GLUING TONES

GROUPING IN MUSIC COMPOSITION, PERFORMANCE AND LISTENING

Edited by Johan Sundberg

GLUING TONES

GLUING TONES

GROUPING IN MUSIC COMPOSITION, PERFORMANCE AND LISTENING

Edited by Johan Sundberg

The figure on the cover is a modified version of figure 9 in W. Jay Dowling's article. It shows the melody *Twinkle, twinkle little Star* (squares) together with distractors.

ISBN 91-85428-73-6
ISSN 0347-5158

Cover and layout: Åsa Wallner
Printed by Tabergs Tryckeri AB, Taberg 1992

GLUING TONES

GROUPING IN MUSIC COMPOSITION, PERFORMANCE AND LISTENING

Contents

Preface

On a sunny day in late May 1991, the Music Acoustics Committee of the Royal Swedish Academy of Music arranged a full-day public seminar. The theme was called *Musician's Tone Glue* and it concerned the phenomenon that when we listen to music we hear that some tones belong together while others do not. This seems to be of fundamental significance to music communication.

Composers carefully mark, e.g. by pitch jumps, tone length and harmonic formulae, what notes belong together and form groups, and where the boundaries are between different groups of tones. These aspects were discussed by Gerald Bennett, Professor of Music Composition at the Zürich Music Conservatory.

When musicians play music, they insert micropauses, lengthenings, shortenings, tempo changes and many other details to show the listener where the boundaries are, which notes and chords are remarkable and which are not. In this way they facilitate the listener's understanding of the piece. This topic I discussed myself in my contribution.

Music is probably the only form of human culture that is appreciated by very nearly 100% of the population in all civilizations. Of course the types of music that people prefer to listen to varies considerably, but almost everybody wants to listen to music every now and then. Further, even small children like music. What is the reason for this? A factor of relevance might be that music and language/speech share significant characteristics. For example, both are communicated by means of acoustic signals that are processed by the auditory system. Also, both have syntax, grammar and hierarchical structure. Are the reasons for all theses parallels that both explore innate potentials of the human perception mechanism? Carol Krumhansl, Professor of Psychology at Cornell University, Ithaca, New York, reviewed her own and her fellow researchers' investigations of infants' perception of musical and speech structures.

The composer's and the performer's hints as to musical structure trigger psychological processes in the listener. These processes were considered by Jay Dowling, Professor of Psychology at the University of Texas in Dallas.

His vast experimental work has revealed that the processes depend on the listener's previous knowledge providing a basis for expectancies.The listener also has a special tool that both composers and performers play on, namely the possi-bility of selective attention, i.e. the focussing of one's attention on one particular tone sequence imbebbed in a great number of other competing sequences.

Last, a panel discussion among the authors was arranged with Professor Hans Åstrand as the moderator. All these contributions are published, together with a phonogram record with sound examples, as the fourteenth volume in this music acoustic series of the Royal Swedish Academy of Music. I am convinced that it will offer musically interested readers fascinating perspectives on the challenging fact that we all want to listen to music.

Vaxholm, June 1992

Johan Sundberg

The Sense of the Phrase - Compositional Grouping in Music

Gerald Bennett

Grouping is something that composers have gotten quite good at over the centuries. In fact, much of composition, as is the case for any other art, has to do with devising interesting ways of grouping the material with which one has chosen to work. Music shares with spoken language - and sign language - temporal grouping; various mechanisms serve to articulate the stream of sound into segments the perception can deal with. I would like to describe here several different mechanisms of grouping in music, illustrating them with musical examples of several different situations. I shall begin by reviewing techniques for grouping that we all know from the common practice period of classical music, but I will also discuss examples of both early and contemporary music, in order to discover techniques which are not specific to any one style. I intend to argue that while there are certain specific mechanisms for the articulation of music in time, the use of these mechanisms almost always conveys an expressive intent and that hence quantitative measurement of these mechanisms of grouping must always be accompanied by a discussion of the expressive intent at hand. I stress that I am not talking directly about psychological strategies of grouping, but rather of specific techniques composers use when writing their music. I am not a scientist, and I intend to leave any broad conclusions about the mechanisms I demonstrate to the musical acousticians and the psychologists of perception.

Figure 1 recapitulates some of the most important techniques of musical grouping. I shall discuss each point separately. Let us consider melody first.

Sound Example 1 shows synthesized tones of very similar amplitude and equal duration and no musical grouping:There is no acoustic clue whatsoever as to how these ones should be grouped, but so great is our need to organize this flow of sound that only the least suggestion is necessary to let us hear a grouping of four, for instance, or of six. Note how much more difficult it is to hear uneven groupings, like five or seven.

Important factors in composition that assist grouping

- **in melodic context:**
  tonic accent
  high tone accent
  long tone accent
  (metric accent)

- **in harmonic context:**
  key, or consistency of pitch material where key does not apply
  cadence (tension-release archetype; open/closed
  opposition) harmonic rhythm (change of harmony)

- **coherence of texture**

- **development of texture**

- **repetition**

- **simple archetypical patterns of experience**
  (tension-release, increase-decrease, etc.)

- **starting and stopping**

- **reference and expectation**

Figure 1. Compositional factors which aid grouping.

Grouping of a succession of tones like this is done by accentuation. There are several kinds of accent; three of the most important are: tonic accent, high tone accent and long tone accent. **Sound Example 2** gives a straightforward example of tonic accent. Every fifth note is a bit louder and we group into units of five notes each.

Sound Example 3 shows high tone accent. Here the notes are all of similar amplitude, but every fourth note is about a minor third higher than its neighbors. Our perception groups accordingly:

Sound Example 4 shows long tone accent.
Of these examples, only tonic accent is accent as we usually understand the word. In the other two examples, all the notes had very similar amplitudes. Any accent we seem to hear is produced by the perception.

This kind of accentuation works to create opposition, and hence to suggest divisions in the flow of sound. But these are specific aspects of melody. A tone lower than its neighbor tones also creates an opposition, but generally receives less of a perceptual accent than does the high note, unless in the lowest voice. A note shorter than its neighbors also serve to mark a group, but the sense of accent is passed on to the next note.

For the music we know best, that written between 1700 and 1900, we must also mention metric accent, which is the tonic accent given to certain beats in each kind of meter (the first beat in a measure of three, first and third beat in a measure of four, etc.).

Very little music in the western tradition is purely melodic, and so we must also consider harmony. A basic concern of harmony is key, and key is certainly the most important help to grouping, at least in music written between 1600 and 1900. The chord built on the fundamental of a scale is called the tonic and is perceived as the tonal center of the passage using that scale and as a stable sound. The chord built on the fifth note of the scale is called the dominant, is usually followed by the tonic and is perceived as relatively unstable. Most phrases in the common practice period end on either the tonic (closed) or the dominant (open). The contrast between open and closed endings is an essential indication of grouping in all harmonic music.

The primary harmonic event is the cadence, often in the form dominant-tonic. The significance of the cadence is the movement from tension to release. This pattern of tension and release is one of the most important means of indicating grouping, also in situations that have nothing to do with harmony. Most often the movement is from calm to tension to release; the release signals the end of a group. This pattern is one we know well from everyday life and is therefore especially important as a means of grouping.

Harmony too can generate accent by the change from one chord to another. In particular, one tends to hear the tonic as accented in the progression dominant-tonic.

Other principles of grouping of such a general nature as to be valid for all music in our tradition are:

• Coherence: A texture or a pattern remains the same over time.

• Development: One or more aspects of a texture or a pattern change regularly over time.

• Repetition: The repetition of a pattern, phrase, etc. marks that which is repeated as a group.

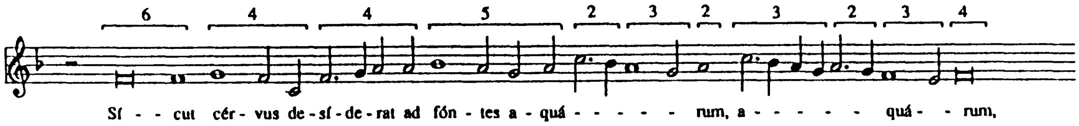
• Increase/decrease: A few patterns of movement seem archetypical and are important in helping to group. The most important I call increase/decrease. This might refer to melodic movement (rise/fall), dynamics (soft-loud-soft), or any other dimension of music.

• Starting and stopping: In much music where conventional aspects like harmony, meter, melody, etc. are missing, the actual beginning and end of sound is an important indication of how to group.

• Reference and expectation: Music can both refer back to past events, thereby indicating grouping, and create expectation, so that one has an indication of how to group what one is hearing. Although these are very important techniques, I shall not speak explicitly about them here, but I do direct you to Jay Dowling's article in this publication, which deals directly with both these aspects of music perception.

I shall now discuss examples of these grouping techniques. The first example is melodic and dates from the middle of the 16th century. It is the soprano voice of a famous motet, *Sicut cervus*, for four voices by Giovanni Pierluigi da Palestrina.

To begin with, this melody is a good example of both high tone and long tone accent. Palestrina wrote without barlines, and the accents divide the melody into groups of varying length. (I should say again that the "accents" are perceived accents, caused by the higher and longer tones, although it is altogether likely that the singers reading this melody would unconsciously add tonic accents of their own.) But there is another important kind of



Score Example 1. The Cantus (soprano) part of the beginning of the motet *Sicut cervus* by Giovanni Pierluigi da Palestrina (middle of the sixteenth century).

grouping going on here. The small, low-level groups are brought together into a perceived larger group, precisely a melody, by development in the pitches and in the movement of the music. Consider the pattern the long and relatively higher notes make: the slow ascent from the fundamental of the key to the dominant, then the balancing, but more rapid descent back to the fundamental. And see, too, how the movement within the melody becomes more rapid towards the end and then slows down again before the cadence. Here we see a clear example of the increase/decrease archetype. The phrase structure of this melody expresses great concern with the intelligibility of the text and attentiveness to a very anthropomorphic sense of balance in the melodic line itself.

Here is a second example, some 120 to 130 years older that the first, the two-part beginning of a motet by Guillaume Dufay, Nuper Rosarum Flores, written for the dedication of the cathedral at Florence in 1436 **(Sound Example 5).**

At the lowest level we see the same high note and long note grouping mecha-nisms as in Palestrina, here perhaps emphasized by the melismatic nature of the music. As with Palestrina, we find similar divisions into groups of unequal length, distributed differently between the two voices. The only place where groups in the two voices have the same boundaries is at cadences, which with their dissonance-consonance pattern move very strongly from tension to release. A difference in Dufay's music is that the groups often become progressively and regularly shorter within one phrase. Although the melodies do not seem to be constructed with the same care to balanced ascent and descent, we do see both acceleration from the first to the third phrase and acceleration within some of the phrases. Besides these similarities, we find here something we would not have found had we analyzed a longer polyphonic section of the Palestrina motet: a structuring

Score Example 2. The beginning of the motet *Nuper rosarum flores* by Guillaume Dufay (1436). The brackets and numbers above the vocal lines show grouping by high tone and long tone accent. The numbers at the right show the number of half notes in each phrase.

into larger groups of equal length, that is, grouping by repetition. This short excerpt from a long composition is a good example of the interplay of several levels and techniques of grouping. It expresses in its phrase structure concern for a more abstract order than does Palestrina's motet. We do not find either the so directly anthropomorphic interest in melodic or rhythmic balance within the phrases.

Here is an example of other techniques of grouping in a style more familiar to you: the beginning of the piano sonata in G major KV 283 by Mozart **(Sound Example 6).**

We have here the first theme and the transition to the second theme of the first movement. The first theme has three parts; the third part is simply a repetition of the second. The first part of four measures is divided in two by the repetition of the rhythm of the right hand and, more importantly, by the harmonic relationship open-closed (dominant ending, tonic ending).

Score Example 3. The beginning of the first movement of the Sonata for Piano KV 283 by Wolfgang Amadéus Mozart (1774).

The second part begins after a texture (and register) change and is also divided into sections, one of four measures (with a cadence on the tonic) and, after another change of texture, one of three measures with a dramatic descent over an octave and a half to the tonic and a stronger cadence form. The end of the first and the beginning of the second section overlap, obscuring somewhat the boundary between the two groups. We find here the same acceleration of movement we found in Palestrina and Dufay. The transition begins after the first real rest in the piece with a change of texture, clearly separating the new motif from its surroundings. The motif is repeated and then repeated three more times, now with no rests and transposed up twice a third instead of a second (i.e. accelerated). In this example, the grouping has both a structural and an expressive function. Clarity of grouping is not necessarily required, some boundaries are clearer than others. Boundaries can be obscured for structural reasons (to create a longer phrase, as in the second part of the theme with its charming and surprising elision) or for expressive reasons (as in the transition, where

Wehmut

Score Example 4. The beginning of the song *Wehmut* from the *Liederkreis* op. 39 on poems by Joseph von Eichendorff by Robert Schumann (1840).

Mozart increases our longing for four-measure phrases in order to call our attention more strongly to the second theme and its regular structure).

But what does the grouping of this example express? The grouping here affects the outer shape of the music, which seems to me to make the music rather public, not intimate. The elision expresses playfulness and suppleness, while the exact repetition of the second phrase conveys a rhetorical formality, creating a certain distance between the listener and the music itself.

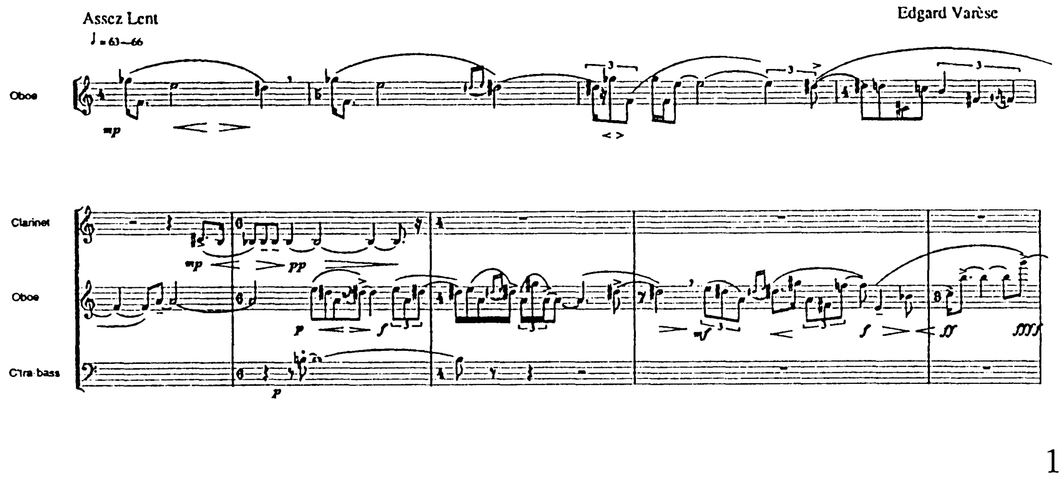
Let us turn now to an example of more complex grouping in the song *Wehmut* from the *Liederkreis* by Robert Schumann (**Sound Example 7**).

Here there are several, often conflicting, markings for grouping within a single phrase. These can best be seen as conflicting accents. The meter suggests a tonic accent on the first beat of every measure. The high tone and long tone accents of the melody suggest accents on the second beat of the second measure and on the first beat of the third and fourth measures. The suspensions in the accompaniment and the harmonic progression (dominant-tonic) suggest accents on the second beat of measure three and on the first beat of measure five. The ascending leap in the middle voice of the accompaniment in every measure suggests further accents. Here the outward phrase structure is much more regular than in the Mozart example: all the phrases in the song are four measures long. But the inner structure of each phrase is much more complex. This way of treating the markings for grouping has no structural function but is purely expressive. Here we find just the opposite situation from Mozart: great outward regularity and extreme inner excitation. This strong contrast and its symbolic significance let us understand the music as intensely personal, highly intimate expression.

Now I would like to discuss five examples of music from the twentieth century. They are of particular interest when thinking about grouping, because many of the techniques associated with earlier music - melody, harmony, regular metrical structure, etc. - take on somewhat different form here. The first example is from the year 1923, the beginning of *Octandre* by Edgard Varèse (**Sound Example 8**).

Obviously, we don't encounter here the subtleties of grouping and phrase that we found in the Schumann example. On the basis of texture we group

Octandre



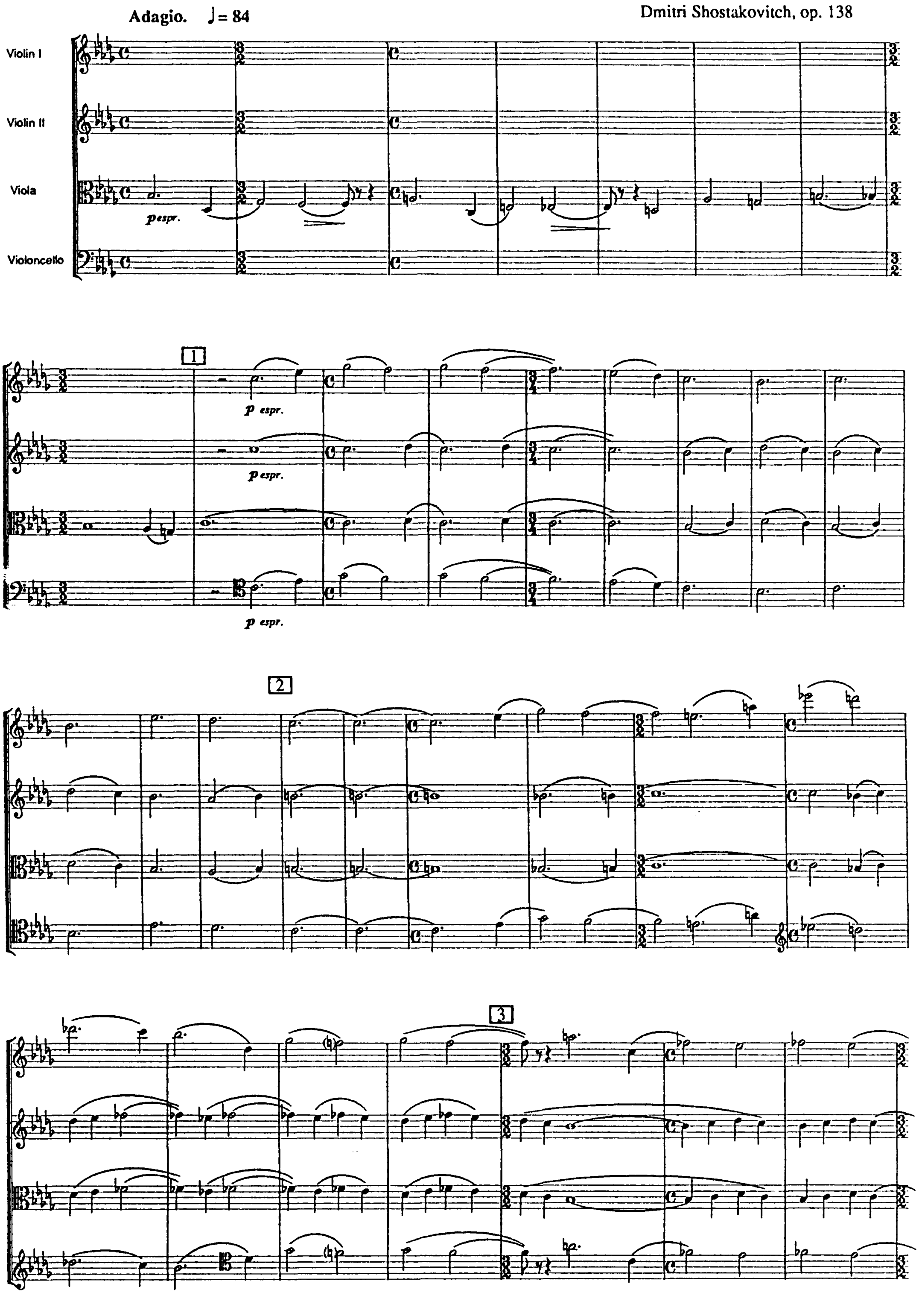
Score Example 5. The beginning of the first movement of *Octandre* for Eight Instruments by EdgardVarèse (1923).

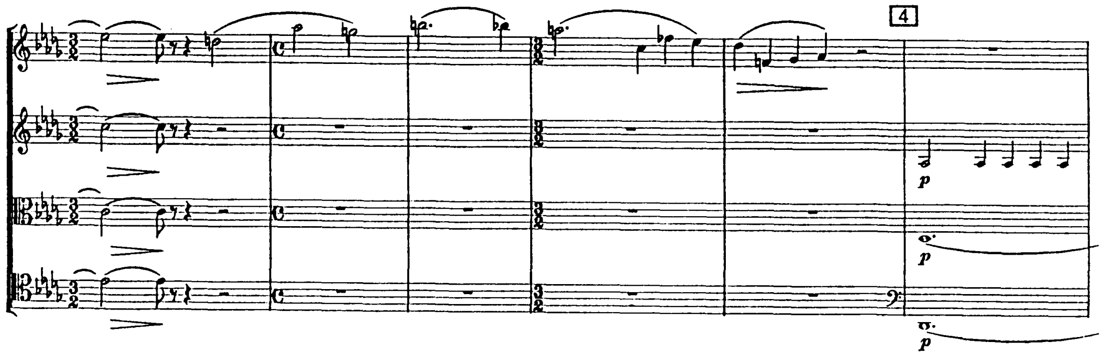
this passage into two large sections: the solo oboe and the rest. In the music for solo oboe we hear two phrases, in the following music we hear three of similar shape: here it is repetition which helps us to recognize phrases. The first phrase of the oboe we divide into three, or possibly four groups: a first motif, its repetition, then a second, varied repetition and extension with a slight pause between repetition and extension. Note that it is the break in the flow of sound at the end of measure one which is the decisive clue to the original grouping, and nothing in the pitches or their durations; the repetition reinforces our guess about how to group. The next phrase is similarly constructed of three or possibly four groups in the pattern: motif, varied repetition, second repetition and extension. The crescendo at the end leaves the phrase open, as though it would end on the dominant, and thus leads to the next large group. The three groups in the second part are articulated by their similar texture (held chord with large dynamic change); the ear takes its clues for grouping from the most prominent instrument, here the trumpet, although I suspect that Varèse counted on the high g-sharp of the flute to unify the two groups. The two short duos in different tessituras in the third group give a further articulation, the crescendo at the very end of the third group leads to the next section. This example is interesting, because it shows how general and undifferentiated the means for marking groups can become in the absence of a clear and strictly followed musical syntax. The expressiveness of this passage depends on things like the rhapsodic structure of the oboe solo with its many repetitions, the similarity to the oboe section in the structure of the ensemble section, the contrast between the solo oboe and the entire ensemble, the contrast in dynamics in the ensemble section and on the magic and evocative power of the pitches of the tutti chords. Most importantly probably, the expenditure of physical energy on the part of the players contributes greatly both to understanding how to group the music and to its expressive force. This is not a music where subtly conflicting signals about grouping lead us into an intimate subjective realm. The hardiness of the grouping mechanisms shows us a mysterious but nonetheless objective music of great physical intensity.

