Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

Research Article

From idea to acoustics and back again: the creation and analysis of information in music[1]

Joe Wolfe

University of New South Wales, Sydney, 2052, Australia
E-mail: J.Wolfe@unsw.edu.au

OPEN ACCESS

Citation: J. Wolfe (2018) From idea to acoustics and back again: the creation and analysis of information in music. *Substantia* 2(1): 77-91. doi: 10.13128/substantia-42

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Abstract. The information in musical signals – including recordings, written music, mechanical or electronic storage files and the signal in the auditory nerve – are compared as we trace the information chain that links the minds of composer, performer and listener. The (uncompressed) information content of music increases during stages such as theme, development, orchestration and performance. The analysis of performed music by the ear and brain of a listener may reverse the process: several stages of processing simplify or analyse the content in steps that resemble, in reverse, those used to produce the music. Musical signals have a low algorithmic entropy, and are thus readily compressed. For instance, pitch implies periodicity, which implies redundancy. Physiological analyses of these signals use these and other structures to produce relatively compact codings. At another level, the algorithms whereby themes are developed, harmonised and orchestrated by composers resemble, in reverse, the means whereby complete scores may be coded more compactly and thus understood and remembered. Features used to convey information in music (transients, spectra, pitch and timing) are also used to convey information in speech, which is unsurprising, given the shared hard- and soft-ware used in production and analysis. The coding, however, is different, which may give insight into the way music is understood and appreciated.

Keywords. Information, music, composition, cognition, coding.

INTRODUCTION

Many digital recordings encode microphone signals as 16 bit numbers, which gives a dynamic range (maximum signal range/digitisation step) of 2^{16} = 96 dB. The signal is sampled at 44.1 kHz. This gives a data transmission rate of 706,000 bits per second or 706 kBaud per channel, not counting error correction bits. A traditional compact disc (CD) can store about a thousand megabytes of data: enough to store several hundred novels, or about eighty minutes of uncompressed recorded music. This raises the questions: Where do all these data come from? How much is provided by the composer, by the players and the instruments?

[1] This paper was originally presented and published as a plenary lecture at the Eighth Western Pacific Acoustics Conference, Melbourne, 2003. (C. Don, ed.) Aust. Acoust. Soc., Castlemaine, Australia.

What happens to that torrent of data when it reaches the listener? The rate delivered by a stereo CD – about one and a half million bits per second – appears to be equivalent to a novel every several seconds. Can our ears and brains cope with such a rate? And finally: Why do we like it? As a composer and physicist, I try here to address these questions from both sides. I suggest some answers, and indicate where research is currently looking for others.

Data compression

Data files can usually be simplified or compressed because they contain much redundancy. For instance, a CD could contain 75 minutes of 1 kHz test tone. This is redundancy on a scale of 1 ms: to a suitably sophisticated receiver, the signal could be sent as the text instruction "p = (1 mPa) sin (2pi*t/ms), 0 < t < 4500 s", which requires only 352 bits in ASCII. For an example of redundancy on a longer scale, consider "house music" in which short sound segments are sampled and repeated many times.

Kolmogorov [1] and Chaitin [2] independently introduced algorithmic entropy to quantify the difference between unpredictable and redundant signals. To paraphrase Chaitin, consider two binary numbers:

1011110010001101010110111011000001101010

and

0101010101010101010101010101010101010101.

The first "looks" random: it was obtained by tossing a coin forty times. The simplest way of transmitting that number is sending the number itself. The second does not "look" random: it can be reconstructed from the instruction "print '01' twenty times". That instruction contains more than forty bits of information, but for a very long predictable number, the reproduction instruction may be rather smaller than the number (e.g. the 208 bit instruction "print '01' a million times" produces a 2 million bit output). The algorithmic entropy is proportional to the number of bits of information in the minimum message needed to reconstruct a signal. (It is thus proportional to the log of the number of permutations and consistent with Gibbs' definition.) The more simple or predictable a signal, the lower its algorithmic entropy and the more it may be compressed. Conversely, the richer in information, the higher the entropy, and the more it resembles a random signal – at least to a receiver that cannot decode it. When sound signals are stored to be heard by humans, they are often compressed using the MPEG (mp3) algorithms. These take advantage of masking in human hearing: one frequency band may mask others, so the masked sounds are omitted. A reconstructed MPEG waveform produces an auditory illusion: its waveform has little resemblance to the original, but it sounds very similar.

