Abstraction of Two Forms of Underlying Structure in a Tonal Melody

EMMANUEL BIGAND
Laboratoire de Psychologie de la Culture, Université Paris x Nanterre, 162 rue Saint-Charles, F-75740 Paris Cedex 15, France

This experiment, based on the work of Lerdahl and Jackendoff (1983), was designed to test the abstraction of two forms of underlying structures: "reduced structure" and "prolongational structure". A family of four melodies (A1.B1.C1.D1) with the same underlying structure but with different rhythmic-melodic organisations was presented to subjects twice in the first step of the experiment. In the second step of the experiment, these melodies were then interspersed with four analogous melodies (A2.B2.C2.D2) differing in their underlying structure. Listeners had to identify melodies belonging to the first family. By presenting either a real family (A1.B1.C1.D1) or a false family (A2.B2.C2.D2), the subject's success in abstracting underlying structures could be observed. Results indicate that this abstraction does occur for the two forms of underlying structure studied, and that it is performed by non-musicians as well as musicians. This experiment thus reveals a highly elaborate processing of musical phrases thoroughly consistent with the Lerdahl and Jackendoff model.

When attempting to study the perception of musical fragments comparable to passages existing in either popular or serious music, one very quickly realises that there are a great many factors that it is absolutely indispensable to take into account and control. The very diversity of experimental studies carried out in the psychology of music illustrates the breadth of the problem. One can, however, split the factors into two categories.

The first is for all the organisational factors related to the hierarchies established by the tonal system. The second category contains all the organisational factors related to the physiognomic characteristics of the musical surface (melodic contours, rhythms, intensity, articulation, etc.).

This distinction is not new. R. Franèes (1958) considered two forms of musical punctuation: "the one entirely linked to a particular culture, in which the temporal signals of harmony, once learnt, are easily interpreted", "the other is intuitive, without the aid of conventions, based on the varieties of melodic motion, through its pauses or interruptions, subject to misinterpretation of the significance of both" (p. 182). Franèes' Experiment VIII (the structuring of musical discourse) makes it possible to observe the effects of these two forms of organisation in subjects with differing musical education. Let us first consider each of the categories separately.

A great many studies have shown that Western listeners have a certain intuitive knowledge of the organisational principles specific to the tonal system. Thus, studies have shown the internalisation of hierarchical relationships between the notes in a given key (Krumhansl, 1979; Krumhansl and Shepard, 1979; Krumhansl and Kessler, 1982), as well as between chords in a
given key (Krumpolz, Bharucha and Castellano, 1982; Krumpolz, Bharucha and Kessler, 1982).

For example, it has been observed that the psychological distance between chords based on the strong degrees is smaller than between chords based on weak degrees. Bharucha and Krumpolz (1983) observed that these relationships of closeness between chords in the same key vary when the harmonic context changes. C and G chords will be perceived to be closer to each other if the harmonic context is C than if it is F# major. This result points to the internalisation of relationships between keys.

This knowledge of the hierarchies of the tonal system enables the listener to abstract the harmonic structures underlying a musical surface. The capacity to extract harmonic structures is crucial, because it informs the listener about the network of tensions and relaxations that give the piece its structure (cf. Lerdahl and Jackendoff, 1983). It may be added that this ability is not limited to the case of harmonised pieces. Indeed, several experiments in melodic perception show the existence of extraction of implicit harmonic structure (Francès, 1958; Tan, Aiello and Bever, 1981; Bharucha, 1984[2]).

The above-mentioned experiments show that this knowledge exists in listeners "with substantial exposure to music but with little explicit knowledge of musical structure" (Bharucha and Krumpolz, 1984). However, in most of these studies, the stimuli contained no information on rhythm, dynamics, vibrato, etc. They are therefore very different from real listening situations. Thus, this knowledge of the rules of the tonal system exists tacitly (cf. Bharucha, 1984[2]) and it would be necessary to study how it would be put into application in more complex musical situations.

Now that is precisely where things become complicated. At musical surface level, there are other types of information which are independent from the pitch hierarchies established by the tonal system, and which make the temporal organisation of music possible.

