The Contribution of Local Features to Familiarity Judgments in Music

Emmanuel Bigand, Yannick Gérard, and Paul Molin

LEAD-CNRS, Université de Bourgogne, Institut Universitaire de France, Dijon, France

The contributions of local and global features to object identification depend upon the context. For example, while local features play an essential role in identification of words and objects, the global features are more influential in face recognition. In order to evaluate the respective strengths of local and global features for face recognition, researchers usually ask participants to recognize human faces (famous or learned) in normal and scrambled pictures. In this paper, we address a similar issue in music. We present the results of an experiment in which musically untrained participants were asked to differentiate famous from unknown musical excerpts that were presented in normal or scrambled ways. Manipulating the size of the temporal window on which the scrambling procedure was applied allowed us to evaluate the minimal length of time necessary for participants to make a familiarity judgment. Quite surprisingly, the minimum duration for differentiation of famous from unknown pieces is extremely short. This finding highlights the contribution of very local features to music memory.

Key words: **music; memory; local features**

Several models have been proposed to account for object recognition in cognitive psychology. The primary distinction is that between bottom-up and top-down models. Bottom-up models emphasize the influence of sensory-driven processes, which use all the features of a stimulus to secure recognition, matching them to known templates stored in memory. The top-down model emphasizes the role of context and the subject's expectations. A given stimulus will be more easily recognized in a context in which it is expected than in a context in which it has never or rarely occurred. In some cases, in different contexts, the same stimulus may be recognized as different objects. For example, a round orange object would be likely to be identified as a fruit when seen on a kitchen table, but as a tennis ball when seen on a tennis court. However, if the object is placed close to a bottle of water on a chair at the side of the tennis court, when local context would lead to its interpretation as a fruit, global context would lead to the interpretation of a ball. Bottom-up processes are assumed to contribute more to object identification when external conditions are good for perception, whereas top-down processes are more influential when stimuli are degraded or the experimental setting is detrimental to perception.

The present study deals with the influence of bottom-up processes on familiarity judgments in music (see Bigand and Tillmann[1] for a review of the influence of context on music perception). Bottom-up processes can be driven by different aspects of a stimulus, and corresponding cognitive models differ on the importance accorded to them. Template theories assume that the whole form of the stimulus drives recognition. The overall form is matched to a miniature copy of the stimulus (a template) in long-term memory. An object is recognized on the basis of the template providing the closest match to the input stimulus. The main problem with such template models is that objects at different times will often occur in different orientations, spatial positions, and even shapes, as is

Address for correspondence: Emmanuel Bigand, LEAD-CNRS UMR 5022, Université de Bourgogne, 2, Esplanade Erasme, Pôle AAFE, Université de Bourgogne. F-21000 Dijon, France. bigand@u-bourgogne.fr

The Neurosciences and Music III—Disorders and Plasticity: Ann. N.Y. Acad. Sci. 1169: 234–244 (2009).
doi: 10.1111/j.1749-6632.2009.04552.x © 2009 New York Academy of Sciences.

the case, for example, with alphanumeric stimuli. Given the enormous variations in visual and auditory stimuli, we would need to have stored in memory a considerable number of templates to recognize even a single object. One way to solve this problem is to suppose the integration of a normalization process (for orientation or size) in memory alongside the specifications for an object's shape. When an object is encountered with new metric specifications, the recognition process engages in transformations of the current stimuli in order to find a match to a stored representation. When the stimuli correspond to a new viewpoint of a known object, the transformation is a mental rotation. When a new size is encountered, the transformation is a mental zooming. Current template models would thus predict costs in time or accuracy in recognizing objects when metric specifications and angle of view have been changed. Larger changes would result in more extensive transformations. In addition, template models would anticipate a failure of recognition when an object is presented only in part or when object components are reorganized.

An alternative approach assumes that object recognition rests on the processing of local features. Any stimulus may be decomposed into local and independent attributes, such as the nose, eyes, ear, and mouth of a face, or the different lines of a letter of the alphabet (i.e., two oblique straight lines, and a connecting crossbar, for the letter A). The process of pattern recognition is assumed to begin with extraction of features from the visual stimulus presented, these features then being compared to those of the objects stored in memory.[2] The letter "B" could be coded by the following features: one vertical line, two continuous curves, and a small horizontal line. As a consequence, a given stimulus is more easily recognized when it occurs alongside objects that do not share the same features. That is, the letter "B" should be recognized faster when it occurs among letters containing vertical and horizontal straight lines, such as W, T, and X, than among curvaceous letters[3] (0, G, or R). Such feature theory models were successfully applied to word recognition,[4] and to auditory word identification.[5] In the former case, each letter of a word was supposed to activate word "nodes" that contain the letter. The most activated word node corresponds to the recognized word. This model provides a nice account of confusion errors in lexical decision tasks, and explains why words sharing similar letters are easily confused. The feature theory would also predict that scrambled objects could be recognized, as long as the scrambling does not alter the components of the object. Scrambling objects encountered in everyday life indeed does not prevent object identification. Feature theory also explains why objects varying greatly in size, orientation, or minor details remain identifiable as instances of the same template.