Recorded music has relatively small algorithmic entropy. Indeed, its underlying order, at several different levels, is one of its attractions. At the lowest level, there is high redundancy in the waveform. A note with a definite pitch is quasi-periodic: one cycle with the pitch period is followed by many others very like it. Of course, in real, interesting instruments, the periodicity is only approximate: transients and vibrato lead to varying waveforms, as do non-harmonic components in percussion and plucked strings.

Systems of music notation take advantage of this redundancy. In standard (Western) notation, vertical positions of notes on the staff plus accidentals specify pitches and thus, approximately, frequencies. A discrete set of note symbols, plus a few other data (tempo and articulation), specify durations. Some information about the type of waveform, and much else, is contained in a word at the beginning of the music: the name of the instrument that is to play it. From this relatively small data set, performers and instruments construct complete waveforms.

The information content of written music is relatively easy to quantify because written music is digital in pitch and in time: relatively small sets of discrete pitches and durations are used. In contrast, performed music is only approximately digital: musicians make fine adjustments to the durations and timing and, except for keyboard players, adjust the pitch slightly according to context. These adjustments contribute to musical interpretation, to which topic we shall return.

Fig. 2 shows a short example: the first two phrases of the theme of the slow movement in Mozart's clarinet concerto. One way of coding it is to sample the pitch regularly in time. The lowest suitable sampling frequency is the metronome marking times the lowest common multiple of its subdivisions. Most simple themes could be adequately sampled at a rate of order 10 Hz. Five octaves (61 notes) covers the range of most orchestral instruments and can be coded with 6 bits (i.e. $61 < 2^6$), so the notes and rests could be coded at about 60 bits^{-1} (60 Baud).

Most notes are longer than the sampling time, however, so this signal can be compressed by coding for the durations of the notes as well as their pitch. Traditional notation does just this, *inter alia* (Fig. 2b). The bar lines appear to be redundant, but to musicians they also give contextual information relevant to musical expression [3]. They also provide a correction mechanism for accumulated errors in duration decoding.

