreover, motive boundaries seem to play a role in increasing the confidence level (i.e., at the RP), although only for musicians. Similarly, motive boundaries affect the familiarity judgments of nonmusicians at an earlier point (at the FEP) than for musicians (at the FP). This pattern of results suggests that the motive structure has an impact at different stages of the recognition process in musicians and nonmusicians. It seems that the motive unit facilitates memory access in nonmusicians and tune selection in musicians. This differential impact might reflect the difference in melody cohort size, alluded to earlier, that would distinguish the memory of musicians and nonmusicians.

Further analyses of the data were performed in order to uncover other bottom-up processes, such as the principles defined by Narmour (1990; for a review, see Schellenberg,

1996), that may affect the time course of the recognition process. However, these attempts proved unsuccessful. No particular property could be identified.

CONCLUSIONS

In the present study, we sought to explore the time course of the music recognition process by means of the gating paradigm. Altogether, the results reveal that approximately 3 sec (e.g., at the IP) are required for a melody to be recognized. Both familiarity and musical training are found to affect the recognition process. However, the divergence between musicians and nonmusicians does not seem to arise from qualitatively different recognition systems. Both musicians and nonmusicians seem to segment melodies into motives to facilitate recognition. These results can be easily accounted for by a cohort model. The differences in initial activation levels of melody representations in a cohort according to their prior frequency of experience can account for the observed effects of familiarity. The effects of musical expertise can be attributed to the larger melody cohort that musicians are likely to have in memory, as compared with nonmusicians, and to the higher degree of confidence in their musical skills. This difference may, in turn, explain why the motive structure of melodies has an earlier impact on the recognition process in nonmusicians than in musicians. The motive would serve as an index to the melody cohort for nonmusicians and as a selection criterion among alternatives in the melody cohort for musicians. Hence, the cohort model represents a good point of departure for modeling the time course of the music recognition process. However, the model needs further specification of its inherent properties to provide a detailed account for the time course of music recognition. The cohort model was developed originally to account for spoken word recognition and, more recently, has been applied to visual word recognition (see Johnson & Pugh, 1994). Insofar as music and language present remarkable differences, it is obvious that only the general properties of the original model (e.g., the multiple access to the candidates of the cohort) can be transferred to the music domain. For instance, we found that the segmentation of melody into motives orient and affect the recognition process. Which other bottom-up processes govern the recognition process? How do these processes interact with top-down information, derived from the activation of memory representations? Moreover, at what stage of the recognition process does this interaction occur? These and similar questions cannot be answered with the current general conceptualization of the cohort model and must await further investigation.

In conclusion, the gating paradigm has proven to be an adequate tool for studying the music recognition process. This paradigm should compensate for the lack of a methodology well suited to capture the evolution of recognition over time, thus stimulating more research into the nature of the information that governs the time course of music recognition.

REFERENCES

Altmann, G., Wathanasin, S., Birkett, T., & Russell, P. (1992). *Experimenter program.* Brighton, U.K.: University of Sussex, Network Analysis Ltd.

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology*, **84**, 147-166.

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 501-515.

Berthier, J. E. (1979). *1000 chants* [1,000 songs]. Paris: Presses de l'Ile-de-France.

Bigand, E., & Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychological Research*, **62**, 237-254.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, **109**, 204-223.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, **25**, 257-271.

Cotton, S., & Grosjean, F. (1984). The gating paradigm: A comparison of successive and individual presentation formats. *Perception & Psychophysics*, **35**, 41-48.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff`s grouping preference rules. *Music Perception*, **4**, 325-360.

Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics*, **11**, 411-412.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, **85**, 341-354.

Dowling, W. J. (1984). Musical experience and tonal scales in the recognition of octave-scrambled melodies. *Psychomusicology*, **4**, 13-32.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, **49**, 524-531.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition.* Orlando, FL: Academic Press.

Dowling, W. J., & Hollombe, A. W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception & Psychophysics*, **21**, 60-64.