The next example is from the year 1970 and is very different. It is the beginning of the 13th String Quartet in b-flat minor by Dmitri Shostakovich **(Sound Example 9).**

String Quartet No. 13

Dmitri Shostakovitch, op. 138





Score Example 6. The beginning of the first movement of the *String Quartet No. 13* in *b-flat minor* op. 138 by Dmitri Shostakovich (1970).

I have played quite a long section, the entire first theme, in order to show something of the complexity of the grouping in this piece. The opening statement in the viola is clear enough: the rests, similarity of motif and phrasing, the return to the note b-flat and perhaps the aab form as well mark very clearly the grouping the composer intended. From rehearsal number 1 on, however, there is much that is unclear about the grouping. Cello and first violin play a melody clearly related to the viola's at the beginning, which cadences perhaps in the fourth measure after 1 (long note), or perhaps two measures later (lowest note), overlapping with the beginning of the next phrase. This phrase has its melody in second violin and viola. Like the viola solo, it has an aab structure with a clear cadence at 2 (long note). But what a cadence, holding the half-tone dissonance for nine beats! Four measures of transition lead to a reprise in first violin and cello of the opening melody, with the descending half-tone repeated at the end of the first two phrases. The second violin and viola continue the minor thirds of the measures before 2, first from d-flat to f-flat, then from b-flat to d-flat. In its final phrase alone, the first violin echoes the viola solo, but instead of slowing its movement to cadence on b-flat it speeds up and turns to d-flat to prepare the entrance at rehearsal number 4.

In this example many of the clues to grouping are obscured: the unclear ending of the melody after 1, the cadence at 2 with sharp dissonance instead of release of tension, the second violin and viola 3 measures before 3, which continue their accompaniment figure across the return of the first theme.

In comparison, Schumann's obscuring of group boundaries in the *Wehmut* seems quite harmless. I find this way of dealing with formal grouping extremely expressive. The main formal purpose of the obscuring of boundaries is to create longer sections. In this slow tempo, the relatively seamless music between the two solos is extraordinarily long in absolute time. This unusual length creates the ever stronger expectation that a phrase will finally be closed. The avoidance of clear boundaries also has a symbolic meaning which I will not try to paraphrase, but in which waiting and unresolved tension certainly play a role. Of course, other aspects of the music, particularly the harsh bareness of the two-voice texture with its stark octave doubling, the constant minor seconds used both melodically and harmonically and the insistant, hopeless, empty movement within the minor third are important in defining the expressive content of the music, but their effect would not be so strong if the phrasing of the music allowed one to relax between numbers 2 and 4. If I have emphasized the dissociative aspects of grouping in this example, it was to show that clear grouping is not necessary in good music. It is important however to point out that the obscuring of grouping boundaries presupposes the expectation that boundaries will be set. In this example the underlying formal movement is quite evident to analysis; its clarity is the prerequisite which allows the obscuring of boundaries to work at all.

Finally, I would like to play you three examples of electroacoustic music. Because most traditional compositional criteria do not apply to these pieces, they are particularly interesting in a discussion of grouping. The first example is both the oldest and the most radical, *Williams Mix* by John Cage, from the year 1952 **(Sound Example 10).** Here the composer has taken care to make grouping impossible, except when the sound actually stops for an instant.

This music is made of eight layers of sound, each of which consists of short fragments of sound chosen at random. The texture is so dense, and the fragments so short, that it is impossible to divide sections into shorter groups than those between two pauses in the sound itself. What effect does this have? By making symbolic interpretation impossible by avoiding traditional means of grouping, Cage turns our perception inwards towards the sounds themselves and not towards any relationships between them or to any meanings they might have.

The next example is by Pierre Schaeffer, the founder of *musique concrète*. It is taken from the piece *Etude aux sons animés* from the year 1958 (**Sound Example 11**).

This excerpt has three parts: the first is a kind of "curtain" that announces a new section, the second is a two-part episode, the third is the "curtain" again. The "curtain" consists of two main elements: a grainy tone, and a sound that seems to be the quick rubbing of a cymbal, perhaps with a metal rod. It ends with a crescendo of the cymbal in just the same gesture we heard in *Octandre* by Varèse. In the first part of the episode the sound of the cymbal is prolonged with continuous changes of pitch. Towards the end of the first part, we hear two single percussive sounds or groups of sounds. The first part is grouped together by a single tone having the same pitch as the cymbal from the "curtain"; its gradual disappearance signal the end of the first part. The second part continues the glissando idea of the first but uses the "rubbed" sound quality from the "curtain", along with other percussive sounds. Here too a single pitched sound leads directly to the repetition of the "curtain".

In this example, three factors from the list in Figure 1 determine grouping: coherence, development and the use of characteristic dynamic gestures. Coherence is given by the ostinato (the long, high-pitched tones in the episode) and by the use of only two timbres. The development of the section takes place in the realm of timbre, as Schaeffer finds ever new variations on both the grainy and the rubbed sounds of the "curtain". The characteristic dynamic gestures (in particular the fading in and out of the ostinati and the crescendi at the end of the first "curtain" and the episode) are archetypical movements we know not only from other musical contexts but also from everyday life. The remarkable thing about this example is how timbre plays a major structural role in defining groups. The grouping mechanisms are quite "classical" and straightforward, even if the materials on which they operate are strange to us. There is no conflict of signals here, no non-fulfillment of expectation, and while the piece seems to me quite expressive, the directness of the grouping makes for a very different expressiveness than in the Schumann song or the Shostakovich quartet.

The last example I would like to discuss is taken from the beginning of Region IV of the piece for tape *Hymnen* by Karlheinz Stockhausen from the years 1966/67 (**Sound Example 12**). In this example there are fragments from the Swiss national hymn sung by a choir over a background of very slowly moving sound. The Swiss hymn becomes more and more distorted as the piece goes on. This example is a bit longer than the others I have played, although I have shortened it considerably from its original form.

It seems to me very difficult to speak about grouping in this music. The time scale for the passage is given by the sound in the background: apparently aimless at first, it soon begins to descend, and every new entrance of background music will turn out to descend in a very slow, never-ending glissando. There is no help for grouping here; there are entrances of new music, but nothing in the background ever finishes. Only the fragments of the Swiss hymn, progressively more distorted, are bounded "on both sides", so to speak, and they too eventually become endless as later entrances are extended indefinitely, become part of the background and merge with the noise of the universe. Much of the poetry of this passage comes from its avoidance of grouping, and the techniques used by Stockhausen serve a very precise compositional and expressive intent: clear entrances, unclear endings, movement from foreground to background, continuous descent, progressive distortion, all of these have strong symbolic connotation. In the absence of grouping - we saw this in the Cage example - one's attention is directed to the sounds themselves, but if Cage took pains to strip the sounds of *Williams Mix* of sentimental or emotional content, Stockhausen does just the contrary, composing seemingly mechanical sounds of limited bandwidth and quite threatening character. Unable to group these sounds meaningfully, we are exposed to their full symbolic and expressive power.

The music's threatening character is certainly in part dependent on the sounds themselves with their narrow bandwidths and the suggestion of repressed energy. But what does the grouping express? The constant non-fulfillment of expectation casts a pall of desperation over this music. Not just the non-fulfillment itself is important - we have seen that in other examples. The continual impossibility of grouping is very distressing: things become boundless, and boundless things are frightening and worrisome. It would be interesting to reflect on the difference between the Shostakovich and the Stockhausen examples in terms of the non-fulfillment expressed in each. I personally find the Shostakovich more expressive, because there is fulfillment, but the "resolutions" are always distorted. I find the distortion more disturbing than the rather crude destructiveness of the Stockhausen.

In conclusion, I would like to stress the importance of conflicting signals about grouping and of the non-fulfillment of expectation about grouping. At the very least, these techniques make us aware of tensions and activity not confined exclusively to the aesthetic realm. Obscured grouping alerts us as well to other expressive aspects of music. We all know what a complex and often nefarious place the world is; the complexity of grouping in most contemporary music only reflects the complexity of the world from which it arose.

I have tried to show very different examples of how composers deal with grouping in their music. Although there are certain basic techniques of creating phrases, grouping in music is never without expressive intent and content. As we continue our research into how grouping works - many of you by examining music, others of us by writing new music - it seems to me essential to listen with great attentiveness to the sense the phrases make.

Musician's Tone Glue

Johan Sundberg

The Problem

Listen to Sound Example 1. This way of performing a piece of music seems far from acceptable from a musical point of view. It is af if one could hear, with a painful evidence, how little the player cares about the piece he is playing. He ignores phrases, articulation and he treats important and unimportant tones alike. When one listens to such performances, the piece appears to consist of an endless, nagging sequence of tones that are completely independent of one another. The piece is moving on eight notes.

What is the problem, then? Our impression of how the performance sounded gives a hint. We seem to lack the marking of harmonic progressions, of phrases, of exciting melodic leaps. The player's *comments* on what is being played is all but missing, and this is a deficiency that irritates in the same way as it is irritating to listen to someone apparently void of compassion read a poem. The result is that the piece of music disintegrates into small fragments and sounds boring.

This is probably a notorious experience for music teachers. In our time the problem has been accentuated by the advent of playing machines, such as computers, sequencers, synthesizers etc. It is now easy to produce music that is played exactly as written in the notes. Such *verbatim* performances are truly boring, no matter how interesting the composition is.

This observation caused the violinist and music teacher Lars Frydén and myself to start a research project in the end of the 70:s. The construction of a singing machine, MUSSE (Music and Singing Synthesis Equipment) had just been successfully completed (Larsson, 1977). It could be played from a keyboard. The vowel timbre was almost flawless, but whatever happened musically in the composition, the machine invariably sounded completely blank. The lack of musical engagement turned out to be a musical disaster.

In his work as a music pedagogue, Lars Frydén was curious to find out if his instructions to students were to be understood literally or artistically. He also asked himself why it was necessary to do all these small tricks when playing? Was it a question of rules that should always be applied, as the teachers tend to pretend? Could a computer be programmed such that it could replace a musician?

The possibilities to attack this problem was very prosperous at that time. An interface was built, so that the singing synthesizer MUSSE could be played from a computer. Rolf Carlson and Björn Granström had constructed a computer program that converted written text in an input file to sounding speech (Carlson & Granström, 1975). They modified their text-to-speech conversion program to a note-to-tone conversion program, so that MUSSE could convert an input note file to singing. In their text-to-speech program there were good provisions for implementation of context-dependent pronunciation rules, and by listening to the result in terms of synthesized speech one could hear if the rules were realistic or not. Similarly, we could implement context-dependent performance rules in the music application. Thus, we could have an automatic, rule-controlled conversion of an input note file to a sounding performance, and we could try different rule formulations and listen to the effects it had on the performance.

This research paradigm is well-known within the natural sciences. It is called *analysis-by-synthesis*. One analyses the sound by synthesizing it. In this way it becomes possible to experience the meaning of different acoustic characteristics from an auditory point of view.

This equipment proved to be an almost perfect tool for analyzing music performance. As the sound quality could easily be adjusted so that it sounded almost exactly as a living singer or as a wind instrument, one expected realistic performances from the synthesis. It was possible to implement various context-dependent performance rules that introduced lengthenings, accents, crescendos etc in the performance, and then listen to the musical effect. During most of the time Lars Frydén and I have worked together with Anders Friberg, pianist and computer programmer. In this article I will present some rules that have emerged from this research (Sundberg, Friberg & Frydén, 1991a). Also some other research results from investigations of music performance will be reviewed.

Rule Categories

The complete set of our performance rules have been presented in detail elsewhere (Friberg. 1991). Table 1 presents an overview in which the rules have been grouped according to the purposes they appear to serve in

Table 1. Overview of the performance rules affecting duration (DR), sound level (L), fine tuning (F), vibrato amplitude (VA), and vibrato frequency (VF) or inserting micropauses at the end of notes (DRO). Some rules have two alternative (alt.) formulations.

RULE NUMBER	RULE NAME	SOUND PARAMETERS CONCERNED
	SEGMENTATION RULES	
1. Duration categories		
DDC 1A.	Durational Contrast	DR
DDC 1B.	Durational Contrast	L
DDC 2A.	Accents	L envelope
DDC 2B.	Double Duration	DR
2. Pitch categories		
DPC 1A.	High Sharp	F
DPC 1B.	High Loud	L
DPC 2A.	Melodic Charge	DR, L, VA
DPC 2B.	Melodic Intonation	F
	GROUPING RULES	
1. Micro level		
GMI 1A.	Leap Articulation	L envelope
GMI 1A'.	Same, alternative formulation L envelope	
GMI 1B.	Leap Tone Duration	DRO
GMI 1C.	Faster Uphill	DR
GMI 2.	Amplitude Smoothing	DR
GMI 3.	Inegalles	L envelope
GMI 4.	Articulation in Repetition L envelope	
GMI 4'.	Same, alternative formulation	DRO
2. Macro level		
GMA 1.	Phrase	DR, DRO
GMA 2A.	Harmonic Charge	DR, L, VF
GMA 2B.	Chromatic Charge*	DR, L
GMA 3.	Final Ritard	DR
	ENSEMBLE PERFORMANCE RULES	
ENS 1.	Mixed Intonation	F envelope
ENS 2.	Melodic Synchronization	DR
ENS 3.	Bar Synchronization	DR

themusic communication process. We have indentified three such purposes. One is to keep order within an ensemble, i. e. to decide when the tones shall start and end, and how intervals and chords should be tuned. These rules are called *ensemble rules.*

Another purpose seems to be to facilitate for the listener to correctly differentiate tones according to pitch into scale tone categories and according to duration into note value categories. By means of these *differentiation rules* the player helps the listener to realize if a tone is remarkable or trivial in its melodic context, if it is an eighth note or quarter note etc. The correct categorization seem important for the listener's understanding of which tones belong together.

A third group of rules, the *grouping rules*, seems to serve this last-mentioned purpose more directly, thus relating directly to the main topic of this article. These rules help the listener realizing which tones belong together and where the boundaries are in the musical structure.

Hierarchical Structure

Almost all music is built according to a *hierarchical principle*. This principle, illustrated in Figure 1, implies that sets of adjacent, small units, or *constituents* form intermediate constituents, which together with other intermediate constituents form greater constituents, which together with other greater constituents form still greater constituents, which together with other such constituents form *still* greater constituents, etc.

The hierarhical structure seems to be somewhat like a finger print or rather, perhaps, as a brain print on most creations of the human intellect. Governement is a typical example of a hierarchical structure, all the way from president or king and down to the cleaning personel. Classical architecture often offers other striking examples of hierarchical structures, starting with the entire house, via the center part and the wings, via roof, walls and souterrain, via windows doors and wall areas and down to the details in terms of bricks etc. Indeed the human mind seems to have difficulties handling structures that are not hierarchical.

Hierarchical structure can be found also in music. Figure 1 illustrates the very obvious hierarchical structure in a simple piece of music, that, typically enough, is written for children. This structure is representative for 16 bar nursery tunes in 4/4 time composed by the Swedish composer Alice Tegnér (Lindblom & Sundberg, 1970; Sundberg & Lindblom 1976). Mostly, the hierarchical structure is less symmetrical and perhaps a little less obvious in other types of music. However, the hierarchical principle is applicable to the structure of almost all types of music. Tree diagrams like that shown in Figure 1 offers a means to efficiently represent the structure (Lerdahl & Jackendoff, 1983).

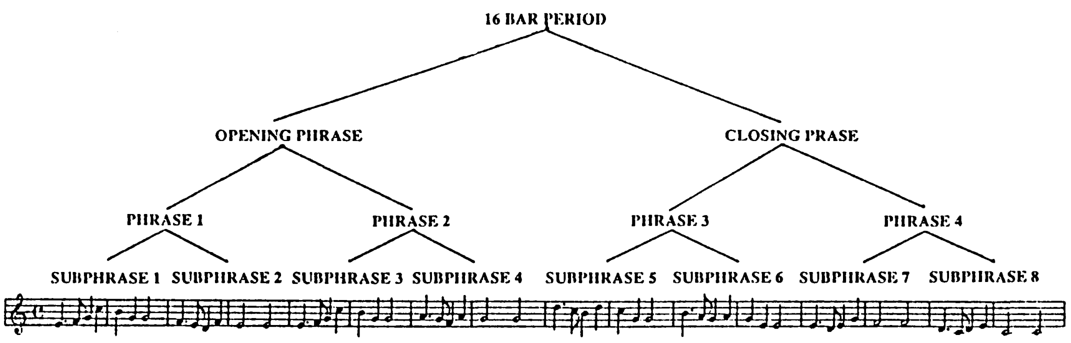


Figure 1. Illustration of a hierarchical structure in a nursery tune by the Swedish composer Alice Tegnér.

The hierarchical structure is the key to many aspects of music performance. As a reverence to this principle, let us start from the bottom and work our way upward.

Performance Rules

The smallest constituent in the musical structure is the single note. But notes belong together in different ways. If they appear as a scale, this scale is often a constituent. According to one of our performance rules, a tone initiating an ascending interval is shortened. In the lab jargon this rule is called *Faster Uphill*, because it increases the tempo in ascending scales. The effect of the rule can be listened to in Sound Example 2. First the music is played exactly as written in the score, without application of any performance rules. Then, the rule is applied, but to an exaggerated extent, so that the effects are easy to notice. Finally the rule is applied at a moderate quantity. Most listeners agree that this rule has the effect of gluing together the tones in ascending scales so that they form a gestalt. A grouping effect emerges.

If all tones in a piece is played as a never ceasing, invariably applied legato, the performance tends to sound boring. Thus, the minute micropauses that players generally insert in the tone flow are important. Another performance rule, *Leap Articulation*, inserts micropauses in melodic leaps. Physically speaking, it makes silent the last portion of all tones that initiate leaps, and the duration of this silence is proportional to the width of the interval as measured in semitones. The effect of this rule can be experienced by listening to Sound Example 3. In the first, deadpan version no pauses occur.

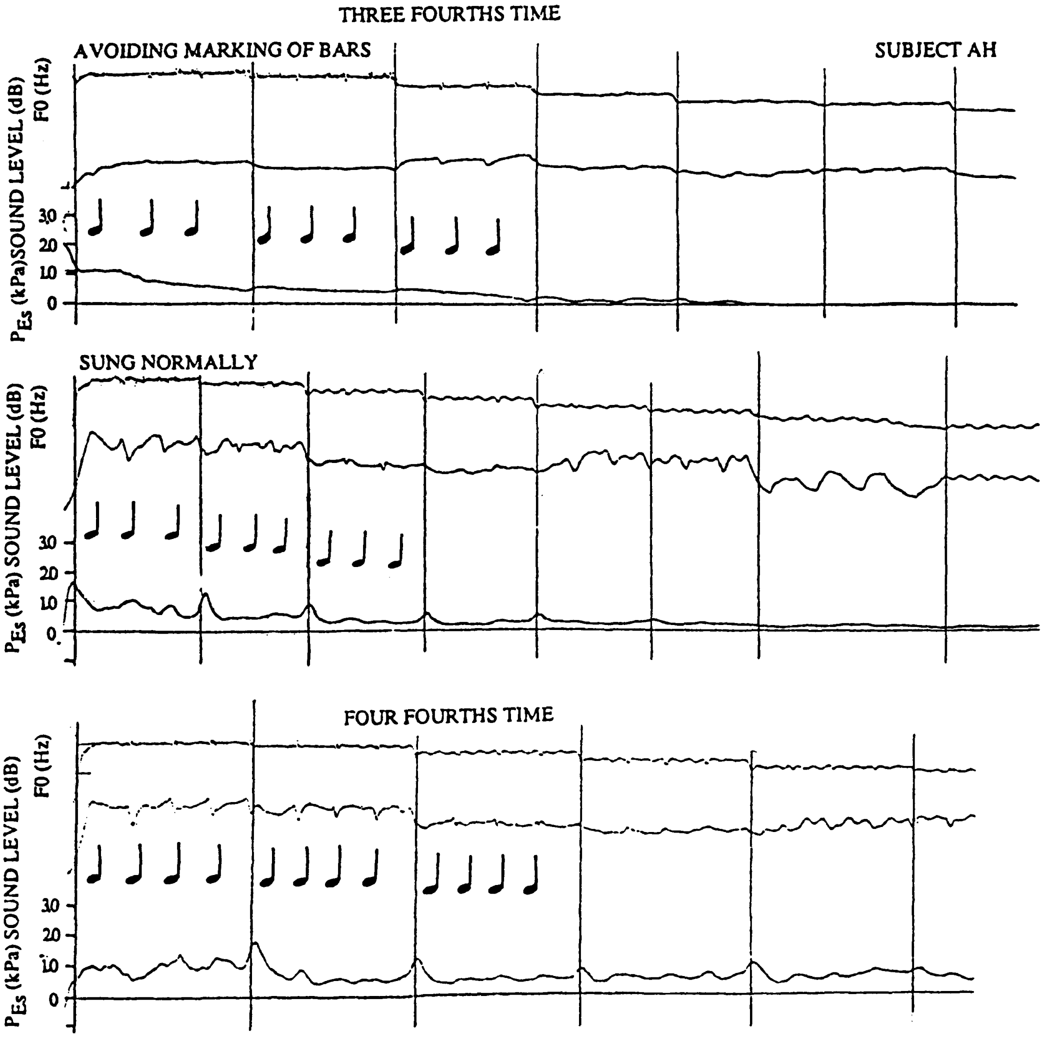


Figure 3. Relative lung pressure in a singer singing descending scales in 3/4 time (two upper curves) and 4/4 time (bottom curve) on the syllable [la], repeating each scale tone 3 and 4 times, i. e. troughout each bar. In the case shown at the top, the singer was asked to avoid the marking of bars.

there are no consonants that could help the listener to identify the boundaries between adjacent tones. Under these conditions singers often use the pitch gesture illustrated in Figure 2; in descending intervals the pitch change is produced with an overshoot, so that the fundamental frequency descends too far and then rises again up to the target. The effect of this is a *marcato*. As has been demonstrated in a synthesis example published previously (Sundberg 1978) this effect can be enhanced by small crescendo-diminuendo gestures on each note.