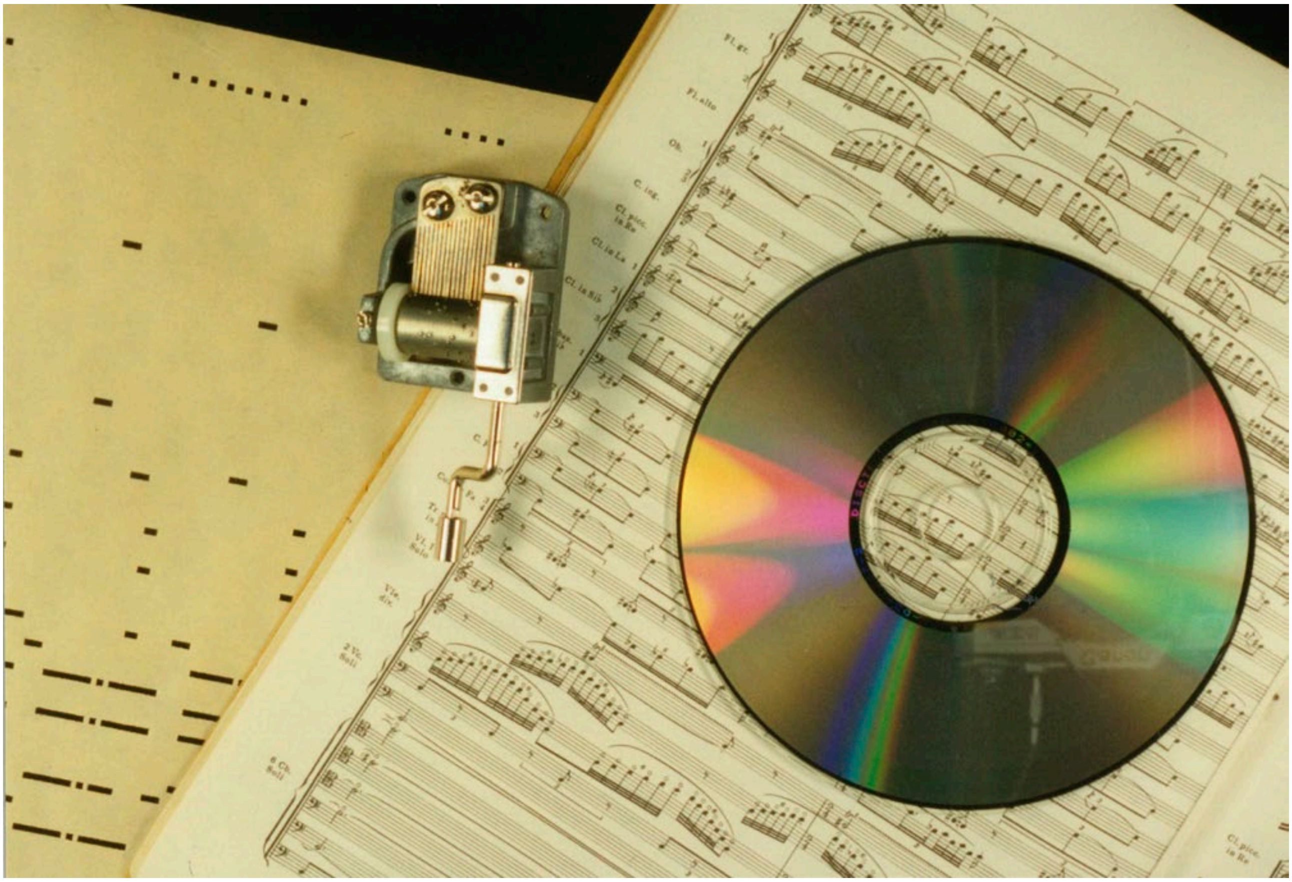


Figure 1. Four digital storage media. (a) The cylinder and comb from a music box play 16 bars from *Lara's Theme* (M. Jarre). The 18 tines of the comb have different masses and thus play different notes when struck by spikes on the cylinder. It has 18 parallel channels – circles round the cylinder. The loudness is binary (spike or no spike, note or no note). The timing is in principle analog, but is here quantised in multiples of 1/12 of a bar. The uncompressed data content of this cylinder is therefore 18 x 12 x 16 = 3456 bits. (b) The pianola roll in the background also has parallel binary channels, but the length of the hole determines the time the strings sound before the damper is replaced. In that sense, both duration and timing could be analogue, but again they are quantised in this example. The uncompressed data content is 35,000 bits per metre. (c) Standard Western music notation is (largely) parallel binary digital coding: each line and space (parallel channels) represents a pitch, though that pitch can be varied by sharps and flats. The time coding is encoded digitally in symbols (see Fig 2). This example (*The Rite of Spring*, I. Stravinsky) has about 30,000 bits on this page, which lasts a few seconds, using a coding somewhat like that in Fig 2c. (d) The CD also carries a binary digital signal ("pit" or "no-pit" in the track) but it is different in all other aspects. The signal is carried in serial rather than in parallel, and it encodes numbers that are proportional to the pressure of a sound wave. This CD records about 5 x 109 bits, not counting error correction bits. The storage efficiencies are approximately: a) 5 x 10^{5} bit.kg1, b) 10^{6} bit.kg^{-1}, c) 10^{7} bit.kg^{-1}, d) 3 x 10^{11} bit.kg^{-1}. The apparatus required for re-creation varies greatly in size: that for (a) is shown (~ 0.01 kg), that for (c) is ~ 10^{4} kg.

Figure 2c shows how a simplified binary parallel coding can represent those aspects of traditional notation used here. This example has a data content of 266 bits and, over a duration of about 13 s, a transmission rate of only 20 Baud. No correlation between the quantity of information and its value is implied, of course: many people consider this 266 bit theme more valuable than, say, a Gbyte of white noise!

The encoding used by music sequencers is close to that of music notation. These, the electronic progeny of the musical automata in Fig 1, are computer programs that output signals to synthesisers via a standard Music Industry Digital Interface (MIDI). The MIDI standard transmits data at 31.25 kBaud in serial form. This permits parallel voices and a range of instructions, and its design allowed bandwidth for further developments. Alternative coding protocols have been proposed [4]. More sophisticated representations include expression – variations in loudness, amount of vibrato, fine adjustments to pitch and to timing [5,6].