In some animal species object recognition was assumed to rest entirely on local-feature processing. But, for human beings, a large body of evidence demonstrates that global features matter. By presenting a large letter (H) made of small letters (either small H or small S), Navon[6] found that recognition of the small letter was faster when it corresponded to the large letter, while recognition of the large letter was unaffected by the nature of the small letter. This finding suggests that the recognition process starts with global, and moves on to local features. The "object superiority effect"[7] also indicates that detecting a given feature is easier when this feature occurs in a coherent rather than less coherent form. Similarly, it was found that detection of specific letters benefits from a "word superiority effect." Recognition of local features remains possible for a scrambled object, but is more difficult than for a normal object. Of interest, similar findings were found in human and some animal species (such as pigeons), showing that the global relationship between local features is helpful to recognition, even in pigeons.

There are several ways to integrate the contributions of local and global features in recognition models.[8–10] According to Biederman's theory[9,10] of recognition by components, the visual system extracts elementary geometric shapes called *geons* and uses them to identify the presented object. Geons are simple volume shapes, such as cubes, spheres, cylinders, and wedges. As with phonemes for spoken language, we need only a limited number (36) of geons for the representation of objects. We can identify an enormous number of objects with this small set of geons, each object representing a specific combination of just a few of them. A cup may be described as an arc connected to the side of a cylinder, and a bucket can be coded by the same two geons, but now the arc is connected across the top of the cylinder. This example emphasizes that the connecting points between geons define the relevant global features that matter for visual recognition. Recognition remains possible as long as these connecting points are preserved. Since the size and orientation of the object do not alter the relationships between geons, recognition remains possible even when objects are seen from different viewpoints. When the connecting points between geons are artificially removed, object recognition becomes extremely difficult.[9] And combining geons in a scrambled way significantly reduces recognition, which remains nevertheless above the level of chance. Thus it is believed that both local features (geons) and global features (the relationship between geons) contribute to object identification in both human and animal species.[11,12]

Whereas local features are generally considered more important than global features in object, visual, and auditory word recognition, the reverse conclusion emerges from face recognition studies. For example, "Identikit" and "Photofit," used by police forces to aid face recognition by eyewitnesses, depend upon a features-based approach. The face of the suspect is reconstructed, feature by feature, by adding the appropriate nose, ears, hairs and eyes. But, though faces may be characterized by specific features, several authors have suggested that face processing is holistic. In the most extreme model, faces are supposed to be coded and recognized as whole templates without reference to their specific parts. In computer vision, many face recognition algorithms process the whole face without explicitly processing facial features. Several items of empirical evidence support such a holistic processing of facial information. When the top halves and bottom halves of different famous faces were closely aligned, participants encountered great difficulty in naming the top part. The difficulty decreased when the two halves were less well aligned.[13] A close alignment seems to create a new overall configuration that interferes with recognition of the two halves of well-known faces.

The effects of face superiority and inversion and well as the "Thatcher illusion" constitute further well-accepted evidence for the holistic encoding of faces. In the face superiority effect, observers better discriminate a facial feature if it is presented in the context of a face, than if presented alone, or in a scrambled face.[14] In the inversion effect, a face is found more difficult to recognize when presented upside-down.[15] The holistic information is no longer available, so that processing by analysis of parts seems necessary, causing a characteristic decrease in recognition speed and accuracy. In the *Thatcher illusion,*[16] the eyes and the mouth of a person (initially Margaret Thatcher) were rotated within the facial image. A seemingly grotesque, strange facial expression resulted. Although obvious when the picture was upright, the strangeness was not perceived when the face was turned upside-down. Inverting the eyes and the mouth within a facial image changes the configural information. Interestingly, these three effects were reported for faces and not for objects. That is, there is no equivalent of the Thatcher and inversion effects for nonfacial objects. Thus, while the object superiority effect was found for objects, faces, and words, it is more pronounced for faces.[17,18] This finding leads us to the assumption that face recognition may be special in comparison to other recognition tasks.