At the most elementary level, this information contributes to the constitution of groups. Grouping is an essential characteristic of perception (Fraisse, 1974), and at this level is based on the organisational principles of the perceptual field as described by Gestalt psychologists. These principles "seem to describe aspects of stimulus organisation that arise automatically from the operation of the sensory systems, without involving more complex cognitive systems such as memory" (Dowling and Harwood, 1986, p. 154).

For example, several notes can be perceived to be linked together, either because of proximity in pitch, or because the succession of pitches defines a steady melodic contour (principle of "good continuation"), or again because of their closeness in time. Establishing these groups is not, however, a straightforward phenomenon, since several grouping principles can conflict with each other. Lerdahl and Jackendoff's preferential grouping rules "establish not inflexible decisions about structure, but relative preference among a number of logically possibilities analyses" (1983, p. 42). I. Deléage (1987) carried out two experiments with musician and non-musician subjects, and observed that "Lerdahl and Jackendoff's grouping preference rules constitute theoretical principles backing up strongly the grouping intuitions of a subject while hearing music" (p. 356).

Another aspect of the grouping phenomenon was observed by Fraisse (1974). When one listens to a succession of identical sounds at regular intervals, one cannot help perceiving them as divided into groups of two or three even though no objective factor justifies such grouping. This subjective organisation strongly indicates the importance of the periodic recurrence of metrical stress. The regular succession of beats seems to constitute a framework comparable to the cognitive frame established by the scale where pitch is concerned (Monahan, 1984, quoted by Dowling et al., 1986). The framework enables the listener not only to appreciate duration patterns (Povel, 1981; Essens et Povel, 1985 (exp. 3)), but also has an influence on the perception of melody. "The closer the stresses are to each other, the more structure they give the melody by virtue of the great number of time indications they give the subject; the rarer they are, the more difficult it becomes to grasp the melodic structure" (Francès, Imbert, and Zenatti, 1979, p. 180).

Thus, two main categories of organisational factors can be defined. Both categories interact in a real musical situation. For example, Francès (1958) observed that even with very experienced musicians, certain cadential melodic contours can suggest the idea of a cadence, even though nothing justifies this harmonically. Sloboda (1985, p. 189) mentions the importance of metrical stress for key attribution in a chromatic passage. Finally, one can point to cases where there is conflict: for example, memorisation of a melody containing modulations is less successful when the rhythmic structure breaks the melody into groups that do not correspond to its harmonic regions (Bignard, 1989).

The problem is therefore to work out how one gets from the perception of the musical structure, made up of sounds of differing pitches, durations, intensities, timbre qualities, to the extraction of underlying harmonic structures, indispensable for proper understanding of tonal pieces.

The model developed by F. Lerdahl and R. Jackendoff (1983) is an important theoretical contribution to this question.

Four components enter into the structural description of the perception of a piece. "Grouping structure" designates "hierarchical segmentation of a piece into motives, phrases and sections." "Metrical structure" designates "the intuitions of the events in the piece are related to a regular alternation of strong and weak beats at a number of hierarchical levels." "Time-span reduction" "assigns to the pitches of a piece a hierarchy of structural importance with respect to their position in grouping and metrical structure". "Prolongational reduction" "assigns to the pitch a hierarchy that expresses harmonic and melodic tension and relaxation, continuity and progression" (Lerdahl and Jackendoff, 1983, pp. 9 and 10).

Let us now deal more particularly with the last two components. The concept of reduction is based on the hypothesis that "the listener attempts to organise all the pitch-events of a piece into a single coherent structure, such that they are heard in a hierarchy of relative importance" (p. 106). The concept of reduction presents formally the way in which the listener manages to separate ornaments from what is important structurally, and points out that "the structurally less important events are not heard simply as insertions, but in a specified relationship to surrounding more important events" (p. 106).
How can the reductions that a listener will infer from the musical surface be described? The model postulates that the rules governing the pitch system are inadequate. It is therefore indispensable to consider the metrical and grouping structures. These structures "serve a double function in constructing reductions: they segment the music into rhythmic domains, and within these domains, they provide rhythmic criteria to supplement pitch criteria in the determination of the structural importance of events" (p. 119). Thus, by integrating these two structures, reductions formalise the interaction of the two main categories of organisational factors mentioned above.

Despite the foregoing, reductions seem to be totally dependent on the segmentation established by the metrical and grouping structures. They can under no circumstances work outside of this framework, which makes them unable to express relationships established between the groups. In order to deal with this problem, Lerdahl and Jackendoff define a fourth category: "prolongational reduction".