This method of marking the boundaries between tones seems logical. Great signal changes typically occur at tone boundaries, at least when the tone corresponds to a complete syllable. Then, there is often a consonant at the beginning of each tone. Consonants imply great changes in the acoustic characteristics of the tone produced, although apparently not great enough to necessarily introduce the impression of a structural boundary. In vocalise singing, such changes are lacking, so the singer exaggerates the only change left, i. e. the fundamental frequency change.

Marking of Bars in Singing

The rules discussed so far were discovered by means of the analysis-by-synthesis method, i. e. by listening to the computer controlled synthesizer's attempts to play music examples in a musically acceptable way. An advantage with the analysis-by-synthesis method is that one can tease out the function of single rules more easily. For example, there are many but quite different reasons to lengthen a tone; a lengthening may mean phrase ending, emphasis and perhaps other things too. If analysis-by-synthesis method is used, different cases of lengthening can be tried in terms of context dependent, tentative performance rules and the effect on the performance can be listened to.

A more commonly used method in music performance research has been analysis-by-measurements on real performances. Thus, one measures e. g. tone durations in the same piece played by different performers. One difficulty with this method is that it is difficult to identify the reason for e. g. a lengthening of a particular tone, as suggested above. Furthermore, different performers may lengthen different tones, because their interpretations differ. Therefore, durations averaged over several performers may be quite difficult to interpret. Nevertheless, measurements are often, of course, quite revealing.

Measurements can be carried out not only on the sound produced but also on the input parameters, i. e. on the musician's gestures. The latter is the case for the three examples shown in Figure 3. They show the changes of lung pressure in a singer singing descending scales and are quoted from a recent investigation (Sundberg, Gramming & Elliot 1991). Singers use this pressure to regulate vocal loudness; the higher the pressure, the louder the tone. In this case lung pressure was captured as the esophageal pressure which faithfully reflects lung pressure changes as long as lung volume is constant. Thus, for adjacent tones, lung pressure changes can be studied in terms of the changes of the esophageal pressure. The esophageal pressure can easily be measured if the subject swallows a thin catheter with a minute pressure transducer.

The figure shows the pressure events during performance of two types of scales, one in 3/4 time, the other in 4/4 time. In each case, each scale tone was repeated throughout the bar. Thus, in the 4/4 time scale, each scale tone was repeated 4 times in each bar and in 3/4 time each scale tone was repeated three times. This figure reveals how this singer marked the bar unit which was obviously a rather prominent constituent in this very simple musical structure.

The graphs show that this singer produced a short pressure accent on the first beat in each bar in both the 4/4 time and the 3/4 time scales. When he was asked to avoid marking the bars in the performance, he stopped producing such pressure accents, the pressure becoming very even, as can be observed on the third graph in the same figure.

The same experiment was carried out also with a female singer. She did not use the same strategy. Her method was instead to produce a very gentle pressure rise throughout each bar culminating on the first beat in the next bar. Thus she made a slight crescendo during each bar, starting on the second beat. In addition, she used articulatory means to mark the bar unit.

These examples indicate that the singer can use different means to mark the bar unit in the musical structure, whenever he/she finds this desirable. One way is to add a pressure accent to the first beat in each bar and another method is to produce a slight crescendo culminating on the first beat in the bars. In both these cases, however, the entering into a new structural unit was to produce a difference in loudness.

Phrasing

Let us now return to our interpretation rules developed by analysis-by-synthesis. The structural constituent closest to the bar is the phrase. In the hierarchical structure there seem to be at least two phrase levels, that can be called *subphrase* and *phrase*. Figure 1 offered an example, a 16-bar melody with subphrases and phrases marked. The structure is perfectly symmetrical; each phrase consists of two subphrases. This seems typical for music designed for young as mentioned, again, the structure belongs to a nursery tune.

According to the *phrasing* rule in our performance rule system, subphrases are terminated by a micropause while the last note of the phrase is lengthened by 40 ms. This means that the subsequent tone is started a little later than would have been the case otherwise. The effect can be listened to in Sound Example 4, which presents the same tune played three times, first without phrase markers, second with exaggerated phrase markers and third with more moderate phrase markers.

Why are the phrases marked in this particular way? The answer probably is that this is what we are accustomed to from speech. When we finish a sentence, we usually lengthen the last syllable, as mentioned. This is a well known phenomenon in phonetics where it is referred to as *final legthening*. It seems to occur in a majority of the languages in the world. Furthermore, we often insert a small pause at important structural boundaries in speech. Indeed, this looks like a smart idea: music listeners understand speech, so therefore let us use the same method of marking termination and boundaries in music as in speech!

In traditional Western tonal music harmonic progressions are frequently used to mark the structure. For instance, the progression dominant-tonic often is used to signal the end of a sdtructural constituent suchg as a phrase. These progressions seem important to music performance. According to one of the performance rules, *marking harmonic charge*, such progressions are marked with crescendos and diminuendos. In short, the rule states that crescendos are introduced culminating on dominants and diminuendos are used to announce the subsequent advent of the tonic chord. Actually, crescendos are used for warning the listener that a remarkable chord is about to arrive, and diminuendos signal the advent of unremarkable chords. In our performance rule system, the harmonic charge of a chord is a quanititative estimate of its remarkableness. This rule seems to serve the purpose of marking the structure; composers use chord progressions to mark the structure and the chord rule marks these chord progressions.

Crescendos and diminuendos also seem to serve another purpose in the marking of structure, namely to glue together the tones involved in the crescendo or the diminuendo. A crescendo implies a gradual increase of loudness that is evenly distributed over a string of adjacent tones. In other words, these tones are united by jointly forming a change of the overall loudness. In a way, one may regard this function of the crescendo/ diminuendo as iconic. The message from the notes to the listener seems to be something like "*We belong to the same constituent, so we all do the same thing*".

Unfortunately, there is no formal evidence supporting this assumption. The reader is, however, invited to experience the effect listening to Sound example 5, that presents synthesized singing. In the first version each individual tone in the tone sequence is marked by the *marcato* signs mentioned before, an overshoot in fundamental frequency in descending intervals and a crescendo-diminuendo pattern for each individual note. Musically experienced listeners tend to agree that the tones are not tied together in this case. Each of them sound in some way autonomous, somewhat as in a military march. In the second version, a crescendo-diminuendo pattern is introduced that covers one pair of bars. Listeners generally find that this reduces the impression of a sequence of autonomous tones.

Timbral Legato

Singers spend much time trying to develop a beautiful legato, i. e. a seemless rendering of a melodic sequence of tones. Ta achieve this goal it seems important that the singer can combine wide pitch leaps with an unchanged vocal timbre; the voice quality should be the same even for tones that are far remote from each other with respect to pitch.

Nothing of this sort is required in speech; on the contrary, voice timbre typically varies with pitch in speech. For example, the larynx normally moves up and down with pitch when we speak. The larynx position affects the voice timbre. This is because an elevation of the larynx shortens the vocal tract and therefore raises the formant frequencies, thus producing a "brigther" timbre and, conversely, a lowered larynx tends to make the vowel sound "darker" (Sundberg, 1987).

If the pitch-dependent larynx height were used also in singing, this would mean that tones differing considerably in pitch also would differ clearly in voice timbre. This would tend to break the legato line at wide pitch jumps.

Therefore it is important for singers to learn how to timbrally bridge wide pitch gaps by avoiding an automatic interdependence of pitch and voice timbre.

Not very surprisingly, similar seems to hold in instrumental music. In a classical sound example, Sound Example 6, previously published by Pierce (1983), the psychologist David Wessel demonstrated that the tones in a sequence could be glued together by giving adjacent tones similar timbres. If, on the other hand, adjacent tones differed in timbre, the melodic line cracked at several places. This effect seems to demonstrate how important it is for singers and other musicians to aquire a complete control over the timbre; if this goal is not achieved, the melodic lines will tend to crack in places where they are supposed to hold.

Quantity

In some of our sound examples, performance rules have been presented with three versions: nil, exaggeration and moderate. What, then is moderate? We have tried to find the answer to this question in two experiments. A detailed account of this research has been published elsewhere (Sundberg & al., 1991b).

For one experiment a special computer program was written that allowed variation of the magnitude of the effect induced by the application of a rule. The tool was a slider on the computer screen that could be operated by means of the computer mouse. In one extreme position the effect was huge, and in the other it was zero. Then we asked 5 professional musicians to adjust the slider for each of five rules such that, according to their musical judgement, the effect of the rule appeared appropriate. For all rules, it was possible to select zero as the preferred quantity. Figure 4 shows the average settings and the 95% confidence intervals obtained for the rules.

The averaged preferred quantities were mostly close to the default values that we had arrived at, when adjusting the complete performance program. More importantly, however, this result indicated that our musicians found that these rules improved the musical quality of the performance; the musicians did not find that a zero quantity of the rule generated effects was preferable.

In another experiment we attempted to determine the threshold quantities for the effects generated by these rules. Here we presented two versions of

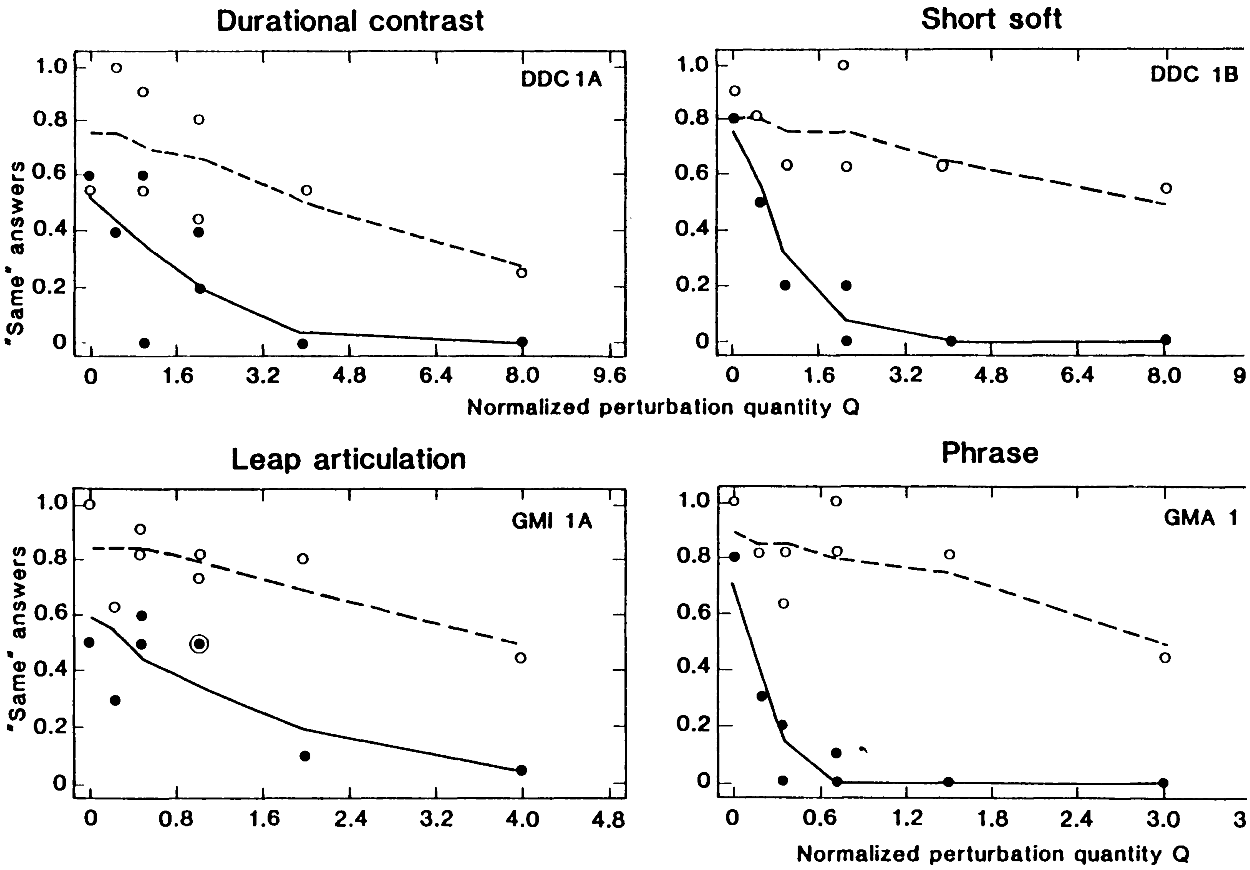
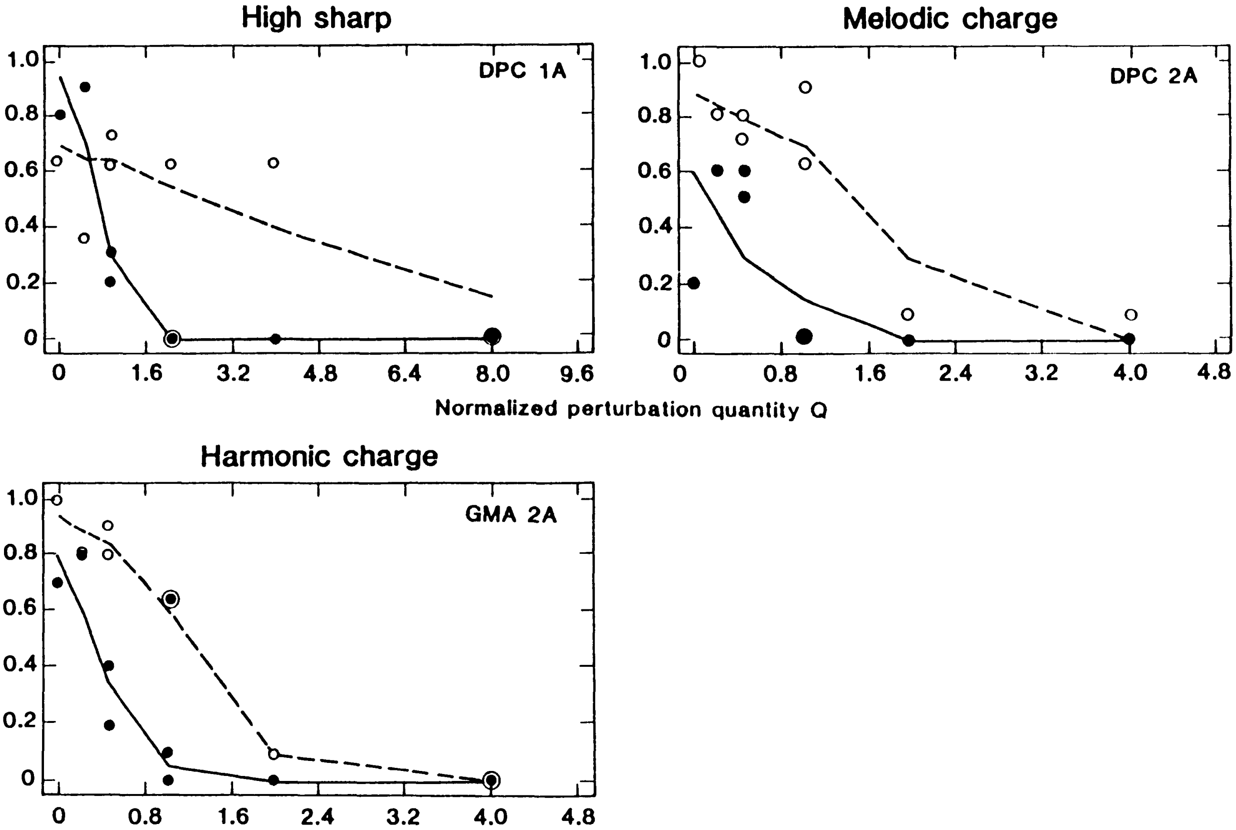


Figure 6. Threshold quantities for effects generated by the 7 performance rules indicated. The graphs show the percentage of "Same"-answers received in a test where musicians (filled circles) and nonmusicians (open circles) compared deadpan and rule-modified versions of the same examples. The abscissa represents the physical difference between the two versions, specified in terms of the quantity parameter Q, which was normalized with respect to the default value used in the program. The curves, derived by processing the data points with a LOGIT program (Bock, 1975), illustrate the overall trend of the data points.

she demonstrates this playing style. On the envelope she writes: "How did one mark the rhythm on organ and harpsichord? It was by stretching those tones on which the beats fall. How much? Well, this of course is a complex question. It depends on the Baroque tone characteristics."

Let us latch on to the last mentioned question regarding how much by



listening to Sound Example 7. It first presents a deadpan version an excerpt from a *Bourrée* from one of J S Bach's suites for solo cello. The computer here plays a sampler synthesizer, and the first time it plays the music deadpan, without any performance rules. The musical impression is the typical mechanical style of playing. Next, a rule distributes 80 ms lengthenings and shortenings to randomly selected tones. This sounds as beginner's playing. Next, the randomly distributed lengthenings and shortenings are reduced to 40 ms. The performance still sounds clearly inacceptable. In the following version the changes are no more than 20 ms. In this case most listeners tend to agree that the performance does not sound as totally mechanical and machine-like, but it is almost impossible to tell if this is because the tones are of different lengths or for some other reason. In the last version the random perturbations of the mechanical meter are no more than 10 ms. Here most listeners find it almost impossible to hear that the tones are not exactly equally long but many pretend they can still hear that there is some difference between this performance and a deadpan version.

This demonstrates a very important phenomenon. In motorical pieces of music, i. e. pieces with marked sequences of equally short notes, lengthenings of about 20 ms are sufficient for evoking an effect. Measurements have shown that lengthenings and shortenings of 20 or 30 ms are common in music performances (see e. g. Palmer, 1988). However, to produce changes that can be correctly recognized as shortenings or lengthenings, much greater effects seem to be needed, something more like 40 or 80 ms.

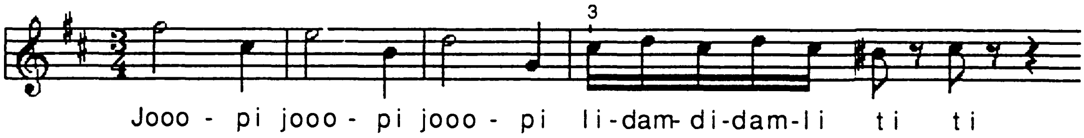
Sound Example 8 presents a sample of Lena Jacobson's organ playing. An anlysis of the magnitudes of her effects she used in the first bars of this piece revealed effects of some 40 to 70 ms, as one perhaps could hear.

It would not be incumbent on science to tell artists how to make art or to decide what is beautiful and what is ugly. On the other hand there is no problem for science to tell what the smallest noticeable difference is. It is quite clear that this organist uses effects that are much greater than the smallest perceptible effects. It seems that in this case, not only perceptible but also correctly analyzable effects are strived for: there is no doubt which tones were lengthened and which were shortened in this example. Had the only purpose been to glue tones together that belong together, much finer effects would have been sufficient.