According to Tanaka and Farah,[19] the brain has separate modules for separate kinds of objects; faces and words fall at opposite ends of a shape-processing continuum. Faces are processed as wholes, and words are processed by parts. This view was supported both by an fMRI study, which found face-specific regions in the brain, and by electrophysiological studies showing a specific negative potential peaking at about 170 ms from stimulus onset (N170) at occipitotemporal sites associated only with face processing.[20] An alternative view would be that the influence of configural information is not specific to face recognition. This influence may result from the subject expertise with stimuli. Since faces matter more than every other object in our social environment, human beings become expert at recognizing faces. If humans had a similar depth of experiencing and identifying other objects (birds, cars, or musical instruments), they would equally recognize these objects by processing their configural features.

Currently therefore, experiments in face recognition evaluate the influence of local and configural information by comparing two types of stimulus transformation: blurring and scrambling. Blurring and scrambling procedures permit the reduction of the contributions of either local or global features separately. Different types of scrambling are used. Scrambling faces was shown to reduce recognition, which remained nevertheless possible. Blurring faces lead to a similar result, suggesting that face recognition involved two routes. Of interest, the same effect was found for both familiar and unfamiliar (recently learned) faces.[21] Finally, further evidence for a dual route to face recognition comes from neuropsychological studies. Patients with prosopagnosia encountered difficulties in recognizing familiar faces, even though they can recognize other familiar objects. Prosopagnosic patients also are not affected by the inversion effect. They actually seem to recognize inverted faces better than do normal subjects. At the same time, there are patients with visual agnosia who are able to recognize familiar faces, but not familiar objects. Moreover, patients with alexia seem not to encounter difficulty in recognizing familiar faces. This overall pattern of data suggests that object recognition involves both local and the configural encoding, the local encoding being no longer accessible in alexia and visual agnosia, and the configural no longer accessible in prosopagnosia.

The relative importance of global and local features has not so far received similar attention in our understanding of the recognition of music. Up to now, studies have focused mostly on the importance of musical parameters (time or pitch, words, melodies), and on the way these parameters are combined in the memory trace. Initial research by Dowling and Fujitani[22] has emphasized the importance of some global features (melodic contour) over more local ones (intervallic features). The respective influences of these features seem to depend upon the time taken for the achievement of the memory task. Melodic contour is more influential over a short time, and pitch intervals over times longer than 20 s. Several findings further demonstrate the importance of global features. For example, playing a melody backward makes it hard to recognize. Such a manipulation might be considered the analogue of the face inversion effect in the temporal dimension of musical stimuli. Empirical research done in the framework of the generative theory of tonal music[23] also demonstrates the importance of global form on melodic recognition. Lerdahl and Jackendoff[23] assume that every tonal musical piece rests on an underlying structure (identified in the time-span reduction component of their theory). This reduced structure is derived from the most important tones of the piece, representing its musical skeleton. Many local features are eliminated in the reduction, and are no longer represented in the skeleton. To some extent, therefore, this musical skeleton might be compared to the blurred faces used in face recognition studies. Playing a skeleton pattern of pitches of well-know tunes suffices for music recognition.[24,25]

The easiest route to music recognition involves, however, local features. Superficial changes in timbre, which do not alter the

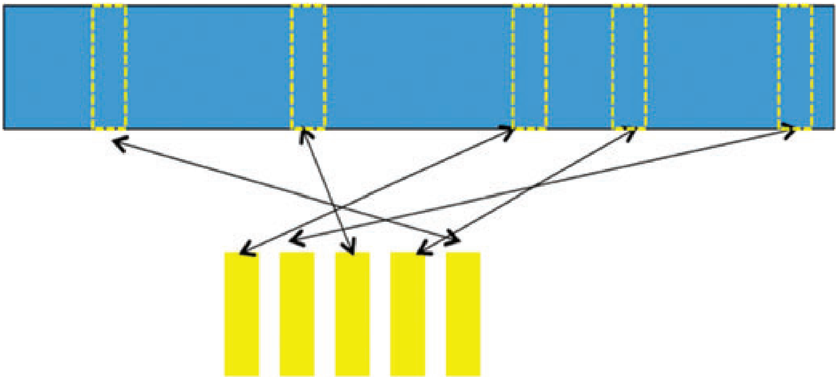

Figure 1. Scrambling music. Short excerpts are selected at random with no overlap, and then linked at random. Each excerpt starts and ends with a short fade-in and fade-out of 50 ms duration. Their duration was manipulated in order to evaluate the strength of local features on music recognition. (In color in *Annals* online.)

overall form of a piece, have dramatic effects on music recognition.[26] Changing the timbre of musical pieces decreases recognition. Other research demonstrates that familiar tunes might be judged as familiar as soon as the first three to six notes are played.[27,28] More challenging findings were reported by Schellenberg *et al.*,[29] who showed that pieces of popular music might be identified after only 100 ms. Recent findings using well-known classical instrumental music provide converging evidence that appropriate familiarity judgments do not need more than 500 ms of music.[30] All of these studies emphasize the importance of the local features identified at the beginning of a musical piece. The suggestion is, therefore, that musical recognition could rely on a cohort process similar to that of auditory word recognition,[5] starting at the very beginning of a piece.[27] It remains an open question whether any short excerpt of music taken at random could similarly activate the memory.