Another crude but pragmatic way of computing data content is to look at the data files of note proces-

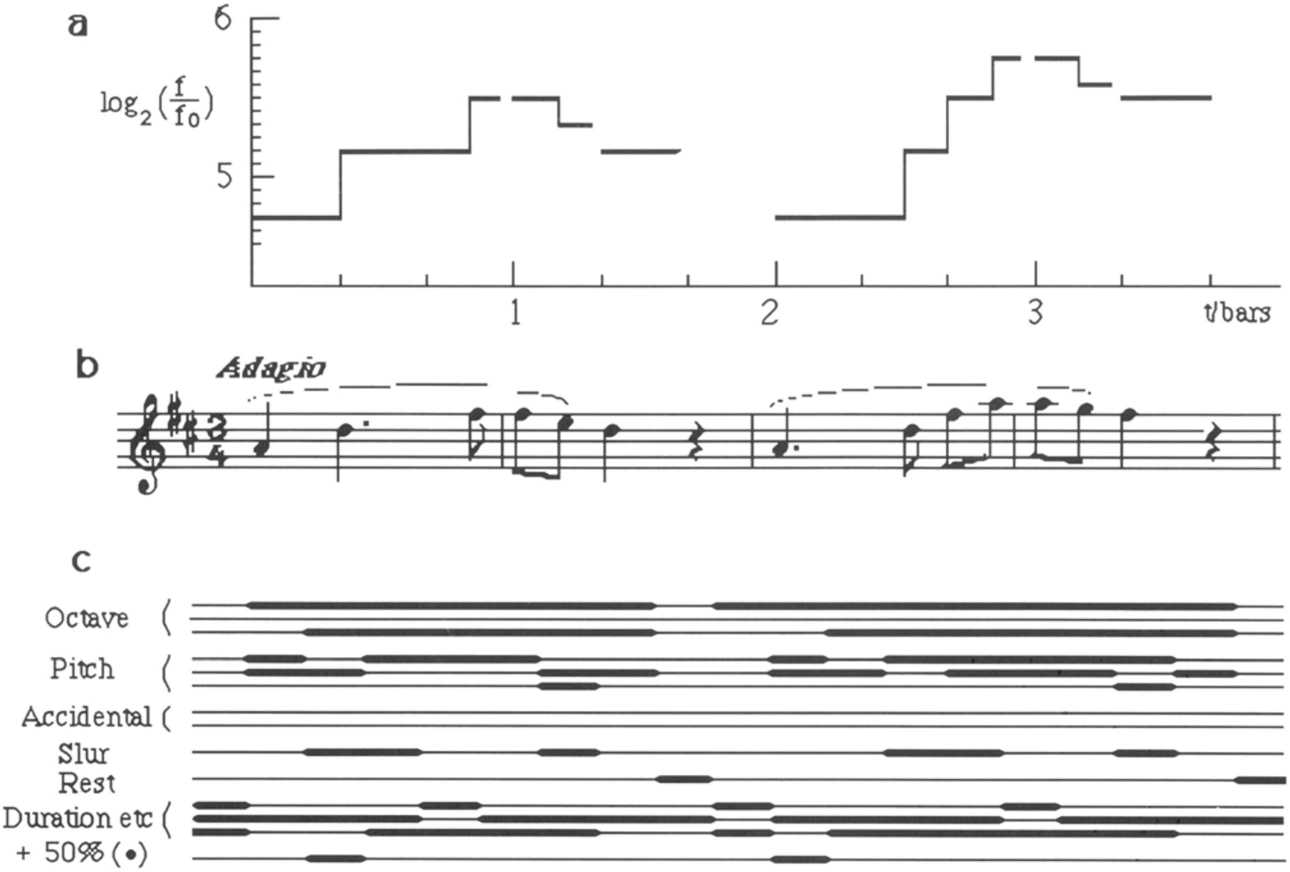


Figure 2. Three ways of coding the first four bars of the theme of the slow movement of Mozart's clarinet concerto. (a) is a semi-log plot of the pitch frequency as a function of time. On the time axis, the larger tics are bars (measures) and the smaller are beats. On the frequency axis, the larger tics are octaves. Notes an octave apart have the same letter name e.g. C5 and C6. The reference frequency is the note called C0, which is currently about 16.3 Hz. The smaller tics are one twelfth of an octave *i.e* frequency ratio of $2^{1/12} \cong 1.059$). These are called equal-tempered semitones: they correspond to the notes on an electronic keyboard. (b) is essentially traditional notation. The vertical and horizontal axes have been adjusted to make it an exactly semi-log plot by varying the spacing between lines, which may represent 3 or 4 semitones. The shapes of notes are a digitised code for duration that has several advantages over the analog time scale used in (a). (c) is a parsimonious parallel binary coding, which is more akin to traditional notation than to (a). The pitches of notes are shown by their octave (top 3 bits) and the note names (next 3 bits) with the most significant bit at the top. The next 2 bits allow for accidentals (sharps, flats and naturals) that are not needed in this example unless the key signature is omitted. The next bit indicates slurs: whether the note is continuous with the preceding one (the curved lines or slurs in (b)). The next bit indicates a rest (silence) of the appropriate length. The next 3 bits show the negative log durations with respect to a whole note. Semibreves, minims, crotchets, quavers and semiquavers (whole, half, quarter, eighth and sixteenth notes) are represented by 000 to 100. 101 is used for a bar line. The final bit allows an increase of 50% in duration (indicated by a dot in (b)). The duration code 111 is reserved as a signal to toggle the coding to text, so that occasional data such as tempo, key signature, expression marks can be added more efficiently. (The unequal spacing of channels is a guide for the eye only).

sors. These are to music what word processors are to text, and are widely used by composers and editors to write and to print music (*Sibelius* and *Finale* are commercial examples). They store written music in digital files that are similar to, but more elaborate than that in Fig 2c. On my hard disc is a 160 kbyte note processor file for a symphonic work. It takes 23 minutes to play, and so its printed score delivers data to the conductor at an average rate of 900 Baud, or 900 bits per second. To achieve the same transmission rate reading this article (not counting figures), one would need to read it at 1100 words per minute. It should be noted that conductors do not absorb all the information in a score in real time.

While comparing written music and written text, it is worthwhile contrasting them as well. One difference is cultural: more people can read text than can read music. Even to those literate in both, however, the aural re-creation is more important in music. Most musicians prefer hearing performances to reading scores, whereas I expect that most text-literate people prefer reading novels (at a rate of several hundred Baud) to hearing them read aloud, at slower rates. In both cases, the auditory transmission contains a great deal more information than does the written version.