Nonsense - Text - Singing

When musicians talk to each other about how they think a theme should be performed, they sometimes sing the theme using nonsense texts. The method is also sometimes used by great conductors instructing the orchestra. Let us imagine that a conductor wants to demonstrate how the *Waltz of the Flowers* from Pjotr Tchavkovski's *Nut Cracker Suite* should be played. Perhaps he would sing the following text:

It is obvious that these syllables are not meaningless. On the contrary, they mean specific things. One way of demonstrating this is to simply change the order of the syllables above and ask if this would be an equally suitable sequence of syllables for the melodic theme:



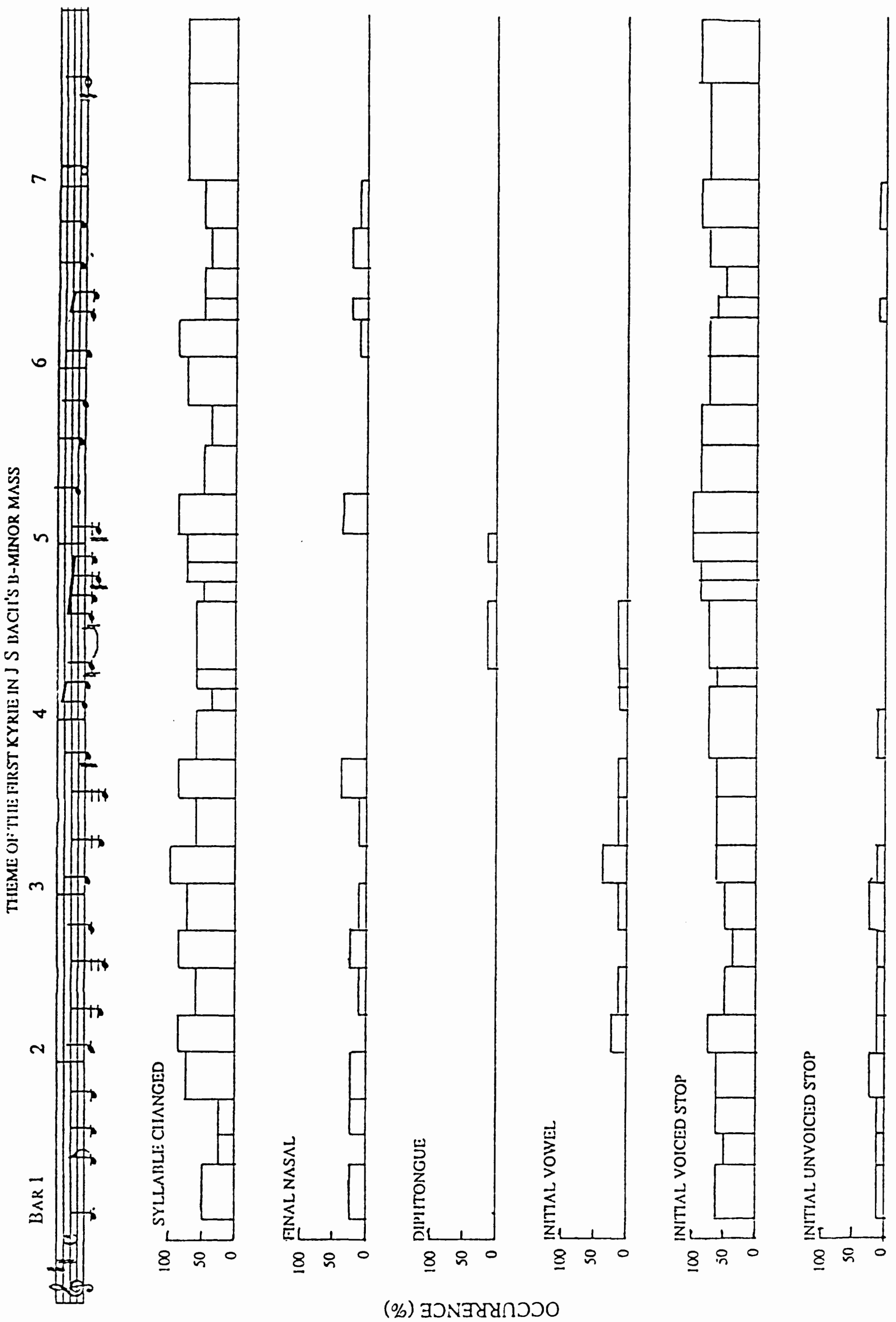
It seems doubtful that this instruction would be understood in the same way as the previous one. If these two nonsense texts are not equivalent, they are cannot be said to be nonsense.

What do they mean, then? Perhaps, one can see traces of performance rules in them? Before starting to look for an answer to these question, one thing should be clearly stated. The choice of syllables does not carry the entire information. The way in which the syllables are pronounced is another obviously extremely important factor, but here we will not analyze that aspect.

An experiment was carried out in which we asked eleven professional musicians to demonstrate how a set of 6 music excerpts should be played according to their view (Sundberg, 1992). We recorded their sung demonstrations on tape and examined their choice of syllables. We found some prinicples that appeared relevant to the choice.

One principle seemed to be to accompany a transition from one group of tones to another by a change of the syllable. In the tune "Santa Lucia" 9 out of 11 musicians changed syllable between the halfnote terminating the first subphrase and the following second subphrase, as can be seen in Figure 7.

Figure 7. Syllables, transcribed to phonetic symbols, used by the eleven subjects for the various tones in syllabling Excerpt Nr 1, Santa Lucia.


THEME OF THE FIRST KYRIE IN J S BACH'S B-MINOR MASS
BAR 1
2
3
4
5
6
7
SYLLABLE CHANGED
FINAL NASAL
DIPHTONGUE
INITIAL VOWEL
INITIAL VOICED STOP
INITIAL UNVOICED STOP
100
50
0
OCCURRENCE (%)

Similarly, very few musicians changed syllable from the first to the second eightnote in the pairs constituting the appogiaturas, a case of clearly contingent eightnotes. The nasal consonant [m] was often used to mark the end of short tone groups. Figure 8 shows where this consonant occurred in one of the excerpts, taken from a String Quartet by J Haydn. It can be noted that a final [m] that it did not appear on short tones, but frequently on tone repetitions. Furthermore, it did not appear between tones that belong strongly together such as appogiatura tones, i. e. when a non-chord note appears on stressed beat and is followed by a chordnote one scale step down.

There are certainly a number of other observations that can also be made from this material. However, for our present purpose the material presented so far seems sufficient. It shows that musicians in spontaneous singing of music themes use nonsense texts that are not at all meaningless. The choice of syllable is used to mark the grouping of tones. This result was not unexpected, but provides another striking example of the fact that marking of group boundaries is extremely important in music performance.

Discussion and Conclusions

This investigation has shown that grouping is marked in many different ways in music performance. One alternative is the rule "faster uphill", which shortens tones initiating an ascending interval, and thus move them closer in time. The rule increases the tempo in ascending scale or triad sequences. The rule "micropauses in leaps" promotes the association of tones that are close to each other in pitch. In cases where the notated bar is a unit of relevance to the musical structure, a singer may mark the first beat in the bars by giving them a small accent produced by a sudden increase-decrease of subglottal pressure. Phrase endings can be marked by lengthening of the final note, while subphrase endings are often marked by a micropause.

Figure 8. Occurrence of various phenomena in eight musicians' syllabling of Excerpt Nr 2, the theme of the first *Kyrie* from J S Bach's b-minor Mass: repetitions of syllable from one tone to the next, syllable with final nasal consonant, with diphtongue, with initial vowel, with initial voiced stop consonant and with initial unvoiced stop consonant.

This helpful marking of structural boundaries is not unique to music. The same happens in architecture, where boundaries between structural constituents are often marked. For instance, the boundaries between windows and walls are often marked by means of colours; window frames are generally painted in a colour different from the adjacent wall. Speech is another example; the importance of marking of structure is a well-known fact in speech science. It is considered a fundamental ingredient in what is generally referred to as prosody. Two examples were already mentioned: micropause for comma and final lengthening for period. Litterature is another example where structure is carefully marked. An extra space is inserted between words, a dot after each sentence, and each paragraph is started on a new line and each chapter is started on a new page.

It seems that the marking of structure is a commonplace phenomenon in human communication systems. Still, it may appear as slightly remarkable that it is considered so important, because even without such signs we tend to realize where the boundaries are. This, precisely, may be the reason why one gets so agonized when one hears a poor musician who reveals from his way of playing that he does not understand what structurally belongs together and what does not, or when we hear someone reading a poem aloud in such a way that it is obvious that the reader does not understand what he/she is reading. The reasons for such reactions is a question for psychology. It may be related to the strong emotional impact that music has on its listeners. For the moment we can conclude that gluing together tones that structurally belong together and marking the boundaries between tones that do not belong together seems to be an important part of musical performance.

References

Carlson, R & Granström, B (1975) "A phonetically oriented programming language for rule description of speech", in G. Fant ed.: *Speech Communication,* Stockholm: Almqvist & Wiksell, Vol 2, 245-253.

Clarke, E (1987) "Generative priwnciples in music performance", in J Sloboda (ed.): *Generative Processes in Music,* Oxford: Clarendon Press, 1-26

Friberg, A (1991) "Generative rules for music performance: A formal description of a rule system", *Computer music Journal 15,* 56-71

Kronman U & Sundberg J (1987) "Is the musical ritard an allusion to physical motion?" in A Gabrielsson (ed.): *Action and Perception in Rhythm and Music*, Stockholm: Publications issued by the Royal Swedish Academy of Music: Nr 55, 1987, 57-68.

Larsson, B (1977) "Music and Singing Synthesis Equipment (MUSSE)", Speech Transmission Laboratory *Quarterly Progress and Status Report* 1/ 1977, 38-40.

Lerdahl, F & Jackendoff, R (1983) *A Generative Theory of Tonal Music*, Cambridge, MA: The MIT Press.

Lindblom, B & Sundberg, J "Towards a generative theory of melody", *Swedish Journal of Musicology 52*, 1970, 71-88, B Lindblom & J Sundberg.

Sundberg, J & Lindblom, B (1976) "Generative theories in language and music descriptions", *Cognition 4*, 1976, 99-122.

Leon van Noorden (1975) *Temporal Coherence in the Perception of Tone Sequences*, Dissertation, Technical Inst, Eindoven

Palmer, C (1988) *Timing in Skilled Music Performance*, Dissertation, Cornell University.

Pierce J (1983) *The Science of Musical Sound,* New York: Scientific American Books (W H Freeman & Co).

Simon, H & Sumner, R (1968) "Pattern in music", in B Kleinmuntz (ed.), *Formal representation of human judgement,* New York: Wiley, 219-250.

Sundberg, J 1978 "Synthesis of singing", *Sw J Musicol 60:1*, 1978, 107-112.

Sundberg, J 1989 "Synthesis of singing by rule", in M Mathews & J Pierce (ed.): *Current Directions in Computer Music Research*, System Development Foundation Benchmark Series, Cambridge, MA: The MIT Press, 1989, 45-55 och 401-403.

Sundberg, J & Frydén, L (1992) "The sense of nonsense. An analysis of the syllable choice in spontaneous nonsense text singing", invited paper at the Second International Conference on Music Perception and Cognition, Los Angeles, Speech Transmission Laboratory 1/1992.

Sundberg, J, Friberg A, & Frydén, J (1991a) "Common secrets of musicians and listeners: An analysis-by-synthesis sturdy of musical performance", in P Howell, R West & I Cross, ed:s, *Representing Musical Structure*, London: Academic Press, 161-197.

Sundberg, J, Friberg A, & Frydén L (1991b) "Threshold and preference quantities of rules for music performance", *Music Perception* 9, 71-92.

Sundberg, J, Gramming, P & Elliot, N (1991) "Subglottal pressure and musical expression in singing", *Scandinavian Journal of Logopedics and Phoniatrics 16,* 43-49.

Sundberg, J (1987). *The Science of the Singing Voice* (author's translation of *Röstlära*), DeKalb (Illinois): N. Ill. Univ. Press.

Sundberg J & Verrillo V (1980) "On the anatomy of the ritard: A study of timing in music", *J Acoust Soc Amer 68*, 772-779.

Grouping processes in infants' music perception

Carol L. Krumhansl

A recent paper by Trehub (1990) began with the following intriguing observations:

"The study of musical pattern perception in infancy might be viewed as a rather esoteric enterprise, of interest perhaps to scholars of aesthetics, music, or other seemingly optional aspects of human endeavor. For reasons as yet unknown, however, music is no less universal than language, appearing in every known culture... There have been suggestions that music and linguistic capabilities use common organizing principles to impose structure on their disparate input and that these principles may be innately specified... Infants are remarkably precocious in (the) domain (of music), exhibiting a predisposition for global, relational... processing of musical patterns.." (p. 429)

This paper will expand on these themes, using research on auditory perception in infants to suggest that a set of basic principles of perceptual organization function in both music and language. This focus on possible commonalities between these two domains of human communication is not to ignore their many important differences. Each is a highly developed system, with complex rules governing the ways in which sound sequences are formed. However, at early stages of perception they share the same auditory mechanism and may be subject to the same general organizing principles. Ironically, these common principles may enable the listener to begin to acquire some of the distinctive features that characterize music and language. In other words, perceptual organization of auditory stimulation by young children may provide cues as to important domain-specific features, such as semantic and syntactic units, musical scales and harmonies.

Even though music and language both consist of auditory patterns extended more or less continuously over time, we experience discrete units which join in a hierarchical fashion to form larger units. In language, phonemes combine to form words, words to form clauses, and clauses to form sentences. In music, tones combine to form harmonies, melodic and rhythmic figures, and musical phrases. In experienced listeners, these perceptual products result from two different kinds of processes: bottom-up processes which are externally (stimulus) driven, and top-down processes which are internally (perceiver) driven. Bottom-up processes respond to acoustic attributes of the stimulus. As summarized by Bregman (1990), studies of auditory segregation indicate that a combination of heuristics are employed. Some of the features that tend to segregate the auditory stimulus into separate units, or "streams", are contrasts of pitch range, timbre, and loudness, and temporal gaps. In contrast, some factors that promote the joining together of the auditory stimulus into a single stream are similarity of pitch, timbre, and loudness, proximity in time, continuous transitions of pitch, and the formation of simple pitch contours.

In experienced listeners, top-down processes act together with the bottom-up processes. The perceptual input is interpreted in terms of knowledge of the rules governing the construction of well-formed sequences of sounds. For example, knowledge of the syntactic roles of words in language or tonal structure in music facilitate identifying the basic auditory units and apprehending how these units function in the larger context. These top-down processes can operate successfully even when, for example, the acoustic input is noisy, or a number of different speakers or instruments are sounding simultaneously. Thus, in the perception of experienced listeners, both bottom-up and top-down processes are engaged. For this reason, it is interesting to examine the perception of listeners with limited experience to discover processes of auditory organization that may operate prior to the acquisition of domain-specific knowledge.

My interest in this approach began a few years ago when I had an opportunity to collaborate with a developmental psycholinguist, Peter Jusczyk. He and his colleagues had been studying pre-verbal infants to see whether they are sensitive to prosodic cues which signal clause boundaries in language. As summarized by Hirsh-Pasek, Kemler Nelson, Jusczyk, Wright Cassidy, Druss, and Kennedy (1987), clause boundaries are often, although not necessarily, marked by such acoustic factors as: 1) longer pauses, 2) lengthening in the syllable preceding the clause boundary, 3) a fall or rise in fundamental frequency, with the former being more frequent, and 4) a stress marking. As I will discuss in more detail later, these cues to clause boundaries may be exaggerated in speech directed to infants.

Hirsh-Pasek et al based their stimuli on a recording of a mother speaking to her 19-month old daughter. From this recording, two kinds of samples were constructed to be used in the experiment. The "Natural" version began and ended at a sentence boundary and was transformed by inserting a 1-sec pause at all the clause boundaries. The "Unnatural" version began and ended in the middle of a clause and the 1-sec pauses were inserted between words in the middles of clauses.

I will describe Hirsh-Pasek et al's (1987) experimental procedure in some detail because we adapted it for our later studies with music. In addition, it illustrates the kind of methodologies that can be used to ask questions about auditory perception in infants. Their experiment used 16 infants, aged 7 - 10 months. During the experimental session, the infant was seated on its parent's lap in the center of a three-sided room. A flashing light on the center wall attracted the infant's attention in that direction. When a hidden observer saw that the infant was looking straight ahead, the flashing light was turned off. Then, lights mounted above loudspeakers on the two side panels began to flash. When the infant made a headturn of at least 30o in one direction, the speech sample began to play from that loudspeaker and continued until the infant looked away for at least 2 seconds. Thus, by their headturn, the infant was able to control the duration of the speech sample. The time the infant oriented in the direction of the loudspeaker playing the speech sample was recorded. For a given infant, the Natural versions were always played on one side, and the Unnatural versions on the other. This was counterbalanced across infants. The experiment began with 8 familiarization trials to acquaint infants with the procedure and the locations of the two different kinds of samples. Then, the experiment continued with 12 trials.

In their first experiment, Hirsh-Pasek et al (1987) found that infants exhibited a clear preference for the Natural samples over the Unnatural samples. The mean orientation time was 15.5 sec to the Natural samples, in contrast to 13.5 sec for the Unnatural samples. To give some indication of the stability of the difference, this pattern was found for 12 out of 16 infants and 11 out of the 12 test samples. Thus, the infants showed a preference for the Naturally segmented samples over the Unnaturally segmented samples suggesting they are sensitive to clause structure. However, this experiment contained a possible artifact. Naturally occurring gaps in the original speech sample would tend to occur at clause boundaries.

In fact, of the relatively long gaps (longer than 450 msec) in the original recording, 72.5% occurred at boundary locations. This meant that after the additional 1-sec pauses were inserted, the total number of gaps in the Unnatural versions was larger than in the Natural versions. Infants might have been responding to this difference, rather than to the clause structure *per se*. Consequently, a second experiment was run in which the recorded speech was first edited to remove any existing gaps longer than 450 msec. That is, all naturally occurring gaps were taken out before introducing the 1 sec pauses to create the Natural and Unnatural versions. This meant that the Natural and Unnatural versions had the same total number of gaps, and that the gaps were of the same duration. Again, infants oriented longer in the direction of the loudspeaker playing the Natural versions (8.1 sec) than in the direction of the loudspeaker playing the Unnatural versions (6.3 sec). Nineteen of the 24 infants showed this pattern, and it held for 11 of the 12 test samples.

Thus, young infants preferred the speech samples in which the pauses occurred at clause boundaries. This suggests that spoken clauses have a kind of perceptual unity for infants. This early sensitivity to clause structure was apparent even in Hirsh-Pasek et al's (1987) second experiment in which natural pauses were removed. Given that the preference remained, it appears that cues other than the naturally occurring gaps that tend to occur at clause boundaries, such as drops in the fundamental frequency and lengthening of final segments, can signal the clause structure to the infants. Alternatively, or in addition, the infants may respond negatively to the gaps in the middle of clauses in the Unnatural samples because they disrupt the prosodic cues that normally unify the phrase. Hirsh-Pasek et al argue that this early sensitivity to clause structure might provide a "perceptual scaffolding on which language-learning strategies can build". They note that many important grammatical rules apply to the level of clauses, and that the ability to segment the auditory stimulus into these units would facilitate the acquisition of these rules. Moreover, they note that adults speaking to infants may cooperate in this process. Infant-directed speech is characterized by exaggerated fundamental frequency contours and short utterances that are separated by long pauses. Indeed, in a subsequent study, Kemler Nelson, Hirsh-Pasek, Jusczyk and Wright-Cassidy (1989) found a preference for Natural over Unnatural segmentations in infant-directed speech, but not in adult-directed speech. In adult-directed speech, the speaker can rely on the listener's knowledge of syntax and semantics, and thus prosodic cues are less important.

Peter Jusczyk and I (Krumhansl & Jusczyk, 1990) wondered whether we might obtain the analogous result for music. If so, then it would suggest that music draws on some of the same processes of perceptual organization as language in young infants. For the musical stimuli, we chose opening excerpts from 16 simple minuets written by Mozart when he was a child (Mozart, Trent pieces faciles pour piano, Bruxelles: Schott Freres). These pieces were chosen because they are stylistically homogeneous, with quite obvious phrase structure. Each of the minuets contained an initial section of either 8 or 10 measures, followed by a repeat sign. The musical stimuli consisted of the initial section and its repetition, so that all stimuli were either 16 or 20 measures in length. The example shown in Figure 1 consists of an initial section of 8 measures, which is repeated to produce a 16-measure long segment. The phrase structure is 2+2+4+2+2+4 measures, as is indicated. In the Natural version we inserted a 2-beat pause (approximately 1 sec) at the end of each phrase, corresponding to the phrase structure of the music. In the Unnatural version, we kept the same pattern of pauses, only began and ended the music at some point other than the first measure, so that the pauses did not correspond to the phrase endings in the music. In the example shown, the Unnatural version began at measure 2, and continued for a total of 16 measures to match the length of the Natural version. The phrase structure was assigned intuitively by the experimenters, and varied from minuet to minuet. These intuitions were subsequently validated by a group of adult listeners.

The subjects in the first experiment were 24 infants of approximately 6 months of age. We followed the same procedure described earlier with one minor variation. As before, the Natural versions were always sounded from the loudspeaker on one side, and the Unnatural versions were always sounded from the loudspeaker on the other side (with the side counterbalanced across infants). However, on any given trial only one kind of sample (Natural or Unnatural) was available, which was signalled by flashing the light on the corresponding side only. An 8-trial familiarization period was followed by 12 test trials. During this main phase of the experiment, half of the infants heard the Natural and Unnatural versions of 6 of the minuets,

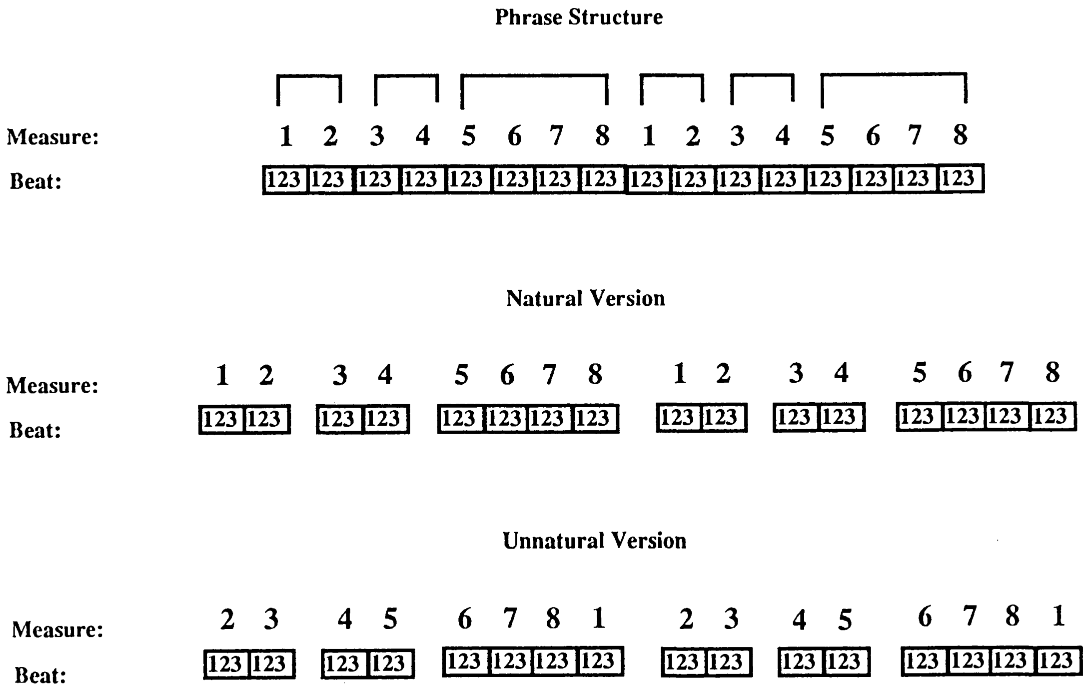


Figure 1. An example of the musical stimuli used in Krumhansl and Jusczyk (1990). This stimulus consists of the initial 8-measure section of a Mozart minuet, repeated twice to produce a total of 16 measures. The top diagram shows the natural phrase structure of the music, which consists of 2+2+4+2+2+4 measures. In the Natural version, 2-beat pauses are introduced between phrases, as shown in the center diagram. In the Unnatural version, the pattern of pauses is identical, as shown in the bottom diagram. However, the music begins at some place other than the first measure (in this example, it begins at the second measure), so that the pauses do not coincide with the phrase endings.

and the other half of the infants heard the two versions of the other 6 minuets. The ordering of the particular stimuli was random with the provision that for three of the minuets the Natural version occurred before the Unnatural version and vice versa. Preferences for the two versions was measured by the amount of time that the infant oriented to the loudspeaker playing each type of sample. The mean orientation times were 10.2 sec and 8.0 sec for the Natural and Unnatural versions, respectively. This difference was highly significant statistically. Twenty-two of the 24 infants showed this pattern, and it also held for 12 of the 12 minuets. Thus, the infants exhibited a strong preference for the musical stimuli in which 1-sec pauses were inserted at phrase endings as compared to the musical stimuli in which the 1-sec pauses were inserted in the middle of phrases.