The present study further addresses the importance of local features with a new experimental method that was directly inspired by face recognition studies. Short temporal windows were taken in a random way in famous and unknown classical instrumental music (Fig. 1). These randomly selected musical excerpts were then randomly linked through short fades-in and -out, thus removing any acoustic clicks. This technique gave us a scrambled musical piece, without global configuration, but retaining all the local features. A similar scrambling method was used by Levitin and Menon[31] to assess brain responses to nonsyntactic (i.e., scrambled) music. Interestingly, a similar scrambling procedure[32,33] was used by the composer Roger Reynolds in his piece "The Angel of Death," seemingly to refresh the memory of listeners at the end of the piece. The scrambling algorithms used were believed to mimic the fast memory processes that apparently occur just before death. In previous studies, we had found that musical target identification is little affected by scrambling.[34] Accordingly, we were anticipating that listeners would still recognize familiar pieces, even though the pieces were scrambled. Our main purpose was to evaluate the minimum size that the temporal window should have to allow appropriate familiarity judgments. As such, our task taps into an implicit level of memory (i.e., familiarity judgments) by contrast to an explicit level of memory that would be addressed by a recognition task. In addition, a similar manipulation was done for French linguistic texts. This allows us to provide new information about the impact of scrambling in recognition of texts and music. This comparison would contribute to the larger issue of music and language comparison.[35]

TABLE 1. Familiar and Unfamiliar Musical Excerpts Used in the Study

Excerpt	Composer	Familiarity	Tempo	Length (s)
Badinerie	Bach	Fam	Allegro	16.1
William Tell Overture	Rossini	Fam	Allegro	14.6
Eine Kleine Nachtmusik	Mozart	Fam	Allegro	18.3
Hungarian Dance No. 5	Brahms	Fam	Allegro	20.5
New World Symphony, 4th movement	Dvorak	Fam	Allegro	20.9
Boléro	Ravel	Fam	Moderato	18.7
Jazz Suite No. 2	Shostakovitch	Fam	Moderato	22.9
Carmen, overture	Bizet	Fam	Moderato	21.7
“The Trout” Quintet	Schubert	Fam	Moderato	14.555
Emperor Waltz	Strauss	Fam	Moderato	15.864
Svenskt Festspiel	Söderman	Unfam	Allegro	21.3
Symphony Opus 11, 4th movement	Olson	Unfam	Allegro	15.3
Les caractèrtes de la danse No.11	J-F. Rebel	Unfam	Allegro	17.1
Symphony No 3, final movement	Berwald	Unfam	Allegro	14.3
Symphony VB 45 (presto)	Kraus	Unfam	Allegro	20.542
I Vadstena kloster, 3rd movement, procession	G. Bengtsson	Unfam	Moderato	18.9
Symphony No.1, final movement	Norman	Unfam	Moderato	20.9
Symphony No. 2	Scriabin	Unfam	Moderato	22.7
Quintet for piano and winds, 2nd movement	Beethoven	Unfam	Moderato	16.4
Midsommarvaka (rhapsody)	H. Alfvén	Unfam	Moderato	15.929
The Valkyrie	*Wagner*	*Train Fam*	*Allegro*	16.1
Wedding March	*Mendelssohn*	*Train Fam*	*Moderato*	11.1
Sonata “a Doi Chori”	*Schmelzer*	*Train Fam*	*Allegro*	13.4
Drapa	*Rubenson*	*Train Fam*	*Moderato*	26.1

Fam = familiar; Unfam = unfamiliar; Train Fam = training items.

Experiment

Method

Participants

Forty-nine musically untrained undergraduate students of the Université de Bourgogne participated in this experiment.

Stimuli[a]

Twelve very well-known pieces of classical instrumental music were used for the familiar excerpts (Table 1). Most of them had already been found to be known even to musically untrained French listeners by Filipic *et al.*[30] and were pretested by Plailly, Tillmann, and Royet.[36] For half of them the tempo was fast, and for the other half moderate. The 12 excerpts were roughly matched with a further 12 excerpts of classical instrumental music little known to the participants. The duration of these pieces varied from 11.1 to 26.1 s. These 24 stimuli were cut in a random way into nonoverlapping fragments of 250, 350, 550, and 850 ms. The fragments of each piece were then linked in a scrambled way, each fragment starting and ending with short fades-in and -out of 50 ms (Fig. 1). In order to avoid any artificial temporal regularities, the durations of the fragments were varied by ±50 ms. Comparative stimuli were provided by the original musical excerpts. In order to determine whether the process of making familiarity judgments may differ between music and language, a similar scrambling procedure was applied to 24 spoken texts of identical duration. Half of these texts were well-known (“Le corbeau et le renard” from Jean de La Fontaine) to French students, most of them having been learned by heart at school.