THE ORIGIN OF INFORMATION IN MUSIC

Melodic and harmonic structures are good examples of redundancy. In a high information/ high entropy signal, all pitches would occur in approximately equal numbers and it would be impossible to predict the next note: a high information signal sounds or looks random. Music is ordered[2], and this order makes music files compressible.

The generation of information is easy to follow in (Western) concert music because it is usually written down at several different stages, which may be (i) motifs; (ii) their extension to melody, their transformation and development; (iii) the addition of other voices (usually in harmony or polyphony); and iv) orchestration or arranging. In formal music, this results in an orchestral score. In less formal music, analogous processes may lead to a score that is stored in one or more person's memory. In improvised music, the entire "score" may never be stored.

A motif is a characteristic phrase of several notes. The opening four notes of Beethoven's fifth symphony is an example, of which more anon. A motif is usually the origin of a musical composition. Several different pitches over a modest pitch range, and allowing for several different note durations, implies a possible information content of a few hundred bits.

Although the production of this information is difficult to study in detail, textbooks on composition give advice on producing motifs from simpler patterns. Schönberg [7], for example, gives numerous examples of how musically interesting phrases can be constructed from the three notes of a major chord by adding passing notes, repetitions, upbeats, appoggiaturas and alterations of notes. Many composers use comparable techniques to produce melodies.

The processes used by human composers are rarely written down, and are difficult to study explicitly [3]. It may seem prosaic to speculate that they are algorithms (as yet unknown) operating on aspects of the composer's background and stimuli, but to do otherwise seems to lead to Cartesian dualism. A range of explicit automata have been devised to create melodies. A famous example is the dice music attributed to Mozart, in which casting a die decides among several possible subunits. In electronic versions, a random number generator replaces the die. Further, while Mozart's subunits are musical phrases, some composition algorithms start with a scale of notes, some random input and a set of rules. Various automatic composers have thus been devised [8] since Harry Olsen created one in 1951 using rules generalised from the songs of Stephen Foster [9]. Michael Smetanin is an example of a contemporary composer who has used simple rules or algorithms to create musical compositions. It is difficult for an outsider to judge the success of such algorithms *per se*, however, because there is usually some discretionary intervention by a human at the input or output stage. In 'Strange Attractions', Smetanin [10] chose a particular algorithm because it gave melodies that he found attractive. An extreme example of choosing an algorithm and then letting nature take its course is 'White Knight and Beaver' by Martin Wesley-Smith [11], in which the composer assigns a note to each of the four bases of the DNA code, and then notates musically a section of the genome of the bacterium *E. coli*[3]. When other examples are given of tunes created by various algorithms, however, it is usually the case that only the 'best' results are presented – so human decision-making has intervened at the output stage.