To what aspects of the musical stimuli might the infants have been responding? Given that the stimuli were complex, naturalistic musical excerpts, any number of musical attributes might have influenced their preference. Although in the present kind of design, it is not possible to determine that some particular, isolated factor is having an influence, it is possible to examine the music for properties that correlate with the orientation-time differences. This approach provides suggestive evidence that certain musical patterns may serve as cues to the infants as to the location of the phrase endings. The music was analyzed for a variety of attributes, which were coded as quantitative variables, and these were correlated with the orientation-time data. Of the various musical attributes examined, three clearly differed between the Natural and Unnatural versions and correlated significantly with the orientation times.

The first attribute was the pitch of the tones surrounding the pauses. The top of Figure 2 shows the average pitch of the last three tones before the pauses, and the first two tones after the pauses. The solid line shows the average values for the Natural versions; the dashed line shows the average values for the Unnatural versions. The values on the vertical axis correspond to pitch height, where C4 refers to middle C on the piano keyboard. As can be seen, there was a sharp drop in pitch height in the last two tones before the pauses in the Natural versions but not in the Unnatural versions. The pitch height of the last two tones before the pauses correlated significantly with the orientation times, suggesting that one of the cues that signals the locations of the phrase endings is a drop in pitch height.

The second variable that correlated with the orientation times was the durations of tones surrounding the pauses. The durations in the melody showed more variability than the bass notes, which tended toward a uniform rhythm. Consequently, only the durations of the tones in the melody were considered in the analysis. The bottom graph in Figure 2 shows the durations of the last three notes before the pauses in the Natural and in the Unnatural versions. As can be seen, the last two notes before the

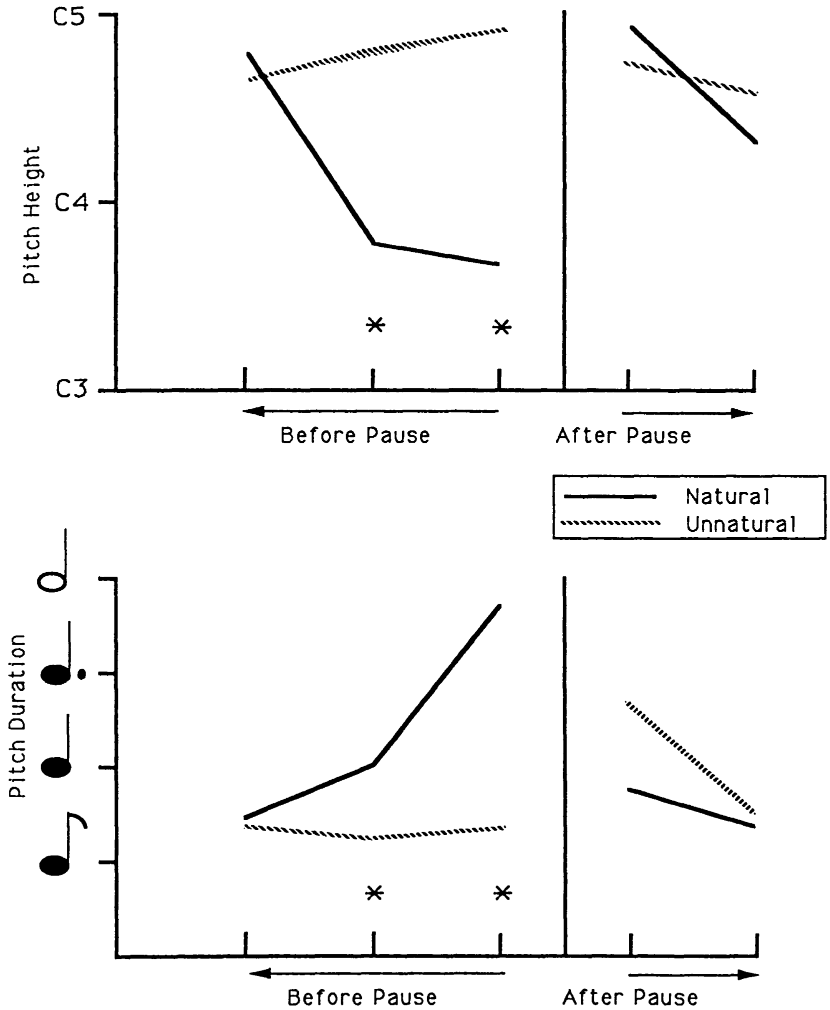


Figure 2. Shows two of the quantitative variables that correlated significantly with the orientation-time data of Krumhansl and Jusczyk (1990). The top graph shows the average pitch height (with respect to C4 = middle C) of the last three tones before the pauses and the first two tones after the pauses in the Natural vérsions (solid line) and the Unnatural versions (dashed line). The bottom graph shows the average durations (with a quarter note = .48 sec) of the same tones in the Natural and Unnatural versions. The percent of final intervals that are octaves also correlated with the orientation-time data. (After Krumhansl and Jusczyk, 1990, with permission of the American Psychological Society.)

pauses increased in duration in the Natural version relative to the same two notes in the Unnatural version. The durations in the last two positions correlated significantly with the orientation times, suggesting that the slowing down of the tones in the melody line might have, in part, signalled the phrase endings. As this variable covaries with the drops in pitch height (shown in the top of Figure 2), this method of analysis cannot be used to differentiate between the two possible cues in their contribution to the preference of the infants to the Natural versions.

The third variable which also correlated with the orientation-time data was the identity of the interval before the pauses, that is, the interval that is formed by the last pitch of the melody and its accompanying bass note. This last harmonic interval before the pauses was frequently an octave (58% of the time) in the Natural versions, but was less frequently an octave (11% of the time) in the Unnatural versions. A quantitative variable was created which was the proportion of intervals preceding the pauses that were octaves. This variable correlated significantly with the orientation-time data. Again, however, octave intervals tended to cooccur with drops in pitch and longer durations of the tones in the melody, so it is difficult to assess the relative contributions of the different musical factors. However, it is possible to find parametric studies in the literature which study these factors in isolation, and these substantiate infants' sensitivity to these musical properties. I will summarize these and other related studies below.

Before turning to that summary, however, let me mention we (Krumhansl & Jusczyk, 1990) ran a second experiment which was a replication of the study I just described, but with younger infants. In the first experiment, the listeners averaged approximately 6 months of age; in the second, they averaged approximately 4 1/2 months of age. Again, 24 listeners were employed and, again, a reliable preference was found for the Natural versions over the Unnatural versions. The mean orientation time was 12.8 sec for the Natural versions, and 10.1 sec for the Unnatural versions. Twenty of the 24 infants showed this pattern, and it held for 11 of the 12 minuets. The correlations with the musical variables were similar, although somewhat less regular. The orientation times correlated significantly with the pitch height of the last tones before the pauses, but not the pitch height of the second to last tones. The orientation times correlated significantly with the durations of the last melody tones before the pauses, but not the durations of the second to last tones. Finally, the proportion of octaves did not correlate significantly with the orientation times, although the correlation was in the same direction and approached significance. In general, however, this second experiment provided additional evidence for infants' sensitivity to phrase structure in music, and provided additional support that these three musical variables may signal the phrase structure to the infants.

Trehub and her collaborators have been engaged in a highly productive program of research on infants' auditory capabilities (see Trehub, 1990, for a recent review). This research has investigated the structural properties of auditory patterns to which infants are sensitive, finding that, in many ways, infants tend to organize auditory patterns similarly to adults. Here, I will summarize only a few of the studies from that program of research, selecting those that have a direct relationship to the study of musical phrase perception just described. These parametric studies are a useful complement to studies using more complex, naturalistic materials, as they serve to demonstrate the effects of isolated variables which may also be involved in the perception of more complex auditory stimuli, such as speech and music.

Before turning to the specific results, a brief description of the experimental method is needed as it differs in certain respects from the orientation-time method described earlier. The method used by Trehub and collaborators is called a conditioned headturn procedure, and it can be used with infants as young as 6 months of age. Basically, infants are trained to turn toward a loudspeaker when they hear a change in an ongoing auditory pattern. The infant sits on the parent's lap with a loudspeaker at 45o to one side. The loudspeaker is continuously playing a repeating melody and, initially, the infant will tend to look toward it. Soon, however, the infant will redirect attention to the tester, who manipulates puppets in front of the infant in order to attract the infant's attention in the forward direction. At random points in time when the infant is looking toward the tester, a trial is initiated which means that a change is made to the pattern that has been repeating continuously. If the change is noticed, infants tend to turn their heads toward the loudspeaker. Such headturns are reinforced by illuminating and activating animated toys near the loudspeaker. If the infant turns when no change has been made in the auditory pattern, there is no reinforcement. An initial training phase informs infants as to these contingencies using patterns with very obvious changes. Then, the experimental trials begin in which the pattern is changed according to the variable being tested in the experiment. The number of times that the change evoked a head turn is recorded, and compared to the number of times a head turn occurs in a similar period of time when the pattern is unchanged.

In one study, Trehub, Thorpe, & Morrongiello (1987) showed that, not only are infants sensitive to melodic contour (the pattern of increasing and decreasing pitch), but that they can classify melodies on the basis of contour. That is, infants can discriminate between *sets* of melodies that have different contours. In one condition, the background stimulus consisted of a 5-tone melody that was repeated over and over, only transposed to different keys. This melody kept the same contour and intervals even though it was shifted on every repetition to a new starting tone. Infants learned to treat this as an unchanging pattern and to respond only when a new melody was introduced with a changed contour. In a second condition, the background stimulus consisted of a 5-tone melody that was transformed, not only by shifting it up and down in pitch range, but by altering the intervals also. They were altered, however, in such a way as to keep the contour pattern the same. And, again, infants learned to treat this as an unchanging pattern and to respond only when a pattern with a new contour was introduced. That performance was approximately equal in these two conditions suggests that contour, rather than precise interval size, is encoded by infants. Dowling (1978, 1988) has emphasized the importance of melodic contour for both adults and for children.

Trehub and Thorpe (1989) obtained an analogous result for rhythms. The background pattern was either a 2-1 (XX X) or a 1-2 (X XX) pattern that was repeated over and over, but at varying tempos. Infants learned to treat this as an unchanging pattern and to respond only when the other, contrasting, rhythm was introduced. Infants were also found to be able to discriminate between 2-2 (XX XX) and 3-1 (XXX X) rhythms despite the varying tempos. These findings argue that infants are able to classify on the basis of rhythmic patterns even across variations in tempo. Thus, both temporal and pitch patterns appear to be perceived by infant listeners as wholes. These wholes retain their identities despite variations of tempo and pitch range, respectively. The relative values of the successive tones within the groups are perceived as unchanging, keeping a fixed rhythm and a fixed contour on the basis of which the patterns can be classified.

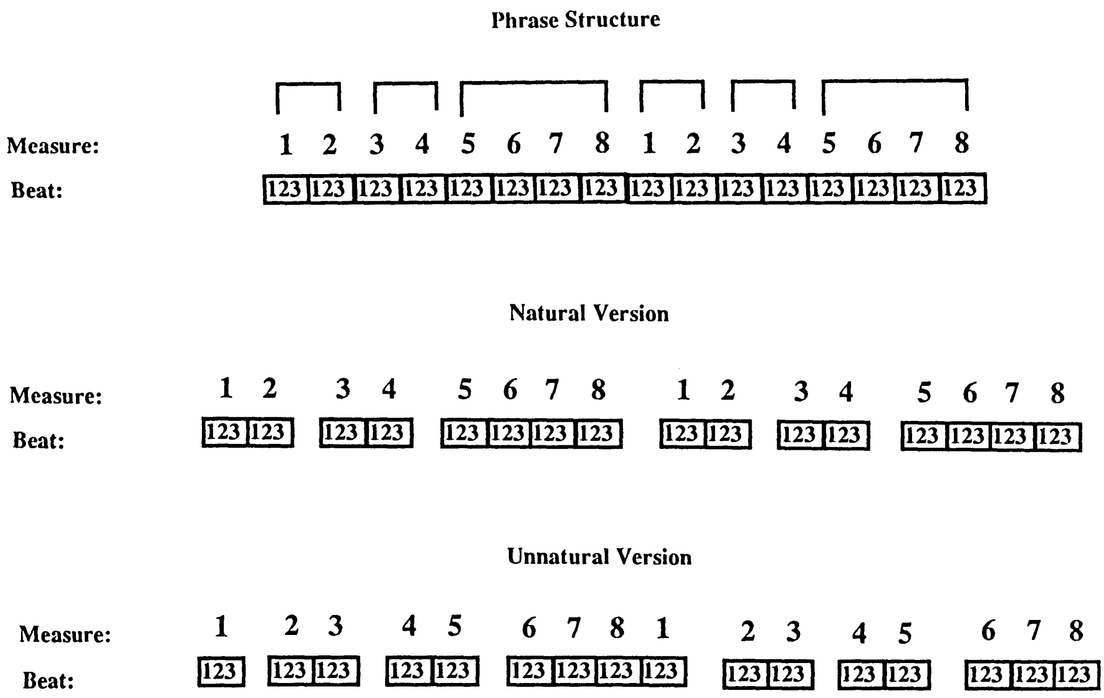


Figure 3. An example of the musical stimuli used in Experiment 1 of Jusczyk and Krumhansl (1991). In the Natural version, 2-beat pauses are introduced between phrases, as shown in the center diagram. In the Unnatural version, the pattern of pauses is shifted (in this example, by one measure to the right) so that the pauses do not coincide with the phrase endings, as shown in the bottom diagram. Unlike the stimuli of Krumhansl and Jusczyk (1990), both Natural and Unnatural versions begin at the beginning of the music and end at the ending of the first main section (after it is repeated).

compared to .66 beats or .32 sec for the Unnatural versions). And, final intervals were more frequently octaves in the Natural versions than in the Unnatural versions (58% and 11%, respectively). Each of these coded variables correlated significantly with the orientation time data, supporting the idea that these three variables help signal the phrase structure to the infants. The main finding of interest, however, was that even though both kinds of stimuli had the same beginnings and endings, the preference for the Natural versions over the Unnatural versions was just as strong, if not stronger, than in the previous study.

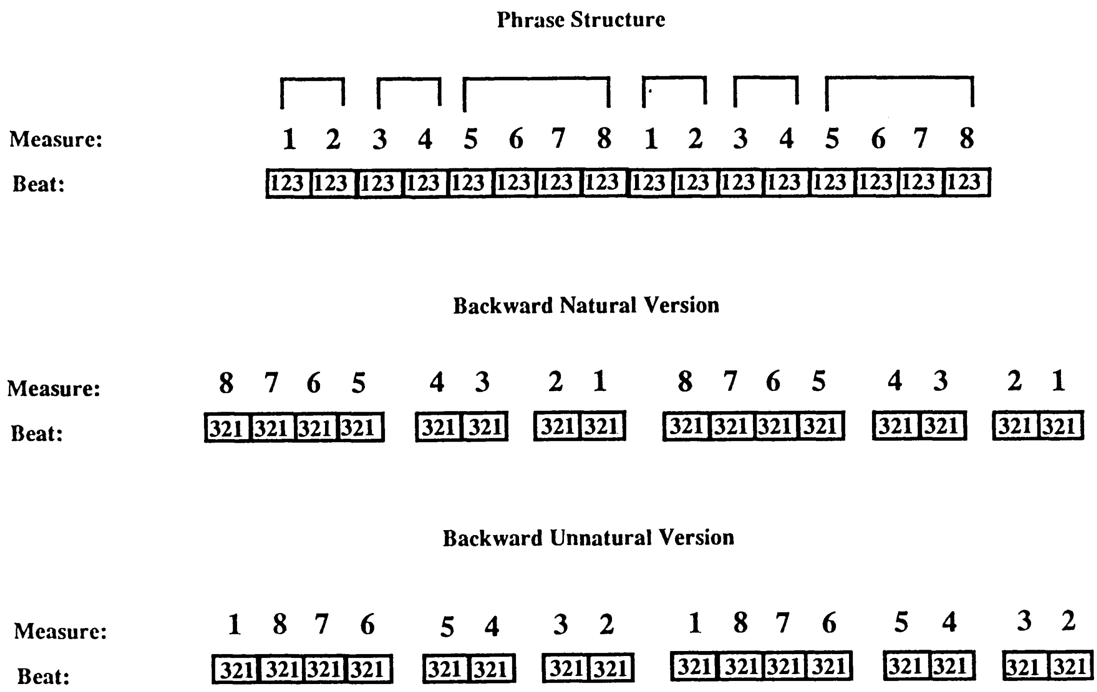


Figure 4. An example of the musical stimuli used in Experiment 2 of Jusczyk and Krumhansl (1991). The Backward Natural version is the Natural version used in Krumhansl and Jusczyk (1990, see Figure 1, center diagram) with the order of all tones reversed. The Backward Unnatural version is the Unnatural version used in Krumhansl and Jusczyk (1990, see Figure 1, bottom diagram) with the order of all tones reversed.

The second experiment took the Natural and Unnatural versions used in the earlier study (Krumhansl & Jusczyk, 1990) and played them backwards. This transformation reversed the order of all the notes of the original music; an example is shown in Figure 4. Two consequences of this transformation are of particular interest. First, the pitch height of the tones before the pauses were approximately equal in the Backward Natural and the Backward Unnatural versions (if C4 is coded 60, the average of the last tones before the pauses in the Backward Natural versions was 69.2 or approximately A4, and the average of the last tones before the pauses in the Backward Unnatural versions was 71.5 or approximately C5). The second consequence of the transformation was that the durations of the tones before the pauses were equal in the Backward Natural and the Backward Unnatural versions.

This is because, in the original music, the first tones after the pauses were always on the first beat of the measure so that, when the music was reversed, the last tones before the pauses were always on the last beat of the measure with a duration of 1 beat (.48 sec).

Thus, reversing the original stimuli had the consequence of equating Natural and Unnatural versions in terms of both the pitch height and the durations of the last tones before the pauses. If these factors are indeed important cues for phrase structure, as we have argued, then the preference for Natural over Unnatural versions should disappear when they are played backward. This was the result obtained. In this study, the mean orientation time was 8.3 sec for the Backward Natural versions and 9.5 sec for the Backward Unnatural versions. This difference is not significant, and tended in the opposite direction from the previous experiments. So, the results support the idea that in the other experiments, the drops in pitch and the longer durations of the melody tones before the pauses functioned as cues to phrase boundaries, because when these variables were equated no preference remained. In the transformed stimuli of this experiment, somewhat more octave intervals occurred before the pauses in the Backward Natural versions (59%) than in the Backward Unnatural versions (32%). This variable correlated negatively with the orientation times, opposite to the effect in previous studies. Apparently, this is a relatively weak cue and may not operate in isolation from the other phrase-signalling cues. I will return to the question of what is known about infants' perception of musical intervals shortly.