[a] Examples of experimental stimuli may be found on the web site http://www.u-bourgogne.fr/LEAD/people/bigand.html.

The other half were other texts, from the same authors ("Le chameau et les bâtons" from Jean de la Fontaine), but little known and not learned at school. These texts were spoken by one of us, whose voice was not familiar to the students.

Procedure

Participants were divided into two groups. Those in the "bottom-up group" (B-UP) started by listening to all the scrambled pieces that had the shortest fragments (250 ms ± 50 ms, block 1). They continued the experiment without pause, with the pieces made of longer fragments of 350 (block 2), 550 (block 3), and 850 ms (block 4). They ended the experiment with the original excerpts in the last block. Those students in the "top-down group" (T-DW) performed the tasks in the reverse order. They started with the original excerpts, and then were presented with scrambled music made of fragments of even shorter duration (850, 550, 350, and then 250 ms). After each stimulus, they were asked to decide, as soon as possible in a binary choice, whether or not the piece was known to them (familiar versus unfamiliar). They had been informed in advance that half of the pieces and texts were considered well-known when played in a normal way. In each of the two groups, half of the students started the study with musical stimuli, the other half starting with the spoken texts. At the end of the experiment, all participants listened for the second time to stimuli in a normal presentation, and judged their familiarity on a 10-point scale, thus allowing us to confirm all stimuli were genuinely known to all participants.

Results

Although the musical and linguistic stimuli had been selected from two groups of familiar and unfamiliar stimuli, participants reported sometimes being weakly familiar with some of the supposedly familiar musical pieces (Schubert's "Trout" Quintet, notably) or French texts. For that reason, the stimuli were sorted on the basis of the familiarity responses participants gave when presented with the normal excerpts. The responses given for the scrambled stimuli were then accordingly coded as correct or incorrect specifically for each participant. Thus, we could assess whether participants were able appropriately to judge the familiarity of scrambled music or text, and determine the minimal length of scrambling fragments which still allowed a correct judgment to be made. The percentages of responses given as "familiar" are reported in Figure 2, for both music and texts, and at each duration. An ANOVA, 2 (B-UP versus T-DW) × 2 (type of stimulus: music versus text) × 2 (type of excerpt: familiar versus unfamiliar) × 4 (duration) was performed with the first variable as a between-participants factor and the others as within-subject factors. The percentage of "familiar" responses defined the dependent variable. The type of excerpt was a significant factor: familiar excerpts received higher percentages of familiar judgments than did unfamiliar excerpts ($F(1,47) = 331.31$, MSE $= 0.07$ $P < 0.001$), and this effect increased with the duration of the fragments ($F(3,141) = 52.05$, MSE $= 0.02$ $P < 0.001$). Scrambled pieces were easily recognized as familiar when the duration of the fragment was long (850 ms). The first critical new finding was that participants could appropriately differentiate familiar and unfamiliar stimuli even when the fragments were as short as 250 ms ($P < 0.001$). Moreover, this effect was found both for music and for spoken text (with no significant differences between the two). Quite surprisingly, T-DW did not outperform B-UP, as attested by the absence of both main effect and any significant interaction with this factor.

A supplementary analysis was run on response time for correct familiar and unfamiliar responses. As shown in Figure 3, participants listened to the scrambled versions for several seconds (between 3 and 10) before giving a response. The same 2 × 2 × 2 × 4 ANOVA was performed, but this time with correct responses time as the dependent variable. The type of stimulus was significant,