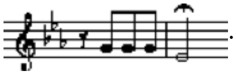
Use of a set of "rules" or fashions to generate combinations of notes and then a decision about which ones to keep is a simple model for the way some human composers work. The "rules" need not be laws (such as "the leading note always rises"[4]) decreed by some authority and observed by composers [12]. Rather they may be habits or tendencies in styles of music. For instance, virtually all composers recognise the octave as the most important and harmonious interval. Even the 'democratisation' of intervals by serialist composers leaves the octave as a very special case [13]. In this case there is a physical explanation: the harmonics of a particular note are a subset of those of the note one octave below, so adding an octave does not, or need not, add any new frequency components. In other cases, the "rules" have more complicated origins: for instance, most compos-

[2] Predictability necessarily implies redundancy. Hearing an unknown piece of tonal music from which some notes had been replaced with obvious blanks, many listeners would be able to guess the missing notes with better than chance scores, just as yo_ cou_d gues_ the _issing lette_s in this sentence.

[3] Does it sound like something that came out of a human colon, one might ask. Well, there are only four notes and they are not discordant. It sounds pleasant and musical, but this listener cannot readily extract a musical meaning.

[4] This rule shows a good example of redundancy: if the leading note were *always* followed by the note above, then an encoding could omit the pitch of the latter, just as one could omit the "u" following "q" in coding English.

ers confine themselves to scales with twelve semitones to the octave. This has a little to do with the physical basis of harmony [14], but it also has to do with what conventional instruments and players can play, what we are used to hearing, and a series of compromises among consonance and keeping the number of notes small. The "rules" for composition in most styles would be difficult to list specifically, but the musical heritage and education of the composer must incline him/her towards some patterns and combinations. Composers have a variety of processes (algorithms) for transforming an old motif into a new one, such as inverting it, changing the rhythm, reversing it, changing one or more intervals [15]. Perhaps the most important stage in producing a good motif is deciding which of many candidates is good. This process, while difficult to analyse, is at least almost universally comprehensible because many music lovers claim an ability to discern a good theme from a bad.

Thus, in one common method of composition, input data and a series of different, often unconscious algorithms generate a short phrase or idea with perhaps some tens or hundreds of bits. This may be developed into a longer melody. In written music, the data content increases in proportion with the length of the melody, but many of the extra data thus produced are redundant, in the scientific sense. The "same" motif may be repeated, transposed, inverted and otherwise transformed to create a much larger work. For one example, note the similarity in the two phrases in Fig. 2. For another, consider the famous opening phrase of Beethoven's fifth symphony: . Much is made of this simple phrase: the motif of three quavers followed by a descent of a third is used dozens of times in the beginning. Simple modifications of it occur in almost every bar of the movement: it is transposed to different positions in the scale, the final interval is changed to a second and sometimes a fourth, the last of the quavers sometimes falls, or the whole phrase is inverted in pitch. Further variants appear in the other movements – a remarkable example of much created from little.

The redundancy or structure that is created by repetition with variation is very common in melodies. In the sixteen bar 'Freude' air of Beethoven's ninth, for example, the phrase of the first four bars is repeated with slight variations in bars five to eight and thirteen to sixteen. This pattern (a,a,b,a) is extremely common, especially in songs. On a larger time-scale, redundancy through explicit repetition is so common that a variety of musical notations exist, including various repeat signs and musical 'goto' statements.

In formal music, there is often a development section in which the original idea is variously transformed: it may appear in different keys, different rhythms, inverted or melodically varied or decorated. The transformed phrase is often sufficiently different that a simple coding cannot easily reduce the length of the simplest representation. The data contained in such sections are thus created by treating the input data (the initial phrase). The existence of important structures with a variety of time scales[5] have made it difficult to formalise or to automate this operation, however. Further, selection among different algorithms and outputs is again an important process. (See the discussions in [3,16].)