The last experiment pitted the Forward Natural versions against the Backward Natural versions; an example is shown in Figure 5. Here, there was a strong preference for the Forward Natural version over the Backward Natural versions. The mean orientation times were 11.7 sec and 8.4 sec, respectively. Eighteen out of the 24 infants preferred the Forward Natural versions to the Backward Natural versions, and this preference was found for 12 out of the 12 minuets. Although these two types of stimuli differ in many ways, let us focus on the tones immediately preceding the inserted pauses. The last tones before the pauses were lower on average in the Forward Natural versions than in the Backward Natural versions (if C4 is coded 60, the average of the last tones before the pauses in the Forward Natural versions was 56.1 or approximately G#3, whereas the average of the

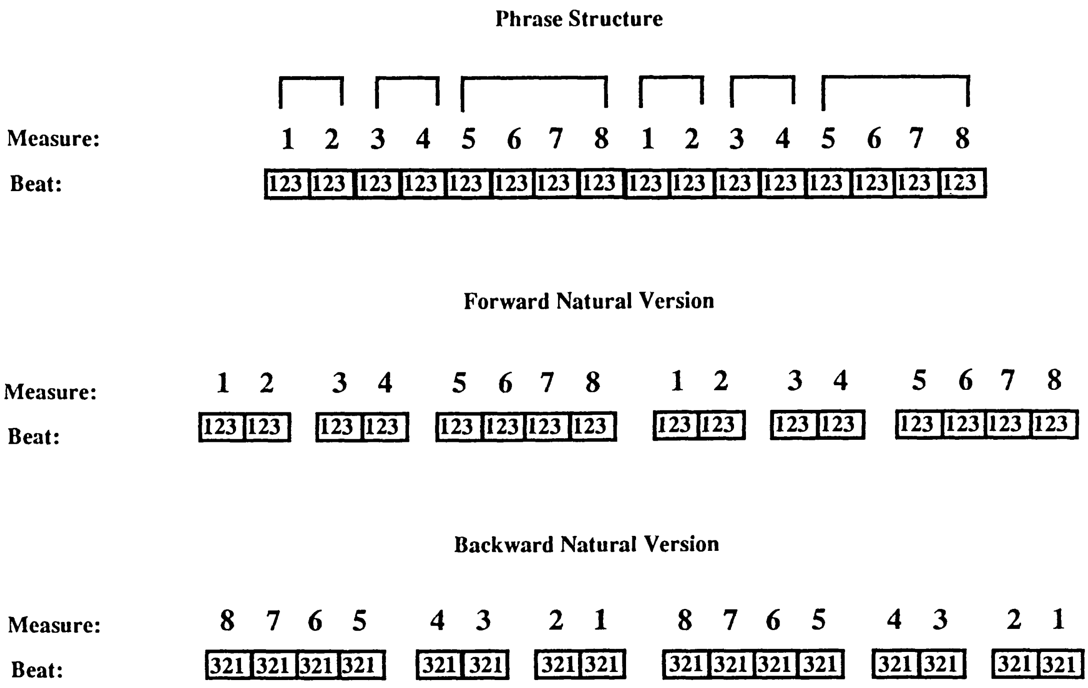


Figure 5. An example of the musical stimuli used in Experiment 3 of Jusczyk and Krumhansl (1991). The Forward Natural version is the same as the Natural version used in Krumhansl and Jusczyk (1990, see Figure 1, center diagram). The Backward Natural version reversed the order of all tones in the Natural version and is the same as the Backward Natural version of Experiment 2 of Jusczyk and Krumhansl (1991, see Figure 4, center diagram).

last tones before the pauses in the Backward Natural versions was 69.2 or approximately A4). This variable correlated significantly with the orientation times. The duration of the last melody tones before the pauses was longer in the Forward Natural versions than in the Backward Natural versions (1.83 beats or .88 sec for the Natural versions as compared to a constant 1.0 beat or .48 sec for the Backward Natural versions). This variable also correlated significantly with the orientation times. The percentage of intervals before the pauses that were octaves was quite similar in Forward Natural versions (64%) and the Backward Natural versions (58%), and this variable did not distinguish between the two versions nor did it correlate significantly with orientation times.

To summarize, we (Jusczyk & Krumhansl, 1991) found in the first of these experiments that infants preferred the Naturally segmented versions to the Unnaturally segmented versions, even when the stimuli all had identical beginnings and endings. This suggests that it is the locations of the interior pauses that are inducing the preference for the Natural versions, and not some characteristics of the beginnings and endings. In the second experiment, the preference for the Naturally segmented versions over the Unnaturally segmented versions disappeared when the music was played backward. The transformation of reversing the order of the tones essentially equates the pitch height of the final tones before the pauses in the two versions, and also equates their durations. In the third experiment, we found that there was a strong preference for the Forward Natural over the Backward Natural versions, and again the pitch height of the tones before the pauses and the durations of the melody tones before the pauses correlated significantly with the orientation times. Thus, these three experiments together with the previous two experiments (Krumhansl & Jusczyk, 1990) provide strong support for the idea that two musical properties are used by infants to segment the music into phrases: drops in pitch contour, and longer durations. The third property, the proportion of octave intervals before the pauses, had inconsistent effects. In two of the five experiments, there was a statistically significant relationship, with longer orientation times for stimuli with more final octaves; in another experiment, the relationship was in the same direction, but was not significant; in another experiment, it was significant, but in the opposite direction; and in the last experiment, the two version tested did not differ in terms of this variable.

That young infants showed less consistent effects of octaves than the other two variables might be understood by the following kind of argument. It seems plausible that infants would have an innate predisposition to organize auditory events according to certain basic principles, such as grouping on the basis of proximity in pitch and time. These principles would have a wide range of application to both music and language, and possibly to other sources of sound. In contrast, specific pitch relationships, such as the octave, appear to be domain-specific. Indeed, marked differences can be found across cultures in the musical intervals employed. Tuning systems and scale structures take different forms cross-culturally, suggesting these are cultural artifacts that need to be learned. In this light, it makes sense that sensitivity to musical intervals may appear later in the developmental sequence than sensitivity to melodic contour and rhythm. A recent study by Lynch, Eilers, Oller, and Urbano (1990) supports this point of view. They tested to see whether listeners are sensitive to a mistuned tone in three different kinds of contexts: a Western major scale context, a Western minor scale context, and a Javanese pelog scale context which has markedly different tunings than the other two. They found that infants of 6 months of age were equally sensitive to the mistuning in all three contexts. In contrast, adult listeners familiar with Western music performed better in the Western major and minor contexts than in the Javanese pelog scale context. This result can be understood if the adult listeners have learned the intervals normally used in Western music and can use this knowledge to help them detect the mistuned tone. The same kind of knowledge would not be available for the unfamiliar tuning system, resulting in poorer performance. In contrast, young infants presumably do not have enough experience with music to have developed this kind of knowledge, and thus they are neutral with respect to the different contexts used in the experiment.

However, the octave may be a special interval. With only one known exception (Ellis, 1965), this interval has been found in every musical culture that has been studied. Moreover, it may have a strong basis in musical acoustics, as it is the interval formed by the fundamental frequency and the first overtone of harmonically rich tones. In addition, Demany and Armand (1984) found some evidence that young infants treat tones separated by octaves as equivalent. In their experiment, infants listened to a repeating melody. After the infants had habituated to the melody, the tones of the melody were randomly shifted to the same notes in different octaves. Some tones were shifted up, others were shifted down. Infants did not dishabituate to the octave-shifted tones. It was as though they treated them as equivalent, in some sense, to the corresponding tones in the original melody. Given this, it is not surprising that the final octave interval is having an effect in some of our studies (Krumhansl & Jusczyk, 1990; Jusczyk & Krumhansl, 1991), although the complex nature of our stimulus materials may override its influence in others.

In a couple of recent studies, Trehub and collaborators have found evidence that infants may be sensitive to other aspects of tonal structure. In one study (Trehub, Thorpe & Trainor, in press), infants of approximately 8 1/2 months of age heard three different background melodies: a "good Western" melody based on a dominant-tonic progression, a "bad Western" melody based on tones of the chromatic scale but not consistent with any diatonic scale, and a "non-Western" melody based on a different tuning system. A one-semitone change was introduced after the infant had habituated to the background melody. The one-semitone change was noticed only in the context of the "good Western" melody. Another study (Cohen, Thorpe, & Trehub, 1987) showed that infants can discriminate between major, minor, and augmented triads. Interestingly, it was much easier for them to notice the change from a major triad to an augmented triad than the change from an augmented triad to a major triad. Similar asymmetries have been found in studies with adults (e.g., Bharucha & Krumhansl, 1983), which show that changes to less expected harmonies are noticed more easily than changes to more expected harmonies. However, Trainor & Trehub (in press) found changes to a nondiatonic tone and changes to a diatonic tone were equally easy for infants to detect, whereas the former were considerably easier for adults to detect. It will be interesting to follow this program of research as it develops. At present, a fair summary would seem to be that basic grouping principles appear to produce much more reliable and robust effects in experiments on infants' perception than tonal structure, suggesting that the acquisition of tonal structure requires more extensive experience.

The question that I would like to turn to now is how this acquisition might occur. What are the cues in music that might enable young children to begin to acquire a sense of its culture's tonal system? Elsewhere (Krumhansl, 1990), I have noted that in Western music, the tones that are considered most structurally significant from a music-theoretic point of view tend to be sounded more frequently and for longer durations. The tonic and dominant appear most frequently, followed by the other scale pitches, and nondiatonic tones appear least frequently. Perhaps sensitivity to these distributional properties may aid in acquiring knowledge of Western scale structure and harmony. Perhaps, in addition, some of the grouping processes revealed in the infant perception studies reviewed here may also play a role. Suppose, for example, that infants are selectively attentive to tones in prominent positions, such as those at the ends of phrases. These tones might form the core of the tonal system that is learned. In support of this, an analysis of the Mozart minuets used in the Krumhansl and Jusczyk (1990) study showed that 69% of the tones preceding the phrase endings were either the tonic or the dominant of the key, and none were nondiatonic tones. Thus, attention directed to the final tones would tend to select the most structurally significant tones in the Western tonal system.

This kind of regularity brings to mind a recent study of infant-directed speech by Fernald and Mazzie (1991). They asked mothers to tell a story using a picture book in which 6 target items were the focus of attention. The story was told to a 14-month old infant and to an adult. The fundamental frequency contours were much more exaggerated in the infant-directed speech than in the adult-directed speech. The infant-directed speech was marked by wide pitch excursions and smooth, simple contours. Moreover, Fernald and Mazzie found that the mothers speaking to the infant quite consistently gave prosodic emphasis to the target words. Approximately 75% of the target words appeared in the utterance-final position, and 60% not only appeared in the utterance-final position but also at a peak in the fundamental frequency. This prosodic emphasis, they argued, might serve two functions. First, it might serve the general purpose of attracting and maintaining the infant's attention. Exaggerated pitch contours have been shown to be effective in this way (Fernald, 1985; Fernald & Kuhl, 1987). Second, the prosodic variations might serve to direct attention specifically to the lexical items of central interest. As such, it may play a role in the acquisition of language structure.

However, long before the infant is ready to begin to acquire the semantics and syntax of its language, infant-directed speech has many of these exaggerated prosodic characteristics. Adults speaking to infants use higher pitch, shorter utterances, and longer pauses. Adults also speak more slowly, and the vowels preceding voiced final consonants are fully twice as long in infant-directed speech as in adult-directed speech (Fernald & Mazzie, 1991). Why do adults modify their speech to very young infants in these ways? Fernald (in press) has argued that even before the language-learning process begins, these characteristics serve a function, tapping into a non-arbitrary, biologically specified, and possibly universal system of communication. In support of this, Fernald (in press) showed analyses of speech to infants in a number of different languages in four different communicative situations: attention bids, approvals, prohibitions, and comfort vocalizations. Both attention bids and approvals were characterized by high average fundamental frequency and wide contours. Approvals were consistently characterized by a rise-fall contour; the contour shape was more variable in attention bids. In contrast, prohibitions and comfort vocalizations had relatively low average fundamental frequencies and narrow pitch contours. These categories tended to differ in their temporal patterns, with prohibitions marked by shorter, more abrupt utterances than comfort vocalizations. Although not yet studied in such detail, similar modifications in language directed to infants have been found in Asian and African languages also. To end on a highly speculative note, perhaps these pitch and temporal patterns found in infant-directed speech with their clear affective messages may provide a clue as to why music carries its powerful affective message and why expressive performance variations are so important for musical communication.

References

Bharucha, J. J., & Krumhansl, C. L. (1983) The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition, 13,* 63-102.

Bregman, A. S. (1990) *Auditory scene analysis.* Cambridge, MA: M.I.T. Press.

Cohen, A. J., Thorpe, L. A., & Trehub, S. E. (1987) Infants' perception of musical relations in short transposed tone sequences. *Canadian Journal of Psychology, 41,* 33-47.

Demany, L., & Armand, F. (1984) The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America, 76,* 57-66.

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

Dowling, W. J. (1988) Tonal structure and children's early learning of music. In J. Sloboda (Ed.), *Generative processes in music* Oxford: Clarendon.

Ellis, C. J. (1965) Pre-instrumental scales. *Ethnomusicology, 9,* 126-144.

Fernald, A. (1985) Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 10,* 279-293.

Fernald, A. (in press) Human maternal vocalisations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* Oxford: Oxford University Press.

Fernald, A., & Kuhl, P. K. (1987) Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10,* 279-293.

Fernald, A., & Mazzie, C. (1991) Prosody and focus in speech to infants and adults. *Developmental Psychology, 27,* 209-221.

Hirsh-Pasek, K., Kemler Nelson, D. G., & Jusczyk, P. W., Wright-Cassidy, K., Druss, B., & Kennedy, L. (1987) Clauses are perceptual units for young infants. *Cognition, 26,* 269-286.

Jusczyk, P. W., & Krumhansl, C. L. (1991) Infants' sensitivity to musical phrase structure. Paper presented at the Biennial Meeting of the Society for Research in Child Development.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Wright-Cassidy, K. (1989) How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16,* 53-68.

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

Krumhansl, C. L., & Jusczyk, P. W. (1990) Infants' perception of phrase structure in music. *Psychological Science, 1*, 70-73.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990) Innateness, experience, and music perception. *Psychological Science, 1,* 272-276.

Thorpe, L. A., & Trehub, S. E. (1989) Duration illusion and auditory grouping in infancy. *Developmental Psychology, 25,* 122-127.

Trainor, L. J., & Trehub, S. E. (in press) A comparison of infants' and adult's sensitivity to Western. *Journal of Experimental Psychology: Human Perception and Performance.*

Trehub, S. E. (1990) The perception of musical patterns by human infants: The provision of similar patterns by their parents. In M. A. Berkley & W. C. Stebbins (Eds.), *Comparative Perception Vol I Basic Mechanisms.* New York: Wiley.

Trehub, S. E., & Thorpe, L. A. (1989) Infants' perception of rhythm: Categorization of auditory sequences by temporal structure. *Canadian Journal of Psychology, 43,* 217-229.

Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1987) Organizational processes in infants' perception of auditory patterns. *Child Development, 58,* 741-749.

Trehub, S. E., Thorpe, L. A., & Trainor, L. J. (in press) Infants' perception of good and bad melodies. *Psychomusicology.*

Perceptual Grouping, Attention, and Expectancy in Listening to Music

W. Jay Dowling

Introduction

Perceptual grouping, attention, and expectancy are three aspects of auditory cognition that I wish to explore in this chapter. I shall contrast selective attention and the application of expectancies, and discuss them in relation to perceptual grouping. This chapter develops further the train of thought begun in a recent article (Dowling, 1990).

LaBerge (1990), following William James (1890), suggests selection and the mediation of expectancies as principal functions of attention, to which he adds the maintenance of focus in information processing. Selection makes an initial separation among incoming stimulus patterns, choosing some for further, more extensive processing, and leaving others unemphasized. Attention links the patterns thus selected with the listener's knowledge represented as expectancies. And attention also includes operations that maintain the selection of a particular type of stimulus pattern over relatively long periods of time.

Here I shall leave aside the function of maintenance and concentrate on selective attention and expectancy, focusing on their relationships to perceptual grouping and to the listener's knowledge. I must caution the reader that for the most part I shall focus on situations in which expectancy is forced to operate on its own, in the absence of the possibility of selective attention. This contrasts with the usual situation in which the two operate hand in hand. In the usual case, expectancies operate on the stimulus patterns selected by the operation of attention, sorting the expected from the unexpected — the matches from the mismatches. In that case matches and mismatches, both already selected in attention, are both available for further processing. But the cases I will consider here involve forcing the application of expectancy processes in isolation from selective attention in situations where attention has been prevented from operating. In these situations expectancies can function in unusual ways, and what I say about expectancies here should not be taken as true of the normal operation of expectancies in the usual case where they are linked to selective attention.

Selective Attention and Expectancy

Selective attention can be characterized as using global features of stimulus patterns to select them for further analysis. The set of stimulus events thus select constitutes a "channel" defined by a particular feature or set of features: a range of pitches, a particular timbre, a spatial location, a loudness, etc. All incoming events in the channel are available for further analysis. In what is often called the "cocktail party" problem, the listener must select one person's voice out of a noisy background of chattering voices. If the voice to be selected differs from the surrounding voices in features such as pitch range, loudness, and spatial location, then the listener will be able to separate it from the background and understand what it says. All that is required is some stimulus feature (or set of features) that distinguishes the target voice from the others.

In music the composer usually (but not always) sets off important streams of information from their surroundings by means of distinguishing features, giving lines important for an understanding of the piece prominence in the musical texture. The listener finds it relatively easy to focus attention on those lines, making use of the structure provided in the stimulus pattern. The listener uses the perceptual groupings provided in the stimuli to focus attention.

When the listener focuses attention on a featurally defined channel, whatever is picked up on that channel can be processed and stored in memory. Suppose I listen to an oboe solo made prominent by its timbre against a background of strings. If the oboe plays a familiar tune I shall be able to recognize it. If the oboe plays a variant of something I have already heard, my expectancies will be brought into play, and I may notice what is now different. And if the oboe plays something novel that I have never heard before, I shall still be able to follow the melodic line. Even with a novel pattern, I shall be able to separate what the oboe is playing from its background, compare it with expectancies, and remember parts of it later.

Two principal aspects of selective attention, then, are that it requires some overall feature defining the stimulus elements to be selected, and that whatever patterns it picks up can be processed further, even when they are novel.

Expectancy processes, in contrast to selective attention, can operate in situations where there is no feature that clearly defines a "channel". In cases where a target pattern is hidden in the musical texture, we can still discern it if we know it is there and what it is doing. For example, I used to play tuba in my school orchestra, and became familiar with the tuba part in a number of pieces. Even though most of the time the composer has carefully hidden the tuba part in the orchestration (so that it rarely pops out at unsuspecting listeners and draws their attention away from more important features), nevertheless I can use my background knowledge to follow the tuba part and discern it clearly. (The reader may imagine that I sometimes hear fairly bizarre versions of Brahms or Stravinsky using this approach.) Expectancy can guide perceptual analysis so that hidden pattern elements that match expectancies can be processed and perceived.

A visual example of the operation of expectancy in the absence of clear stimulus features segregating target patterns from the background is the type of puzzle picture in which we are supposed to "find six lions in the jungle" (Dowling, 1988). There we do not see the lions immediately because the more obvious lion features (tawny fur, light- colored mane) have been removed from the black-and-white drawing. But with knowledge of lion shapes, along with the knowledge that there are six lions to be found, we are going to see the lions after searching them out. Indeed, once we see the lions it will be difficult to ignore them (just as it is difficult for me to hear the *Academic Festival Overture* without an awareness of the tuba part).

Expectancy, then, usually operates in conjunction with selective attention. However, when it operates in the absence of stimulus features that permit selective attention, expectancy can nevertheless succeed in discerning hidden patterns as long as the perceiver is familiar with their shapes and aware of their presence. Pattern elements that more or less match expectancies can be selected for further processing; those that do not match are lost, and not processed further. Without selective attention, expectancy processes operating alone cannot focus on a novel pattern and select it for further processing and memory storage. Expectancy processes involve knowledge the listener already has, and bring about pattern organization in perception on the basis of that internal structure, rather than external features. This

Figure 2. Beginning of the sinfonia to J. S. Bach's Cantata No. 29, Wir danken dir, Gott, an orchestration by Bach of the Partita movement in Figure 1.

demonstration is successful, you will not be able to identify the tunes.

Next, Sound Example 4 repeats the first phrases of the melodies. With each repetition the pitch of one of them is raised by 2 semitones until the two lie in separate pitch ranges. (This example follows van Noorden, 1975.) Now you should be able to attend selectively to either one. The top one is the Swedish nursery tune "Lasse, Lasse Liten" and the bottom one is "Twinkle, Twinkle, Little Star". (I found "Lasse Liten" in an article by Lindblom and Sundberg, 1969).

Now, as the example continues, the pitches of the two melodies move together again. Try to follow one of them as it merges with the other. With a little practice, and with knowing what melody to listen for, you will be able to hear the target melody even when the two melodies fall in the same pitch region. The stimulus pattern at the end of the example is the same as at the beginning, but now your expectancies can be directed precisely to the melody you choose.

In order to fit the melodies together in the interleaved version their rhythm has been regularized and longer notes have been divided up. The melodies are easier to follow when larger portions of them are presented. Sound Example 5 presents a more complete version "Lasse Liten" alone in this stylized rhythm, followed by "Lasse Liten" interleaved with "Twinkle, Twinkle" in Sound Example 6. Sound Example 6 is exactly the same as Sound Example 3, but where before it just sounded like a meaningless jumble of notes, now you can discern one or the other of the melodies because now you know what to listen for. (All things being equal, "Lasse Liten" should be easier to pick out than "Twinkle, Twinkle," since the notes of the former fall on the odd-numbered notes, on the implicit beat in the combined version, while the notes of "Twinkle, Twinkle" fall off the beat; Dowling, Lung, & Herrbold, 1987, Experiment 1).

When the melodies are separate in pitch, you can pick either of them out using selective attention. Then even a melody you have not heard before will be perceived clearly. But when the two melodies are interleaved in the same pitch range expectancy processes must be engaged, and so you must know what melody to listen for in order to discern it. It also takes some practice aiming your expectancies in just the right way. For experiments such as those of Dowling (1973) and Dowling et al. (1987) listeners typically practice hearing hidden melodies for about 1 hr before beginning the experiment.

Until recently I thought that listeners used a temporally guided variant of selective attention to pick out a hidden target melody (Dowling, 1973; Dowling et al., 1987). Now I prefer to view this phenomenon as involving a different type of process based on a series of rhythmic expectancies. The main reason for thinking in terms of a separate process is that selective attention generally succeeds in passing whatever it picks up - novel as well as familiar - on to the next stage of processing. But listeners are unable to attend selectively to every other note in the pattern and identify the melody those notes form, whatever it is. Without being cued to listen for a particular melody (or very small set of melodies) listeners are unable to identify the target. Furthermore, when listeners know one of the interleaved melodies they cannot attend selectively to the notes in between those of the one they know, and identify it without additional cuing (Dowling, 1973). The expectancy process selects only those pattern elements that come close to matching expectancies.

Perceptual Grouping and Attention

Before proceeding to some recent experiments that explore the operation of expectancies in the absence of the possibility of selective attention, I wish first to illustrate the variety of stimulus features that can serve as a basis for selective attention with interleaved melodies. As we have seen, pitch differences can provide selective attention with a basis for "channel" or "stream" selection (McAdams & Bregman, 1979). Other stimulus features, such as differences in timbre, spatial location, and loudness, can also provide that possibility.