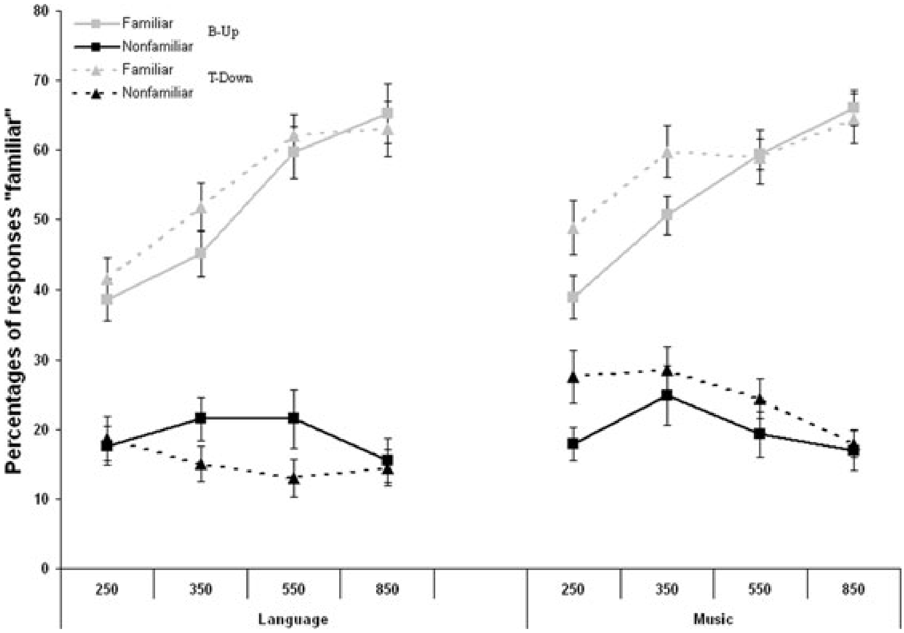

Figure 2. Percentages of responses labeled "familiar" by the participants for familiar and unfamiliar music and text excerpts presented in a scrambled way (250, 350, 550, and 850 ms). The bottom-up group (B-UP) started the experiment with the scrambling by the shortest fragments (250 ms), and the top-down group (T-Down) with the original excerpts.

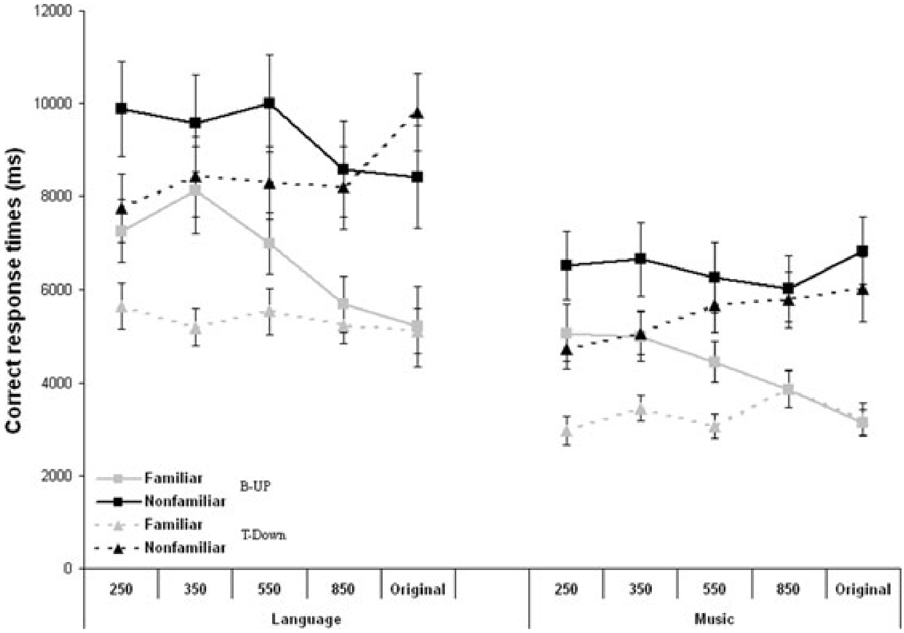

Figure 3. Correct response times for familiar and unfamiliar music and texts excerpts presented in a scrambled way (250, 350, 550, and 850 ms). The bottom-up group (B-UP) started the experiment with the scrambling by the shortest fragments (250 ms), and the top-down group (T-Down) group with the original excerpts.

correct response times being shorter for music than for language ($F(1,47) = 57.78$, MSE = 27,900,000 $P < 0.001$). Correct response times were also shorter for familiar than for unfamiliar stimuli ($F(1,47) = 132.38$, MSE = 15,200,000 $P < 0.001$). These times increased as the duration of the fragments decreased ($F(3,141) = 12.01$, MSE = 17,700,000 $P < 0.001$). An interesting new point was that participants continued to respond faster to familiar than unfamiliar stimuli even when the duration of the fragments was as short as 250 ms. Thus, despite the supposed confusion created by the scrambling method,[30] a memory trace was nevertheless activated which allowed participants to respond faster to well-known excerpts.

Discussion

The influence of scrambling on recognition has been investigated for both object and face recognition. Scrambling usually diminishes recognition, but does not necessarily prevent it. This finding emphasizes the importance of local features (in contrast to global features) to recognition. Usually, in the experiments reported here, the size of the visual units that are scrambled is not manipulated and corresponds to a locally complete feature (a geon, such as nose, ear, or eye). In the present experiment, we investigated whether well-known or familiar music may be recognized as familiar when presented in a scrambled way. Our finding indicates that musical memory, as well as memory for text, is not strongly affected by crude manipulations of the overall organization of musical and linguistic excerpts. This finding is consistent with other ones, but with less drastic changes, reported by Tillmann and Bigand.[33] Thus, even when rendered nonsyntactic, well-known music continues to be recognized as familiar without great difficulty.