Here, the aim is to describe the kind of musical progression that a listener will detect in a given piece: how the music moves from one point to another, where the point of maximum tension (climax) is situated, when a resolution occurs, etc. The main hypothesis is that "intuitions about pitch-derived tension and relaxation in a piece can be expressed in terms of a strictly hierarchical segmentation of the piece into 'prolongational regions', such that (a) each region represents an overall tensing or relaxing in the progression from the beginning to its end, and (b) tensings and relaxings internal to each region represent subordinate and non-overlapping stages in the overall progression" (p. 211).

How is prolongational reduction established? It has to be derived from the time-span reduction—there are two reasons for this. Firstly, in order for such connections to be established between all the elements of a piece in its entirety, it is indispensable for the listener to have an idea of the successive stages and the analytical fields that define the elements available for prolongational connection. Prolongation therefore first considers the elements of the last time-span reduction level. Once they are connected to each other, these elements constitute the basic structure, similar to Schenker's Ursatz. Furthermore, this dependence is in fact necessary, given that the rules of harmony alone do not constitute adequate criteria for establishing connections. Other factors, for example rhythm, must also be taken into consideration.

This great dependence of prolongational reduction on time-span reduction constitutes a major aim of Lerdahl and Jackendoff's theory. "It asserts that the perceived patterns of tensions and relaxations in pitch structure depend crucially on the hierarchy of structurally important events within time-spans defined by meter and grouping" (p. 188). Prolongational reduction is thus derived from time-span reduction, which is itself associated with a particular musical surface.

Thus, the model formalises the interaction between the organisational structures detected at surface level and the more fundamental ones linked to the rules of tonal harmony. It therefore constitutes a vital theoretical contribution to the question raised at the beginning of this study.

To what extent can this theory be used for experimental research?

Consider for example melody a1 (Fig. 1). The metrical and grouping structures segment the musical surface in a non-conflicting fashion. We will consider the groups at the level of individual bars (3rd metrical level). The time-span reduction rules provided by Lerdahl and Jackendoff's model make it possible to predict which elements will be predominant in each group. Rules TSPR1 (metrical position), TSPR2 (local harmony), TSPR3 (registral extremes) and TSPR8 (structural beginning) define the note A as the dominant element of the first bar. However, by virtue of the rule of fusion TSWFR3b, E and C could be superimposed upon the A, which would then give us the A C E triad. Rules TSPR3, TSPR6 (prolongational stability) and
TSPR7 (cadential retention) would select G♯ as the predominant element of the second bar. The same rule of fusion could then add to the G♯ the notes B and D', which would give us an E 7ths chord without the root. Similarly, the note A (or the interval A-C) would be chosen for the third bar, and G (or the interval G-B) for the fourth (particularly as a function of rules TSPR4 (parallelism) and TSPR7). In bar 5, rules TSPR3, TSPR1 and TSPR7 would lead one to choose C (or the G-E-C chord). The choice for the bar 6 would be more ambiguous in the case of the first melody, since TSPR1 and TSPR5 would select G, whereas TSPR4 would lead one to choose B. TSPR3 and TSPR2 would not allow one to choose between them. In fact, it is probable that the rule of fusion would lead one to choose the interval G-B. The last three notes would be A (or the interval A-C), G♯ (or the interval G♯-B), and A.

If we were to analyse the other three melodies, which have very different rhythmico-melodic structures, we would find the same elements: A G♯ A G C G A G♯ A, or, more likely, by virtue of TSWFR3, the chord sequence A minor E 7th A minor G C A minor E 7th A minor. This sequence of elements constitutes what may be called a "reduced structure".

The family relationship between these melodies is basal from the point of view of musical theory, but it certainly is not from the point of view of perception. The melodies are different in melodic contour, in rhythms, in metrical structure, in intervals and in thematic cells. All they have in common, therefore, is this reduced structure. If the listener does not carry out any reduction (or arrives at reductions based on criteria totally different to those expressed by the rules quoted above), he should hear four different melodies. Otherwise, he ought to sense that these melodies are related to each other.