Sound Example 7 presents the beginnings of two Schubert songs interleaved in the same pitch range with the same loudness and timbre. Again, it should be difficult to identify them. In Sound Example 8 a timbre difference is introduced. Now you should be able to hear the beginning of "Die Forelle" ("The Trout") and of "Das Wandern ist des Mullers Lust" from *Die schöne Müllerin*. (The sixteenth notes in the latter have had to be replaced with eighths in this stylized rhythm, and a bit of the accompaniment pattern is used to stretch the phrase to match the length of the phrase from the other

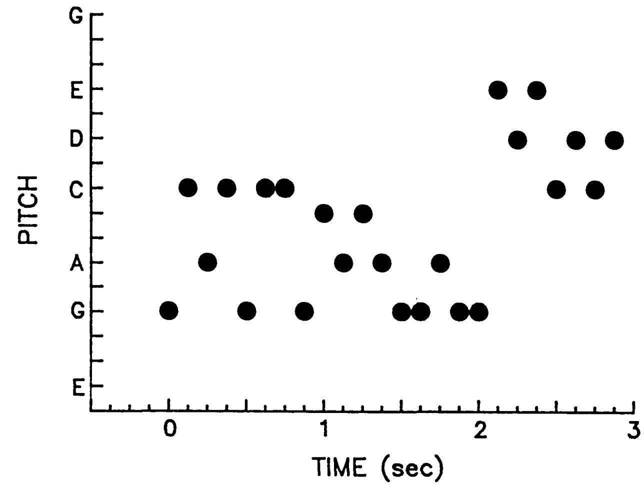


Figure 3. "Happy Birthday" and "Old Macdonald" interleaved in the same pitch range.

tune.) This introduction of a timbre difference has a visual analog in the illustration provided by Figures 3 and 4. In Figure 3 two melodies familiar to American listernes ("Old Macdonald" and "Happy Bitrhday") are interleaved without any cue to distinguish one from the other. In Figure 4 different colors are introduced for the two melodies, and you can pick out either one quite easily

Sound Example 9 presents rhythmically stylized versions of two themes of Richard Wagner's *Der Ring des Nibelungen*: Siegfried's Horn Call and the Ride of the Valkyries. Sound Example 10 presents those themes with the same timbre, but differentiated in spatial location between the two stereophonic speakers. And Sound Example 11 presents them again distinguished by both space and timbre cues.

Finally, Sound Example 12 presents "Lasse Liten" and "Twinkle, Twinkle" differentiated by loudness. Notice that you can attend to the softer of the two melodies as well as to the louder. This is of interest because some models of attention in the 1960s conceived of the focus of attention as boosting the "gain" - of increasing the amplification - of whatever material was selected. Such a mechanism would be counterproductive when the listener is trying to attend to the softer melody - it would decrease rather than increase the contrast. Furthermore, listeners are quite accurate when judging the

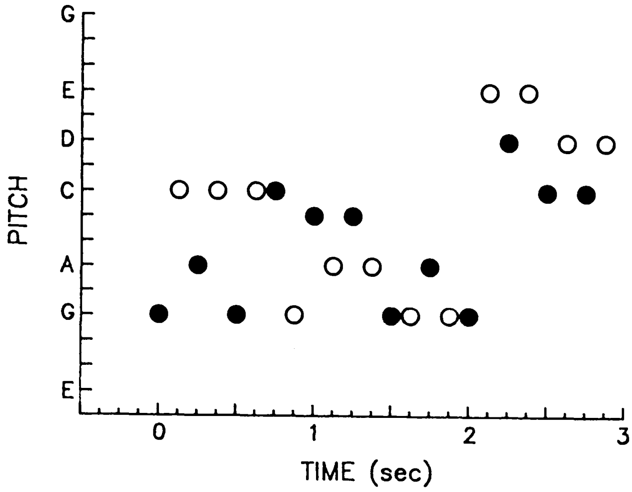


Figure 4. "Happy Birthday" (filled circles) and "Old Macdonald" (open circles) interleaved in the same pitch range.

loudness of the softer of the two melodies when they are attending to it (Dowling, 1973). Selection in focal attention seems to make a stimulus pattern more vivid in perception, but that vividness is not simply a matter of increased loudness or intensity, but rather something subtler and more elusive.

In all of the immediately preceding cases the listener can use selective attention to focus on one or the other of the melodic patterns by means of an overall stimulus feature distinguishing them. Now I shall turn to a pair of experiments dealing with cases in which such a focusing of selective attention on the basis of stimulus contrasts is impossible. In these experiments the listener must use expectancy processes alone, in the absence of selective attention, to pick out the target melodies. Using expectancies involves, of its very nature, using our knowledge of the world as an aid in perception. These experiments indicate ways in which the listener's implicit knowledge of the tonal structure of the music becomes involved in the process of melody perception.

Assimilation of Quarter-Steps to Chromatic Scale Steps

In the first experiment (Experiments 2 and 4 of Dowling et al., 1987, with some additional material) listeners were posed the task of following a

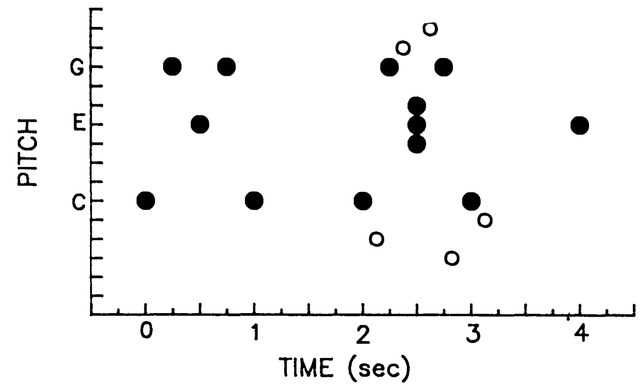


Figure 5. The structure of a trial in the first experiment: A target (filled circles) is presented alone, and then with distractors (open circles). The middle note of the target can move to a new pitch in the second presentation. The listener's task is to follow where the middle pitch moved to, and judge the pitch of the following probe tone in relation to that pitch.

repeated melodic pattern interleaved among distractors. One note (always the same one) of the pattern moved in pitch. The listener had to judge the pitch of that target tone in relation to a probe tone that followed the whole pattern. The result was that when the pitch of the target tone was a chromatic step (as represented on the piano keyboard) pitch judgment was accurate. However, when the target pitch was a quarter step (in between the piano pitches) it tended to be assimilated to neighboring chromatic steps.

The type of stimulus pattern used in this experiment is illustrated in Figure 5. On each trial the listener first heard a standard pattern (C-G-E-G-C), followed by a repetition of that pattern (the solid symbols in Figure 5). When the pattern was repeated the middle tone, which had initially been an E, moved in pitch to other chromatic steps in the neighborhood of the E (D, D#, F, or F#), to quarter steps (which I shall symbolize D+ and F+), or to more remote pitches outside the region in which the listener expected it (the A above or below the pattern). The listener's task was to mentally note the new pitch of the middle tone, and then compare it to the pitch of a probe tone presented following the second pattern. If the probe tone seemed equal in pitch, the listener said "equal;" if higher or lower, the listener said "higher" or "lower."

The first result was that when the patterns were presented without interleaved distractor tones (the open symbols in Figure 5) pitch judgment was quite good no matter where the target tone landed. Performance was equally good for quarter steps and semitones (60 % and 63 % correct saying "equal" when the probe actually was equal in pitch, vs. a chance level of 50 %). Performance was even better (76 % correct) for the remote pitches, which are perceptually quite salient when there are no interleaved distractors.

When distractors were interleaved among the notes of the repeated pattern (as illustrated in Sound Example 13) this picture changed. Performance on the remote pitches in unexpected pitch regions fell to chance. And while performance remained relatively good for chromatic semitones in the expected region, judgments of target pitches that landed on quarter steps became assimilated in most cases to neighboring chromatic pitches. These results can be seen in Figure 6, where the

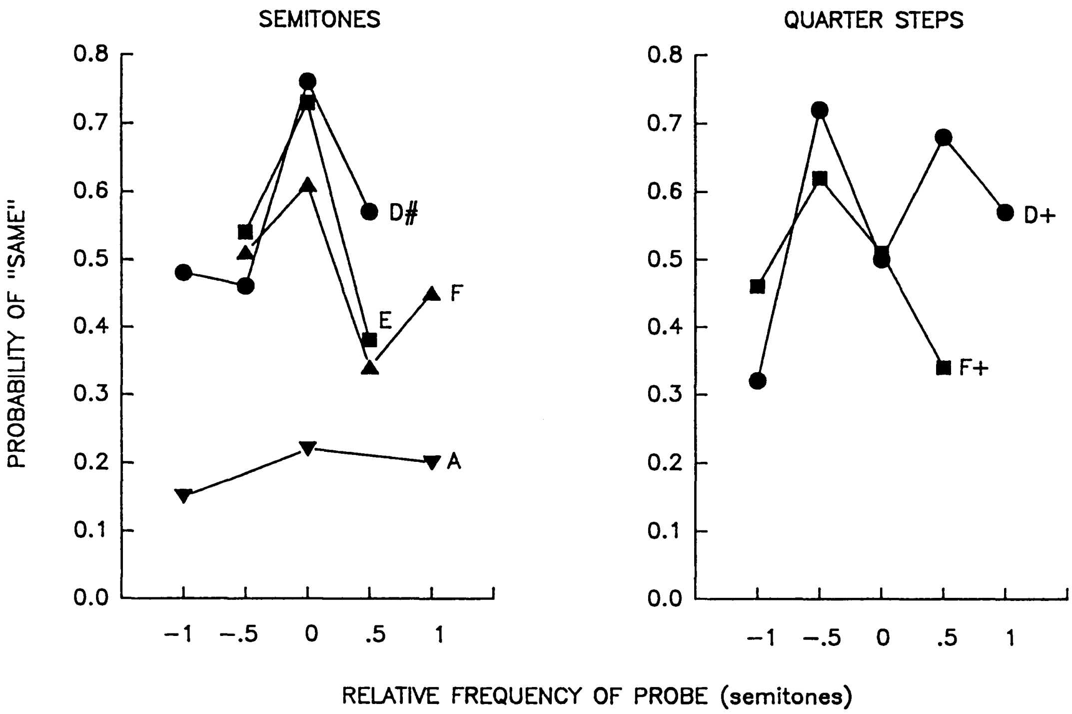


Figure 6. Probability of "same" judgments of probe tones at different pitches relative to the target tone, for target tones that were semitones on the chromatic scale (left panel) and targets tones that fell on quarter steps in between the semitones (right panel).

proportion of "same" judgments is plotted against the pitch of the probe tone relative to that of the target tone. The results for targets that were chromatic semitones can be seen in the left-hand panel, in which separate curves pertain to the different target pitches. First notice the chance level performance for the remote pitches (As). There listeners responded the same whether the probe matched the target or not. The inverted-V shape of the curves for targets near the original E (D, D#, F, F#) indicates better-than-chance performance, since in that case the proportion of saying "equal" when the pitch of the probe really did equal that of the target was greater than when the probe was a quarter step (0.5 semitones) higher or lower. In contrast, the results for targets that were quarter steps are shown in the right-hand panel. There the inverted-W shapes of the curves indicate worse-than-chance performance. With one exception the probe was more likely to be judged "equal" when it was a quarter step higher or lower than the target tone, than when it was actually equal in pitch. Targets that landed on quarter steps became assimilated to neighboring semitones. The one exception was that of the probe F#, which in the key of C is a very salient pitch, and not likely to be confused with anything else.

These results show a clear contrast between selective attention and expectancy depending on the presence or absence of interleaved distractor tones. Without distractor tones, selective attention picks up and interprets whatever stimuli are present in the selected channel, resulting in good performance for remote pitches and quarter steps, as well as expected chromatic pitches. With distractor tones, selective attention becomes difficult or impossible. Judgment of targets remote from the expected pitch region falls to chance, and unusual quarter-step targets within the expected region are assimilated to their more usual semitone neighbors. Thus when forced to rely on expectancies to discern targets the listener's implicit knowledge of the tonal scale system is brought into play at an early stage of processing.

The next part of this experiment is of more direct interest to experimental psychologists than to musicians, since in addressing a methodological issue it simply adds evidence supporting the above conclusions. The methodological concern was that a particular form of response bias might have led to the results shown in Figure 6, without assimilation to semitones occurring in sensory processing. That bias might have affected responses to probes that fell on quarter steps, without any relationship to what the target pitches might have been. This bias would arise if quarter- step probes in isolation sounded a little odd to the listeners, who as a result would be less likely to say "equal" to them. That would explain the lower rates for responses of "equal" seen in both the left and right panels of Figure 6. Therefore to eliminate this possible source of bias we repeated the interleaved portion of the experiment (Dowling et al., 1987, Experiment 2) only restricting listeners to the responses of "lower" and "higher" in judging the pitch of the probe. The usual way to present such results is in a "psychometric' function" in which proportion of "higher" judgments is shown for each pitch of the probe tone. The steeper the function the better the performance in discriminating one pitch from another. Figure 7 shows the psychometric functions for semitone targets

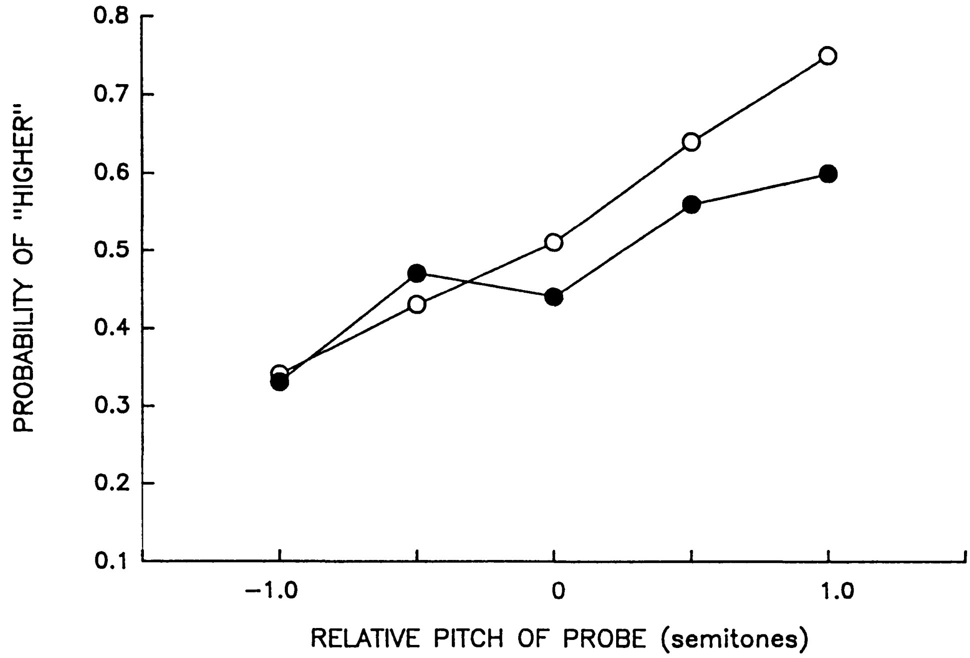


Figure 7. Probability of "higher than" judgments of the probe for target tones at different pitches relative to the target tone for targets that were semitones (open circles) or quarter steps (filled circles).

targets and quarter-step targets. The difference in slopes was statistically significant, with worse performance obtaining for the quarter-step targets

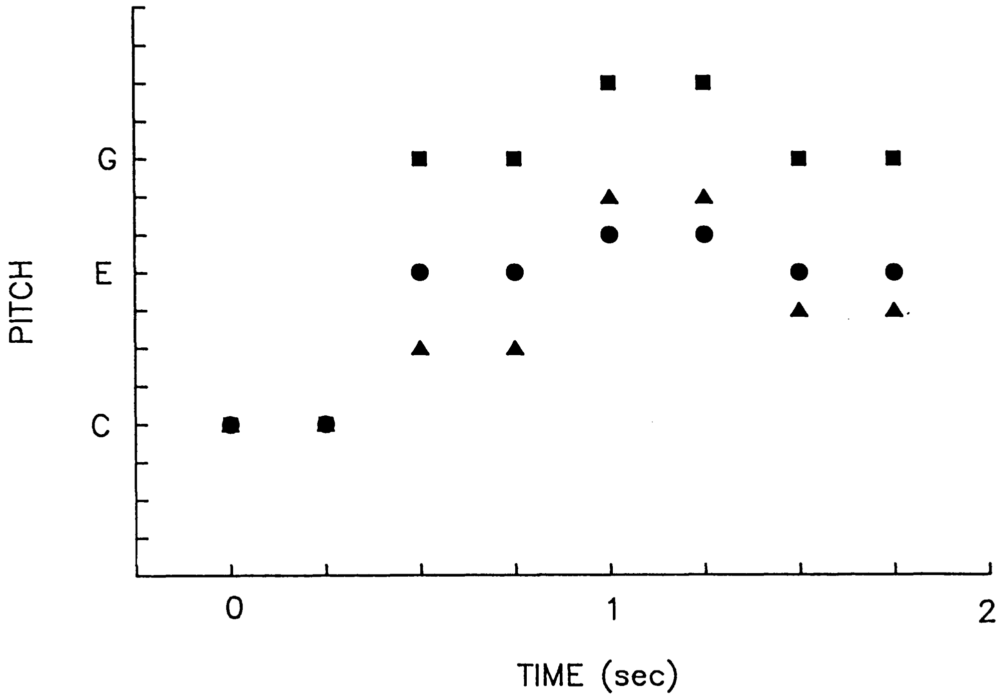


Figure 8. Three versions of "Twinkle, Twinkle": Straight (ST - squares), Tonal Wandering (TW - circles), and Atonal Wandering (AW - triangles).

as before. Furthermore, the quarter-step function itself was not as smooth and regular as the function for semitones, with kinks arising from the type of assimilation effects seen in Figure 6.

(Sound Example 14 was created as a joke, since it seemed amusing to imagine this experiment carried out with vigorous brass sounds on each trial, instead of gentle flute sounds.)

Target and Background Tonality

The second experiment varied the tonality relationships of target melodies and interleaved distractors under conditions that permitted selective attention and conditions that prevented selective attention by thoroughly hiding the target amid the distractors (Andrews & Dowling, 1991). The interesting result in the present context was that the effects of the tonality relationship between target and distractors varied depending on the possibility of selective attention to the target.

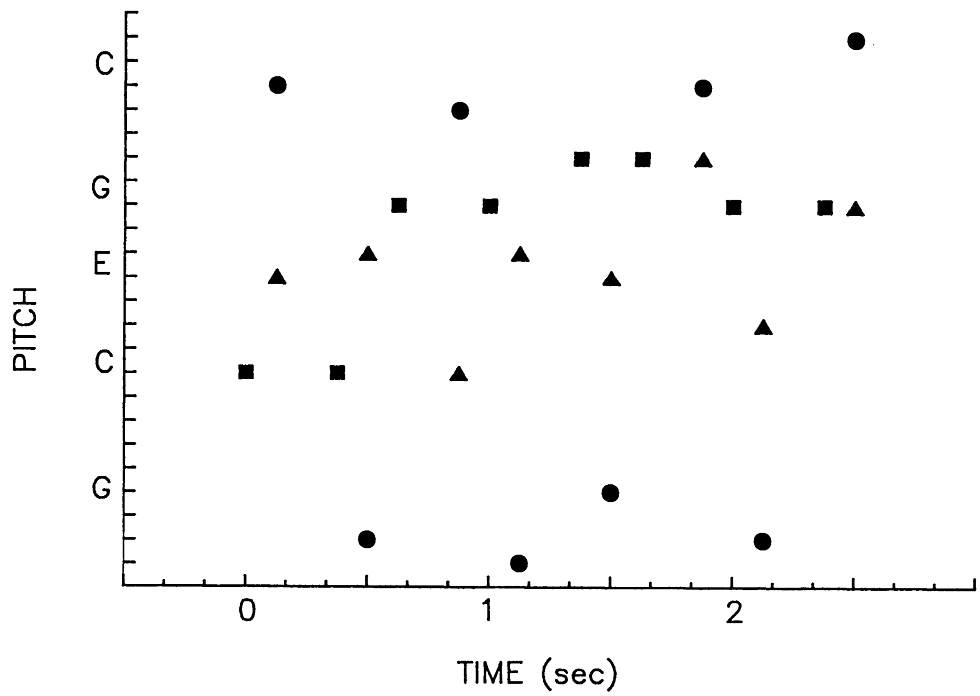


Figure 9. A straight version of "Twinkle, Twinkle" (squares) hidden among distractors outside its pitch range (circles - the Hidden/ Out condition) or inside its pitch range (triangles - the Hidden/ In condition).

Part of our interest in doing this experiment was in the development of attention, and so we carried it out with subjects of ages 5 through 10 years, as well as with adults. We used target tunes familiar to the youngest listeners: "Old Macdonald" and "Twinkle, Twinkle." We were curious whether children at the different ages would be able to follow the target tune if it wandered in pitch while preserving contour (the pattern of ups and downs of pitch), and so introduced wandering as well as straight versions of the tunes. The wandering versions either stayed within the original key ("tonal wandering" - TW) or deviated from it ("atonal wandering" - AW). Figure 8 shows three such versions of "Twinkle, Twinkle."