The critical point of the study was our showing that familiarity judgments for both music and spoken texts can be made on the basis of excerpts as short as 250 ms. In language, this duration is approximately that of a phoneme. In music, this duration allows the perception of between one to three notes, depending upon the tempo of the piece. Thus, music recognition might indeed rest on extremely local features, as initially reported by Schellenberg *et al.*[29] This finding is impressive when considering that the musical stimuli used in the experiment derived from classical instrumental music, that is, from a musical repertoire that is not most familiar to which musically untrained listeners probably listen less often than to the pop tunes (in contrast to the pop tunes used by Schellenberg *et al.*[29]). Moreover, the finding is not specific to music, since a similar result was obtained for well-known spoken texts. A further surprising aspect of our data was that there was no difference between the bottom-up and top-down groups. In this second group, participants started by listening to the unscrambled original excerpts, both famous (familiar) and unknown (unfamiliar). This initial listening was expected to have primed their musical memory, resulting in some advantage in decoding the subsequent scrambled excerpts. No such effect was found, suggesting that recognition of well-known musical and linguistic stimuli is little affected by top-down factors, and more strongly by local features.

Finally, the finding of the sufficiency of a duration of only 250 ms was impressive, suggesting that participants might probably still discriminate well-known from unknown music, in even shorter durations, and even when scrambled. Repeating the experiment down to shorter durations indicated that participants actually continue to differentiate familiar from unfamiliar music, even when scrambled, in fragments as short as 100 ms and 50 ms.[37] In contrast, however, participants found it impossible to differentiate familiar from unfamiliar scrambled texts in fragments shorter than 250 ms. The findings thus shed new light on one specificity of musical memory. At such durations, there is no consistent musical unit that may be identified in scrambled music. Only the "colors" of the sound can be perceptible. By "color of sound" we mean the complex of local features that includes not only the timbre, but also the harmonic style, the voicing, and the orchestration. We suggest that the color of sound provides a very fast route for accessing musical memory traces. This color feature has no counterpart in the recognition of well-known texts, and is therefore one difference between music and language (as long as the voice pronouncing the text is not famously associated with the content of the text). The color of sound may, however, be a component of the quality of a speaker's voice, and it might be identified extremely rapidly. Voice recognition seems to need only a very short slice of auditory information: excerpts as short as 25 ms permit correct identification of the speaker.[38] To some extent, our finding parallels voice recognition finding, and suggests that the speed at which recognition of voice occurs is not specific to voice, but could also be found in music. Further research should develop this comparison, by evaluating how many random fragments of music we need to have to make appropriate familiarity judgments. On the basis of our current work,[37] this number and its duration might be shown to be very small.

Acknowledgment

This paper is supported by a grant from the Agence Nationale de la Recherche.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Bigand, E. & B. Tillmann. 2005. Effect of context on the perception of pitch structures. In *Pitch Perception.* C. Plack & A. Oxenham, Eds: 306–351. Springer Verlag. New York.
2. Selfridge, O. 1959. Pandemonium: a paradigm for learning. In *Symposium on The Mechanization of Thought Processes*. Her Majesty's Stationery Office. London.
3. Neisser, U. 1964. Visual search. *Sci. Am.* **210:** 94–102.
4. McClelland J.L. & D.E. Rumelhart. 1981. An interactive activation model of context effects in letter perception. Part 1. An account of basic findings. *Psychol. Rev.* **88:** 375–407.
5. Marslen-Wilson, W.D. & L.K. Tyler. 1980. The temporal structure of spoken language understanding. *Cognition* **8:** 1–71.
6. Navon, D. 1977. Forest before trees: the precedence of global features in visual perception. *Cogn. Psychol.* **9:** 353–383.
7. Pomerantz, J.R., L.C. Sager & R.G. Stoever. 1977. Perception of wholes and their component parts: some configural superiority effects. *J. Exp. Psychol. Hum. Percept. Perform.* **3:** 422–435.
8. Marr, D. & H.K. Nishihara. 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. B* **200:** 269–294.
9. Biederman, I. 1987. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* **94:** 115–147.
10. Biederman, I. 1990. Higher-level vision. In *An Invitation to Cognitive Science: Visual Cognition and Action.* D.N. Osherson, S. Kosslyn & J. Hollerbach, Eds.: 41–72. MIT Press. Cambridge, MA.
11. Kirkpatrick-Steger, K., E.A. Wasserman & I. Biederman. 1996. Effects of spatial rearrangement of object components on picture recognition in pigeons. *J. Exp. Anal. Behav.* **65:** 465–475.
12. Kirkpatrick-Steger, K., E.A. Wasserman & I. Biederman. 1998. Effects of geon deletion, scrambling, and movement on picture recognition in pigeons. *J. Exp. Psychol. Anim. Behav. Process.* **24:** 34–46.
13. Young, A.W., D.J. Hellawell & D.C. Hay. 1987. Configural information in face perception. *Perception* **16:** 747–759.
14. Tanaka, J.W. & M. Farah. 1993. Parts and whole in face recognition. *Quart. J. Exp. Psychol.* **46:** 225–245.
15. Farah, M.J., J.W. Tanaka & H.M. Drain. 1995. What causes the face inversion effect? *J. Exp. Psychol. Hum. Percept. Perform.* **21:** 628–634.
16. Thompson, P. 1980. Margaret Thatcher—a new illusion. *Perception* **9:** 483–484.
17. Yin, R.K. 1969. Looking at upside-down faces. *J. Exp. Psychol.* **81:** 556–566.
18. Pelli, D.G., B. Farell & D.C. Moore. 2003. The remarkable inefficiency of word recognition. *Nature* **423:** 752–756.
19. Tanaka, J.W. & M. Farah. 1991. Second-order relational properties and the inversion effect: testing a theory of face perception. *Percept. Psychophys.* **50:** 367–372.
20. Bentin, S. & L.Y. Deouell. 2000. Structural encoding and identification in face processing: ERP evidence