Now, what would happen if these melodies were played in different keys? The elements constituting their reduced structures would be different in each case (see Fig. 2). However, it is probable that a listener (especially if he is a musician) would sense the relationship between them. It is therefore necessary to define a different, more abstract form of underlying structure, based on the degrees of the implicit harmony underpinning the melodies (see Fig. 3a).

\[
\begin{array}{cccccccc}
I & V & I & V & I & V & I & I \\
T1 & T3 & \\
(a) Implicit harmony of F1.
\end{array}
\]

\[
\begin{array}{cccccccc}
I & V & V & I & I & V & VI & V \\
T1 & T3 & T5 & \\
(b) Implicit harmony of F2.
\end{array}
\]

Fig. 3

Implicit harmony of F1 and F2.

This structure no longer takes into account only what happens within each group. Notation in degree form displays the relationships between the groups. This notation is very much more than a simple musicological algorithm: it represents the complex network of tensions and relaxations in our family of melodies. This network can be represented in graph form (see Fig. 4a).

This network indicates that the main tension of the melody appears in the fourth bar, and is gradually resolved onto I. Another harmonic framework (see Fig. 3b), just as plausible musically, would present a different pattern (see...
This group of melodies has rhythmico-melodic contours very similar to those in the preceding family \((a_1\) similar to \(a_2, b_1\) to \(b_2\), etc.), but what Group 2 melodies have in common is a different underlying harmonic structure, \(F_2\).

The melodies were recorded by a professional oboist who was asked to use exactly the same phrasing and tempo for each pair: \(a_1\) to \(a_2\), \(b_1\) to \(b_2\), etc. The melodies were all first recorded in the same key \((A)\). Then they were recorded in 12 different keys in the following way: The Family 1 melodies were recorded in four different keys (\(a_1\) in \(A\) minor; \(b_1\) in \(E\) minor; \(c_1\) in \(B\) minor; \(d_1\) in \(F\) minor), then again in four other keys (\(a_1\) in \(C\) minor; \(b_1\) in \(E\) minor; \(c_1\) in \(D\) minor; \(d_1\) in \(G\) minor). The Family 2 melodies were recorded in the remaining keys (\(a_2\) in \(C\) minor; \(b_2\) in \(E\) minor; \(c_2\) in \(B\) minor; \(d_2\) in \(A\) minor). The recordings were then played to three judges (musicians and musicologists), not acquainted with this research, who considered that the melodies were comparable from the point of view both of esthetics and performance.

**Procedure**

Family 1 \((a_1, b_1, c_1, d_1)\) was presented twice to the listeners; after that, all melodies, mixed Families 1 and 2 in a controlled order. Each listener was given the following instruction: “You are going to hear a family of four melodies twice. The melodies appear to be different, but in fact they are all related to each other in the same way. Listen to them carefully, because afterwards you will hear them again, but interspersed with four other melodies that sound very similar to the ones in the family, but which are not in fact related to them in the same way. You must then indicate whether each melody belongs to the family, or is an outsider.” There is a pause of 15 seconds after each melody. Each listener therefore gave eight answers, and the number of errors he made was recorded.

However, if each melody in Family 1 is compared to the corresponding one in Family 2 \((a_1\) to \(a_2\), etc.), it can be observed that the look-alike melodies do not always have strictly identical rhythmico-melodic contours. Such distortions are inevitable if one is to respect the harmonic framework and at the same time give the melody a plausible musical character. However, these distortions are at no time systematic. The slight differences in contour between \(a_1\) and \(a_2\) do not recur between, for example, \(b_1\) and \(b_2\). These distortions do not therefore constitute a criterion for discriminating between the two families. Furthermore, it must be noted that they are minimal compared to the great differences in contour observable in the melodies within each individual family. This remark is also true for comparisons based on interval size. These differences are not systematic, but they facilitate memorisation of individual melodies. This phenomenon must therefore be neutralised by the experimental design.

**Experimental design**

Three factors are studied:

**Factor A Mode of presentation**

\(A_1\) True Family (TF). In this case, a true family is presented twice in a row, then all eight melodies are presented, mixing the families.
A2 False Family (FF). Here, a false family \((a1, b2, c1, d2)\) is presented twice in a row, and then the eight melodies are presented mixed. Thus, there is a relationship between the underlying structures in TP presentation, and no relationship in FF. Thus, individual memorisation is controlled for.