The listener's task on each trial was to identify the tune. The tunes were presented in one of three degrees of hiddenness among interleaved distractors (shown in Figure 9). "Salient" targets were in a different timbre from the distractors and also louder. In that condition selective attention could easily

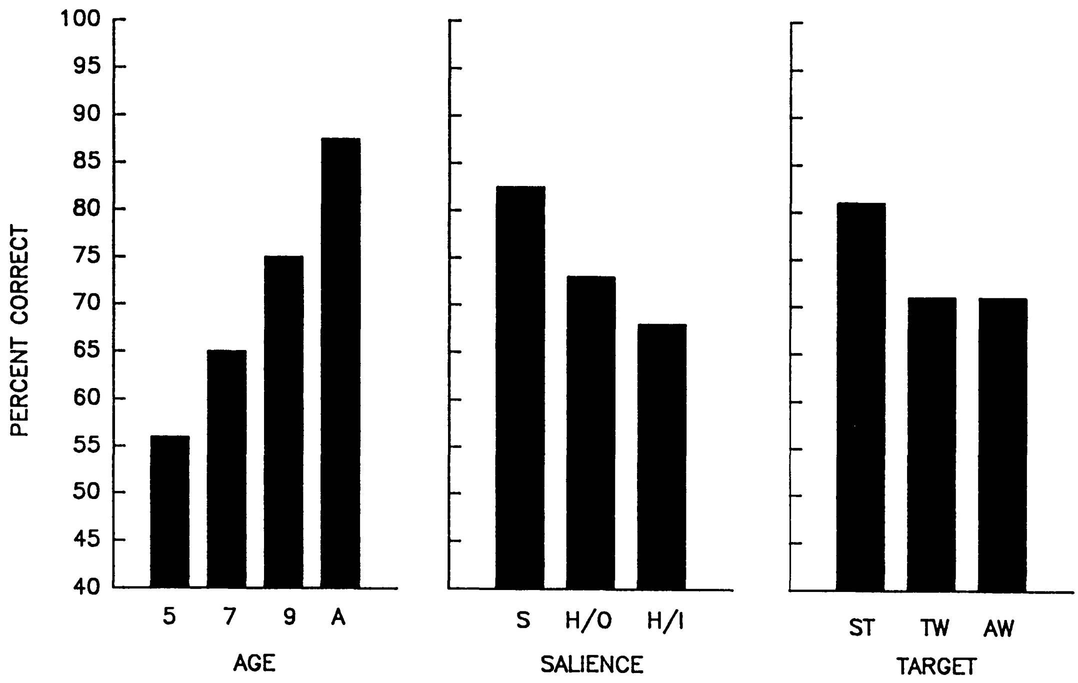


Figure 10. Percent correct in the second experiment as a function of age (5-6 yr, 7-8 yr, 9-10 yr, adult), target salience (Salient - S, Hidden/Out - H/O, Hidden/In - H/I), and target type (Straight - ST, Tonal Wandering - TW, Atonal Wandering - AW).

be directed to the target. "Hidden/Out" targets were of the same timbre and loudness as the distractors, but in a different pitch region. Selective attention could be used to focus on the pitch region of the target, but it was only by the age of 9 or 10 years that the children were able to do that effectively. "Hidden/In" targets were in the same pitch region as distractors, as well as of the same loudness and timbre. That condition was intended to prevent selective attention and require the use of expectancies, as in the previous experiment.

The interleaved distractors were either in the same key as the straight and tonal-wandering targets ("tonal" distractors), or in a distant key from them and closer to the keys into which the atonal wandering targets wandered ("atonal" distractors). The question here is under what conditions will similarity vs. contrast of tonality between targets and distractors be effective.

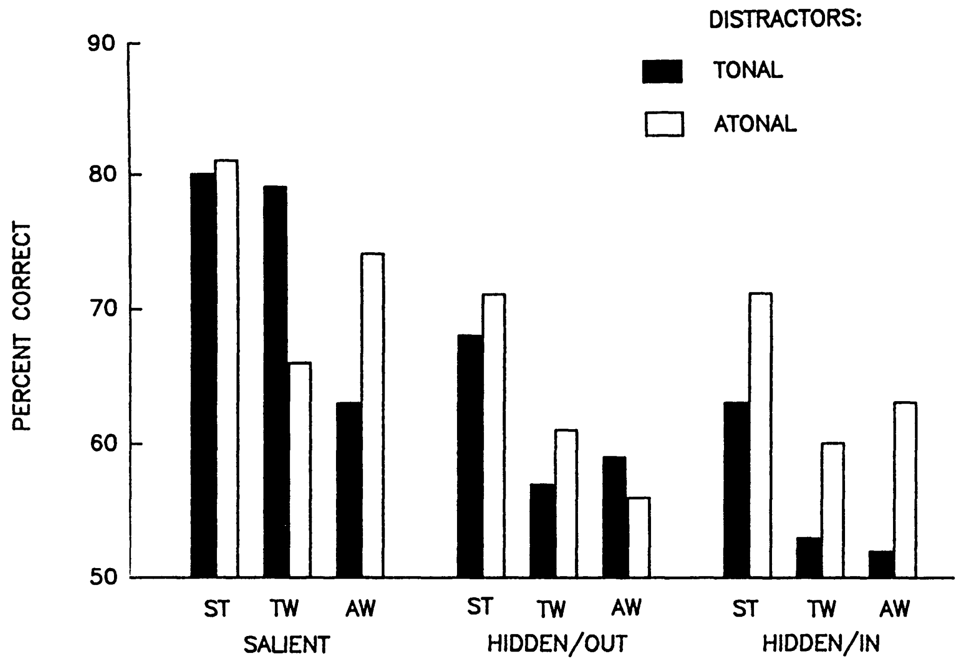


Figure 11. Percent correct for children in the second experiment as a function of target salience, target type, and tonality of distractors (tonal distractors - filled bars, atonal distractors - open bars). Note that there is an interaction of target and distractor tonality in the salient and hidden/out conditions, but that performance with atonal distractors is uniformly better in the hidden/in condition (where selective attention cannot operate).

In broad outline the results were as we expected: performance improved with age and target salience, and was worse with wandering targets (Figure 10). The more
interesting results in the present context involved the interaction of target salience with the tonality relationship between targets and distractors. The pattern of that interaction for the children is shown in Figure 11. (The pattern for the adults was more complicated, involving a further interaction with experience - see Andrews & Dowling, 1991, - Fig. 8.) With salient targets (the left-hand cluster of bars in Figure 11) similarity of tonality between targets and distractors produced the best performance. With Straight and TW targets performance was better with tonal distractors, and with AW targets performance was better with atonal distractors. With

James, W. *The principles of psychology.* New York: Holt, 1890.

Krumhansl, C. L., & Keil, F. C. Acquisition of the hierarchy of tonal functions in music. *Memory & Cognition,* 1982, 10, 243-251.

LaBerge, D. Thalamic and cortical mechanisms of attention suggested by recent positron emission tomographic experiments. *Journal of Cognitive Neuroscience,* 1990, 2, 358-372.

Lindblom, B., & Sundberg, J. Towards a generative theory of melody. *Quarterly Progress and Status Report.* Stockholm: Speech Transmission Laboratory KTH, 1969(4), pp. 53-86.

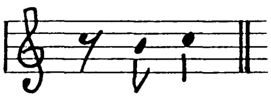
McAdams, S., & Bregman, A. Hearing musical streams. *Computer Music Journal,* 1979, 3(4), 26-43 & 60.

van Noorden, L. P. A. S. *Temporal coherence in the perception of tone sequences.* Eindhoven, The Netherlands: Institute for Perception Research, 1975.

On Gluing Tones

Hans Åstrand

Very few persons would be satisfied with this universal signal without its final cadence:



Yet, moving to a slightly more complex musical phrase, we should not be too sure that a listener recognizes how a group of notes glue together and, more specifically, how many notes belong together, and in what way.

At a conference ten years ago, I had something like a shock, but also a kick, from professor Michel Imberty's conclusions of a series of experiments with children concerning their "perceptual structuring of musical time", resulting in the bold statement that "an uncultivated adult conserves a [musical] judgment and a perception close to those of a child", meaning bluntly that most adults stay at a stage of musical infantility. They would, frankly speaking, feel comfortable with

as a finished musical structure...

One of the points in Imberty's reasoning is that music is "an inborn capacity", but that this capacity certainly "does not manifest itself completely unless constantly solicited by the environment and by an efficient education" - hear, hear! This may sound very self-evident, but it gives food for thought, and it seems eminently connected with Johan Sundberg's theme for this seminar on the gluing of tones. We may have some inborn feeling for how to glue them, but we certainly need much solicitude from the cradle to the grave.

Another quite shocking experience was of recent date: to celebrate the retirement of my dear colleague, the research secretary of the Academy Gunnar Larsson, in February of this year a seminar was arranged, and he was asked to set the theme. He asked the intriguing and impossible question: *Quid est sonus? Quid est tonus?*, and his secret delight was probably to leave behind him such an impossible problem. It did not help at all that he brought with him an expert on quantum chemistry, a professor of mathematics, a leading psychoanalyst, etc. - of course there was no definite reply telling us what the tone really is - Gunnar implies that there may be some sinister dark horses - call them quarks, black holes or otherwise - that actually propel sound waves from the sound source to our poor ears, not just simple air molecules. This could imply that John Cage's dream of hearing all music ever produced all over the world in all times, and at the same time, could be made true somewhere out there among the black holes, once you find the net that can catch the black little devils carrying the sounds and tones...

But even considering the tones as we believe to know them already, there are immense problems defining them. Let us remember Pierre Schaeffer's ambitious efforts to investigate the musical objects in his monumental *Traité des objets musicaux*, and the rather obvious criticism - for instance in the volume *le traité des objets musicaux dix ans après* edited by his pupil Michel Chion - against its isolation of single sound objects, thus not considering the combination - or gluing! - of contiguous *objets sonores*.

And in this context - *la musique concrète* -, we find ourselves approaching the overwhelming masses of electro-acoustical sounds, evading standard discreteness describing "notes" by their *glissandi*, clusters, envelopes and sophisticated frequency modulations. Here the composer as well as the listener confront many black holes to investigate.

As a first halt in this simplified sample of tone-glues, I would like to point at the two axes to consider; we mostly think of a temporal parameter, moving in a linear direction leftright, but glue is equally vital, and tricky, for the vertical axis, not only for consonant versus dissonant chords in nice occidental art music but for all simultaneous vertical agglomerations of discrete notes to dizzy frequency clouds, whether they hide black holes and nasty quarks or not...

So much for adding complications to the proceedings of today; now I would like to ask the panel, and the audience, for other considerations and comments on the tricky gluing procedures of notes and sounds. Provocatively, I should perhaps point out that what has been said today clearly indicates that music is *much more basic* in human life than is often supposed, that music psychology is maybe the central domain of psychology and should be considered as such. We musicians always knew that in the beginning was not poor *logos* but *sonology*, *id est* MUSIC.

The summary of the ensuing discussion between members of the panel (i.e. the lecturers, Gerald Bennett [GB], Jay Dowling [JD], Carol Krumhansl [CK], Johan Sundberg [JS], and myself [HÅ]) and participants "on the floor" concentrates on some interesting points raised during the preceding papers:

1. GB took up the thread of logos vs. sonus/tonus, expressing his fascination with the idea that language and music share common processing mechanisms at a prelingual stage of a child's development. Later on, language seems to continue to use the apparently inborn grouping rules in a most strict manner, while in music, and to some extent in heightened uses of language like poetry, the bending of these rules takes on expressive significance.

JD stressed differences in the observation of his own children developing tendencies to "visit" various stations of partials and intervallic stepwise movements, contrasting with speech glissandi, also music's rhythmicism as opposed to rhythmical structures of speech.

2. JS pondered GB's example with Robert Schumann's *Wehmut* (from *Liederkreis*) and Schumann's assumption that the different grouping codes implicated in the voice part vs. the piano "accompaniment", including their conflicting "loading", were actually perceived by the listeners; were those codes more or less "inborn" or the result of a slow historical educational process? If the latter, was this an ongoing process towards ever greater sophistication? GB saw in the individual a very definite tendency towards ever greater sophistication in performing and listening to music. Historically, this tendency is clearly present as well, but another generation can be interested in sophistication of tonal structure rather than grouping, for example, and one aspect of music having received much attention earlier

can suddenly become greatly simplified.

CK reported the experience of an experiment with music (seemingly) lacking cognitive structures, where Karlheinz Stockhausen's *Klavierstück IX* was played and the subjects were asked to locate structural boundaries, which they did find (in changes of pitch, structure, character, repetitions, etc., thus "well-known" elements in occidental art music) without much difficulty and were very much agreed on where to place them.

3. Bengt Edlund (Lund University) asked from the floor whether JD in his distinction between selective attention and expectancy considered them to operate isolated from each other or together (as mostly described by other music psychologists, such as Leonard B. Meyer); JD found them both active, mostly at the same time, but that in experiments one of them could be more or less isolated and partly cut off.

4. HÅ asked whether there is a tendency in infants to gravitate towards tonality structures; JD considered such a development to be clearly gradual, although octave perception occurs at an early stage.

GB asked what possible difference a propensity towards diatonic scales could make in understanding the perception of music. Because art always requires distance from and extension of natural perceptual and intellectual proclivities, and because there is already so much important and exciting music not based on diatonicism, the interpretation of evidence of an inborn diatonic propensity would have to be undertaken with care.

5. On a question from the floor concerning nonoccidental music perception, CK related an experiment involving a psychologist born in India, with a developed bicultural training; in Indian music there are tonal hierarchies analogous to those in occidental (polyphonic...) music, although differing in detail; both Western and Indian listeners showed sensitivity to the style-appropriate hierarchy.

6. HÅ spoke of occidental polyphony as a unique development not really to be found in other highly developed musical cultures.

In this connection, JS pointed to the fact that harmonic spectra may be considered as an origin of the diatonic scale, with its consonance/dissonance balance, intervallic division of the octave, etc.

HÅ insisted that the partials might have a basic importance in the "training" of intervallic structuralization, pointing to the normal experience of the opening measures of Beethoven's ninth symphony, with their open fifths on d - a without a third, giving a very strong and wellknown impression of the fifth partial, the major third, that is f sharp and not the later resultant tonality of d minor with a natural f - maybe an intentional ambiguity on Beethoven's part? How important is the "harmony of the spheres" of the antiquity and medieval music theory for the development of such intervallic structures as are implied in the partials?

JS then spoke of the Bohlen-Pierce scale, demonstrated by Max Mathews and John Pierce (cf. M. Mathews, J. Pierce & L. Roberts, "Harmony and new scales", in J. Sundberg (ed.): *Harmony and Tonality*, Royal Swedish Academy of Music: Publication No. 54 (1987), p. 59-84), whose pitches have an inharmonic spectrum. Music written with this scale can be made to show dissonance/consonance relationships similar to those we know from the diatonic scale. GB sees the importance of these experiments for composers in the insight into the intimate relationship between the harmonic ordering of sounds in a most general sense and the inner structure of the sounds themselves.

To a question from the floor concerning equal temperament vs. other ("natural") temperaments, JS pointed out that musicians often seem to use a quasi-Pythagorean tuning at the onset of chords and successively and slowly approach a more harmonic tuning, if there is time for it, i.e., if the tone is long enough; in instruments with fixed tuning, equal temperament must be considered the most adequate and "operative".

To conclude this summary, I might be allowed to go back to Gunnar Larsson's provocative question *Quid est sonus/tonus?* Of course, neither he, nor Pierre Schaeffer meant investigating just isolated sounds/tones. It takes at least two tones to create a rhythm, or a "chord" - glued aggregations. It takes two "parts" - at least - for the other vital aggregation, sound producer and receiver, although not glued together in the same way. There is yet another bipolarity behind all glued and perceived tones: innate capacity and - more or less conscious and felicitous - "training". Whether harmonic or inharmonic, glued tones of infinite varieties have formed and will form all kinds of "music", this very basic expression of human experience. We are told that it starts in the prenatal stage - as all vital human expressions probably do -, but it should not be allowed to remain in its infant structures. The better we glue together, the more "adult" our music will be.

List of Sound Examples

1. Captions to Bennett's Sound Examples

Sound Example 1.A string of synthetic notes of equal pitch (368 Hz, approximately f-sharp above middle c) and equal duration (250 ms from the beginning of one note to the next).

Sound Example 2.A string of synthetic notes of the same pitch and duration as Example 1. The amplitude of every fifth note is about 4.5 dB higher than that of the other notes.

Sound Example 3.A string of synthetic notes of equal amplitude. Every fourth note has a pitch of 441.6 Hz; the others have a pitch of 368 Hz.

Sound Example 4.A string of synthetic notes of equal amplitude. Every sixth note has a duration of 500 ms, the others have a duration of 250 ms.

Sound Example 5. The beginning of the motet *Nuper rosarum flores* by Guillaume Dufay (1436).

Sound Example 6. The beginning of the first movement of the Sonata for Piano KV 283 by Wolfgang Amadé Mozart (1774).

Sound Example 7. The beginning of the song *Wehmut* from the *Liederkreis* op. 39 on poems by Joseph von Eichendorff by Robert Schumann (1840).

Sound Example 8. The beginning of the first movement of *Octandre* for Eight Instruments by Edgard Varèse (1923).

Sound Example 9. The beginning of the first movement of the *String Quartet No. 13 in b-flat minor* op. 138 by Dmitri Shostakovich (1970).

Sound Example 10. The beginning of *Williams Mix* for tape alone by John Cage (1952).

Sound Example 11. The beginning of *Etude aux sons animés* by Pierre Schaeffer (1958).

Sound Example 12. The beginning of *Region IV* of *Hymnen* by Karlheinz Stockhausen (1966/67).

2. Captions to Dowling's Sound Examples

Sound Example 1. Partita No. 3 for solo violin of 1720 (BWV 1006).

Sound Example 2. Sinfonia from J S Bach's Cantata No. 29, *Wir danken dir, Gott*, where the solo violin line appears more or less intact in the right-hand organ part.

Sound Example 3. Two familiar melodies interleaved with each other in the same pitch range and with the same loudness, timbre, and spatial location.

Sound Example 4. Repeated presentations fo the same melodies repeated as in Sound example 3. With each repetition the pitch of one of them is raised by 2 semitones until the two lie in separate pitch ranges. (After van Noorden, 1975.)

Sound Example 5. A Tegnér's nursery tune "Lasse, Lasse Liten" presented alone in stylized rhythm

Sound Example 6. The melody of Sound Example 5 interleaved with "Twinkle, Twinkle".

Sound Example 7. Beginnings of two Schubert songs interleaved in the same pitch range with the same loudness and timbre.

Sound Example 8. Same as Sound Example 7, but a timbre difference is introduced that makes it easier to recognize the melodies, "Die Forelle"

("The Trout") and of "Das Wandern ist des Müllers Lust" from F Schubert's Die schöne Müllerin.

Sound Example 9. Rhythmically stylized versions of two themes of R Wagner's *Der Ring des Nibelungen*: Siegfried's *Horn Call* and the *Ride of the Valkyries*.

Sound Example 10. Same themes as in the preceding example presented with the same timbre, but differentiated in spatial location between the two stereophonic speakers.

Sound Example 11. Same themes as in the preceding example distinguished by both space and timbre cues.

Sound Example 12. "Lasse, Lasse Liten" and "Twinkle, Twinkle" differentiated by loudness.

Sound Example 13. When distractors were interleaved among the notes of the repeated pattern (as illustrated in) this picture changed.

Sound Example 14. was created as a joke, since it seemed amusing to imagine this experiment carried out with vigorous brass sounds on each trial, instead of gentle flute sounds.)

3. Captions to Krumhansl's Sound Examples

Sound Example 1. Minuet #14: Natural version, used in Krumhansl and Jusczyk (1990)

Sound Example 2. Minuet #14: Unnatural version, used in Krumhansl and Jusczyk (1990)

Sound Example 3. Minuet #13: Natural version, used in Juszcyk and Krumhansl (1991, Experiment 1)

Sound Example 4. Minuet #13: Unnatural version, used in Juszcyk and Krumhansl (1991, Experiment 1)

Sound Example 5. Minuet #12: Backward Natural version, used in Juszcyk and Krumhansl (1991, Experiment 2)

Sound Example 6. Minuet #12: Backward Unnatural version, used in Juszcyk and Krumhansl (1991, Experiment 2)

Sound Example 7. Minuet #18: Forward Natural version, used in Juszcyk and Krumhansl (1991, Experiment 3)

Sound Example 8. Minuet #18: Backward Natural version, used in Juszcyk and Krumhansl (1991, Experiment 3)

4. Captions to Sundberg's Sound Examples

Sound Example 1. Deadpan performance of an excerpt from Franz Berwalds String Quartet

Sound Example 2. Illustration of the effect of rule *Faster Uphill.* The same excerpt is played three times, first exactly as written in the score, second the rule is applied, but to an exaggerated extent, and third the rule is applied at a moderate quantity.

Sound Example 3. Illustration of the effect of rule *Leap Articulation.* The same excerpt is played three times, first exactly as written in the score, second the rule is applied, but to an exaggerated extent, and third the rule is applied at a moderate quantity.

Sound Example 4. Illustration of the effect of rule *Phrasing.* The same excerpt is played three times, first exactly as written in the score, second the rule is applied, but to an exaggerated extent, and third the rule is applied at a moderate quantity.

Sound Example 5. Synthesized singing. In the first version each individual tone is marked by an overshoot in fundamental frequency in descending intervals and a crescendo-diminuendo pattern for each individual note. In the second version, a crescendo-diminuendo pattern is introduced that covers each pair of bars.

Sound Example 6. Significance of timbre for gluing tones together. A series of tones are played twice. The first time adjacent tones differ significantly in timbre; as a result, the tone sequence disintegrates into a series of unrelated tones. The second time, the timbral differences between adjacent tones is small, and as a result, the tones create a melodic line.

Sound Example 7. Significance of magnitude of duration effects. First a deadpan version is presented of a Bourree from one of JS Bach's suites for solo cello. In the following versions lengthenings and shortenings of 80 ms, 40 ms, 20 ms and 10 ms are randomly distributed among the notes.

Sound Example 8. Sample of organ playing according to the "Alte Speilweise".

GLUING TONES

GROUPING IN MUSIC COMPOSITION, PERFORMANCE AND LISTENING

Why do we hear that certain tones in a melody belong together while others do not? How come that we feel when a phrase has come to its end?
These are signs of grouping, a reflection of the propensity of the human brain to form groups out of sequences.
There seem to be at least three actors that help the music listener in grouping the tones: the composer, the performer and the listener himself. They all apply their own special glue to the tone so that they stick together. What is this glue?
In this book four scientists discuss the components of this musical tone glue: Gerald Bennett, composer and music theoretician, Johan Sundberg, music scientist and Carol Krumhansl and Jay Dowling, psychologists and experts in music cognition.

Publications issued by the Royal Swedish Academy of Music
No 72, Stockholm 1992