Frauenfelder, U. H. (1996). Computational models of spoken word recognition. In T. Dijkstra & K. de Smedt (Eds.), *Computational psycholinguistics: AI and connectionist models of human language processing*. London: Taylor & Francis.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, **28**, 267-283.

Grosjean, F. (1996). Gating. *Language & Cognitive Processes*, **11**, 597-604.

Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1995). Aging and experience in the recognition of musical transpositions. *Psychology & Aging*, **10**, 325-342.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Hébert, S., & Peretz, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory & Cognition*, **25**, 518-533.

Idson, W. L., & Massaro, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics*, **24**, 551-565.

Johnson, N. F., & Pugh, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology*, **26**, 240-346.

Kallman, H. J., & Massaro, D. W. (1979). Tone chroma is functional in melody recognition. *Perception & Psychophysics*, **26**, 32-36.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, **100**, 609-639.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 34-53.

KRUMHANSL, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.

KRUMHANSL, C. L., & SHEPARD, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 579-594.

LERDAHL, F., & JACKENDOFF, R. S. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

MARSLEN-WILSON, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, **25**, 71-102.

MARSLEN-WILSON, W. D., & WELSH, A. (1978). Processing interactions during word-recognition in continuous speech. *Cognitive Psychology*, **10**, 29-63.

NARMOUR, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.

PALMER, C., & KRUMHANSL, C. L. (1987a). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology : Human Perception & Performance*, **13**, 116-126.

PALMER, C., & KRUMHANSL, C. L. (1987b). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics*, **41**, 505-518.

PERETZ, I. (1989). Clustering in music: An appraisal of task factors. *International Journal of Psychology*, **24**, 157-178.

PERETZ, I., BABAÏ, M., LUSSIER, I., HÉBERT, S., & GAGNON, L. (1995). Corpus d'extraits musicaux: Indices relatifs à la familiarité, à l'âge d'acquisition et aux évocations verbales. *Canadian Journal of Experimental Psychology*, **49**, 211-239.

PERETZ, I., GAGNON, L., & BOUCHARD, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, **68**, 111-141.

PERETZ, I., GAUDREAU, D., & BONNEL, A.-M. (1998). Exposure effects on music preference and recognition. *Memory & Cognition*, **26**, 884-902.

PEYNIRCIOĞLU, Z. F., TEKCAN, A. I., WAGNER, J. L., BAXTER, T. L., & SHAFFER, S. D. (1998). Name or hum that tune: Feeling of knowing for music. *Memory & Cognition*, **26**, 1131-1137.

PROTOPAPAS, A. (1999). Connectionist modeling of speech perception. *Psychological Bulletin*, **125**, 410-436.

SCHELLENBERG, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition*, **58**, 75-125.

SCHELLENBERG, E. G., IVERSON, P., & MCKINNON, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review*, **6**, 641-646.

SMITH, J. D. (1997). The place of musical novices in music science. *Music Perception*, **14**, 227-262.

SMITH, J. D., NELSON, D. G., GROHSKOPF, L. A., & APPLETON, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition*, **52**, 23-54.

TYLER, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics*, **36**, 417-427.

TYLER, L. K., & WESSELS, J. (1985). Is gating an on-line task? Evidence from naming latency data. *Perception & Psychophysics*, **38**, 217-222.

WALLEY, A. C., MICHELA, V. L., & WOOD, D. R. (1995). The gating paradigm: Effects of presentation format on spoken word recognition by children and adults. *Perception & Psychophysics*, **57**, 343-351.

WHITE, B. W. (1960). Recognition of distorted melodies. *American Journal of Psychology*, **73**, 100-107.

NOTES

1. New terms are introduced that better reflect the isolation and the recognition points in the development of a feeling of knowing in a familiarity decision task.

2. In a few cases, although the participants correctly produced three notes beyond the presented melody, it became clear later that they did not recognize the melody but were, rather, improvising on the basis of their musical knowledge. In similar cases, confidence ratings usually remained very low. Recognition was verified by questioning the participant at the end of the trial, in order to decide whether to include the response in the analysis.

(Manuscript received July 25, 2001;
revision accepted for publication February 9, 2003.)