Factor B Form of the underlying structure

B1 Reduced structure and prolongational structure. In this case, all eight melodies are played in the same key. The relationship between them is defined by similarity in both reduced structures and prolongational structures.

B2 Prolongational structure alone. Here, all eight melodies are played in all 12 keys. First, Family 1 is exposed twice in a row in four initial keys. Then, it is presented in four new keys, mixed in with the Family 2 melodies, which are themselves played in the four remaining keys. The relationship can only be established on the basis of similarity in prolongational structures.

Factor C Subjects

C1 Non-musician subjects. These were university students with no musical knowledge, and who had never learnt to play a musical instrument.

C2 Musician subjects. These were students, all of the same age, in their last year of “solfège” (sight-singing and dictation) and musical analysis, at the National Conservatory in Boulogne (near Paris).

We have here a 2x2x2 factorial design with eight independently-processed groups, and 20 subjects per group; the experiment was therefore carried out on 160 subjects.

Experimental Hypotheses

The main hypothesis deals with Factor A: “If listeners perceive the similarity in underlying structure, then fewer errors will be observed in A1 (where this similarity exists) than in A2 (where it does not exist).” If no difference between A1 and A2, this could be interpreted in two different ways, depending on whether the error rate is close to chance level: a case like this could mean that the subjects find nothing allowing them to distinguish between the melodies, and therefore that they do not manage to go beyond superficial organisational structures. On the other hand, a significantly lower error rate would mean that the subjects do manage to distinguish between the melodies, not, however, on the basis of the underlying structures, but on the basis of individual memorisation of the melodies.

Factor A acts in a sense as a test, so the effects of the other factors will be studied as a function of the main effect of A and interactions with it. In particular, one would expect the effect of A to be greater with musicians than with non-musicians (AC interaction). Similarly, one would expect the effect of A to be greater with B1 than with B2 (BC interaction).

Effect of Factor A: When a true family is presented, the subject makes fewer mistakes. The presence of an underlying structure common to all four melodies improves the subjects' results. Thus, the effect of this factor reflects the phenomenon of underlying structure abstraction.

Effect of Factor B: The effect of this factor indicates that when the melodies are transposed, the subjects are less successful. However, the absence of interaction between Factors A and B means that this perturbation effect is the same whether a true or a false family is presented. Thus, transposition had a globally perturbing effect. On the other hand, this lack of AB interaction indicates that the effect of Factor A does not vary with the form of the underlying structures, which means that the listeners manage to abstract both structure types successfully.

Effect of Factor C: Musician listeners do better in the experiment than non-muscians. Here too, however, the absence of interaction means that the effect of Factor A remains identical for both populations: the abstraction of the underlying structures was successfully performed by both musicians and non-musicians. The absence of second order interaction makes it possible to affirm that this is true for both forms of underlying structure. Musical training globally facilitates success in this experiment, though it is not indispensable to the abstraction process.

These results require thought on another point. If one looks at the results of the FF group, one notes that they are different from what they would have been if people had answered at random, which indicates that the subjects do manage to find reference points enabling them to recognise the melodies. Given that there is no single dimension common to all four melodies, it is probable that the listeners remembered traces of each of them. The results obtained with FF could then be explained by individual memorisation of the melodies. The experiment does not make it possible to detail the precise
nature of the memory traces, and that there would probably be many
different kinds of them (absolute pitch of the notes, specific timbres of certain
melodic cells, amplitude of the intervals, etc.). Given that this phenomenon is
observed with FF, it is possible that it would occur with TF as well, and that it
would be superimposed on the process of abstraction of underlying structures.
This, the latter would not be the only process involved in performing the
task.

Interpretation

A listener is thus able to abstract an underlying structure common to four
melodies that have different rhythmico-melodic contours. Moreover, he can
distinguish between two families of melodies with comparable surfaces, on
the basis of their underlying structures. These results also mean that a listener
can go beyond the level of superficial organization to reduce the musical
surfaces to a minimal underlying structure representing them.

The experiment makes it possible to postulate the existence of two kinds of
underlying structure: one of them, the "reduced structure", is defined by the
predominant elements within each of the melody's groups; the other, the
"prolongational structure", is defined by the network of tensions and relaxa-
tions induced from the harmonic framework.