for separate mechanisms. *Cogn. Neuropsychol.* **17:** 35–54.

21. Schwaninger, A., C.C. Carbon & H. Leder. 2003. Expert face processing: specialization and constrainsts. In *Development of Face Processing*. G. Schwarzer & H. Leder, Eds.: 81–97. Hogrefe. Göttingen, Germany.
22. Dowling, W.J. & D.S. Fujitani,1971. Contour, interval and pitch recognition in memory for melodies. *J. Acoust. Soc. Am.* **49:** 524–531.
23. Lerdahl, F. & R. Jackendoff. 1983. *A Generative Theory of Tonal Music*. MIT Press. Cambridge, MA.
24. Bigand, E. 1990. Perception et compréhension des phrases musicales [perception and understanding of musical phrases]. Ph.D. thesis, University Paris X, ISSN 0294-1767, No. 09882/90.
25. Mélen, M. & I. Deliege. 1995. Extraction of cues or underlying harmonic structure: which guides recognition of familiar melodies? *Eur. J. Cogn. Psychol.* **7:** 81–106.
26. Poulin-Charronnat, B., E. Bigand, P. Lalitte, *et al.* 2004. Effect on instrumentation change on the recognition of musical materials. *Music Percept.* **22:** 239–263.
27. Dalla Bella, S., I. Peretz & N. Aronoff. 2003. Time course of melody recognition: a gating paradigm study. *Percept. Psychophys.* **65:** 1019–1028.
28. Schulkind, M. 2004. Serial processing in melody identification and the organization of musical semantic memory. *Percept. Psychophys.* **66:** 1351–1362.
29. Schellenberg, G., P. Iverson & M. McKinnon. 1999. Name that tune: identifying popular recordings from brief excerpts. *Psychon. Bull. Rev.* **6:** 641–646.
30. Filipic, S., B. Tillmann & E. Bigand. Judging familiarity and emotion from very brief musical excerpts: investigating the time-course of music processing with the gating paradigm. Submitted for publication.
31. Levitin D. & V. Menon. 2005. The neural locus of temporal structure and expectancies in music: evidence from functional neuroimaging at 3 Tesla. *Music Percept.* **22:** 563–575.
32. McAdams, S. & R. Reynolds. 2002. Problem-solving strategies in the composition of "The Angel of Death." 7th International Conference on Music Perception and Cognition. Adelaide, Australia. Causal Productions. July 17–21.
33. Reynolds, R. 2004. Compositional strategies in The Angel of Death for piano, chamber orchestra, and computer-processed sound. *Music Percept.* **22:** 173–205.
34. Tillmann, B. & E. Bigand. 1998. Influence of global structure on musical target detection and recognition. *Int. J. Psychol.* **33:** 107–122.
35. Patel, A.D. 2008. *Music, Language, and the Brain*. Oxford University Press. New York.
36. Plailly, J., B. Tillmann & J.-P. Royet. 2007. The feeling of familiarity of music and odors: the same neural signature? *Cereb. Cortex* **17:** 2650–2658.
37. Bigand, E. & Y. Gerard. How many milliseconds to recognize familiar music? Submitted for publication.
38. Schweinberger, S., A. Herholz & W. Sommer. 1997. Recognizing famous voices: influence of stimulus duration and different type of retrieval cues. *J. Speech Language Hearing Res.* **40:** 453–461.