Both underlying structures can be abstracted by musicians and non-
musicians alike. Musical education does lead to better performances globally,
but it does not seem to provide musician listeners with a specific competence
absent in non-musicians.

The lack of great differences between the two populations can perhaps be
explained by the fact that the underlying organizational principles are fairly
clear: the harmonic frameworks are very different, and there is total
concordance between the rhythmical and metrical groups, the dominant
elements in each group, and the network of tensions and relaxation. The
stimuli used are in fact straightforward: their underlying structures are quite
without ambiguities, which could explain the good performances of the non-
musicians. However, the reactions of both groups are very similar: all the
subjects found the experiment difficult, and very few of them manage to put
into words what determines their answers: the majority recognize that they
answered intuitively. This is fairly surprising, since it was expected that the
musicians would make conscious use of their theoretical knowledge to
understand how the melodies were constructed. This observation confirms
the fact that what is obvious from the point of view of theory is much less so
from the perceptual point of view.

It also seems to indicate that both populations use the same listening and
analytical strategies. All the subjects who participated in this experiment
waited for the end of the melody before deciding on their answer, and in most
cases only wrote it down a few seconds later. Going on the first tests carried
out for the preparation of this experiment, I observed that the lapse of time
between the end of the melody and the moment the subjects wrote down their
answers could be fairly long, thus justifying a pause of 15 seconds between the
melodies. This observation leads to the conclusion that all the types of
information gathered in the course of the melodies are necessary to determine
the answer.

Discussion

The results of the experiment confirm the general hypothesis on reduction
established by Lerdahl and Jackendoff: the listener manages to hear melodies
belonging to the same family as variations on an underlying pattern which is
more important than the differences observed on the musical surface.

But that is to consider only the most banal aspects of the hypothesis: these
results do not confirm the postulate that all the elements in a piece are heard
in terms of a strict hierarchy of structural importance, nor that the tensions
and relaxations are organised in accordance with a strict hierarchy. They
simply indicate the abstraction of two kinds of underlying structures at one
level of the hierarchy. They therefore confirm a very elementary theoretical
point, which had to be verified before envisaging experiments that would test
the model in greater detail.

Future experiments should develop these results: to begin with, the
hierarchical character of each of these structures should be determined.
Thereafter it would be important to determine the role played by the rule of
fusion TSRWFPR3b in the reduced structure of the melody (see Bigand,
1988). Finally, one ought to examine the psychological validity of the rules
enabling one to describe and predict the reductions made by a potential
listener. That was not the initial aim of this study, but the results of the
experiment lend a certain degree of psychological validity to the rules used for
the construction of the melodies. However, the rules played by metrical and
grouping structures must still be checked. Lerdahl and Jackendoff stress that
"the listener's understanding of pitch connections in a piece is a function of
how he segments its surface" (p. 183). The tonal functions of certain notes or
intervals are determined by the positions of these notes in the rhythmical and
metrical cells. Thus, small rhythmical modifications carried out on the same
series of notes could alter the underlying structures considerably.

![Fig. 5](image-url)

Melody c1 and its conflicting "reduced" and "prolongational" structures.

For example, what would happen if the rhythm cells were changed as in
Fig. 5? Clearly, the groups are different, and some of them do not tie in with
the metrical structure. There is much conflict in choosing the dominant
element. The reduced structure is very different from what it was in c1. The
G♯ in the first bar seems more ornamental than structurally important. Generally speaking, all notes structurally important for prolongation are less important within their groups. This conflict situation gives rise to an ambiguous prolongational structure. Changes of this kind can be carried out on all the melodies. From the point of view of Lerdahl and Jackendoff's theory, the underlying structures would no longer be considered similar. In that case, the same experimental task would probably be much more difficult to perform (such an experiment is in preparation).

Conclusion

This study provides elements with which to answer the question raised in the introduction. It shows that in both musicians and non-musicians, very elaborate musical phrase processing takes place. When the various musical components do not conflict with each other, the activity of processing enables the listener to abstract underlying structures that are clear enough for him to discern relationships between melodies with different rhythmico-melodic structures and phrasing. This aptitude is crucial for the understanding of music, since without it, it would be impossible, for example, to grasp the relationship between a theme and its variations, nor to appreciate, on the other hand, the specific contribution of each individual variation.

References