Adding harmonies and counter melodies to a principal melodic line adds more data, but in some instances the extra data have relatively great redundancy. A canon is an extreme case, in which the original melody accompanies itself with a time lag, so the only extra information required is the period of the delay. In a fugue, the same or a similar melody enters with a delay, and often a symmetry operation, i.e. transposed or inverted in pitch, with doubled or halved tempo. In these cases, and in polyphony, several parallel channels of melody are of approximately equal importance. In much music however, there is one melody (or foreground) of pre-eminent importance and a harmony or accompaniment (middleground and background).

In many musical styles, the harmony is subject to rules of varying strictness, which to some extent limit the freedom of other voices and thus introduce further redundancy. Students of traditional Western harmony will agree: it often seems that the combination of strict harmony rules and voice ranges, when applied to the melody set in a harmony exercise, allow only a small number of possible 'solutions'. In many styles of music, the second most important line is the bass. If strict harmony rules are applied to a given melody and bass line, the possibilities for further parts is severely limited. Altos and tenors in choirs, or the players of second violin or viola sometimes feel that theirs are the 'left over' notes and that the result is a part that both more difficult and less satisfying than the top or bottom lines. Strict rules are extreme examples [12], but it is rare that harmony or polyphony is without rules, whether formal or informal, rigorous or fuzzy. Thus the generation of the harmony or accompaniment is often aided by the operation of algorithms on the information in the melody [18,19]. Sometimes the harmony is coded in a com-

[5] For example, the use of time-series analysis to predict the next note from the previous several notes may work well for short time scales, but is prone to wander rapidly among keys. Reviewed by Dubnov and Assayag [17].

pact but inexplicit way, such as chord symbols or figured bass. Some of its information (*e.g.* the chord symbol) is sufficiently important that the composer chooses to specify it, but the octave in which the notes occur, or their timing, is left to the performer.

Information other than notes, including articulation, ornamentation and expression marks, may be written above or below the musical staff, to convey information about pitch and duration (*e.g.* trill, staccato etc.) in ways that are more compact and legible than the explicit notation. Others carry information about loudness, articulation and tempo (*pp*, *sfz*, *accel.* etc). Others, particularly in contemporary music, contain instructions about timbre or tone colour [20]. Schönberg proposed the development of Klangfarbenmelodie (tone colour melody) in which changing patterns and structures of timbre would attain a status similar to that of changing pitch in traditional melody. Achievement of this aim might require extra data at a rate of tens or hundreds of bits per second. Some contemporary concert music contains highly specific instructions for performance, sometimes even several instructions per note. Where pitch intervals less than a semitone (microtones) are explicitly required, this is indicated by further qualification (half flat *etc.*). The requirement for slight pitch adjustments is usually implicit: many musicians do not play exactly tempered scales but, according to musical context, make fine adjustments.

One of the most important instructions about timbre is the name of the instrument that plays each part. Orchestration, the process of distributing the parts among the instruments of the orchestra, adds further information. However, there is sometimes a high redundancy when the same notes are played by different instruments.

How many data are stored in an orchestral score? Stravinsky's "The Rite of Spring" [21] provides an example of high content: it is written for a large orchestra and often the parts are relatively independent. In some sections, there are more than 40 distinct musical lines, although of course at any instant there is doubling of notes (Fig 1c). Coding just the notes of this score by sampling in time (*cf* Fig 2a) would require high transmission rates – over 100 kBaud – because of the complicated rhythms. Traditional coding (Fig 2b) is more economical, and requires only several thousand Baud[6].

So a transfer rate of up to several kBaud (equivalent to a few hundred words per second) is available to the conductor of such a work, from the score alone. Not all of this is discernible: if one player in a tutti failed to accent a note, or if the bass clarinet and second bassoon exchanged parts, this would probably pass unnoticed. When one is *not* conducting nor listening to a performance, there is no need to read a score in real time, and one may spend minutes reading carefully a single page of score, which is played in several seconds.

The performer: information input and output

Orchestral players usually read only one line, so they receive and process their written parts at rates of up to a few hundred bits per second. Other visual inputs come from the movements by other musicians, especially the conductor's baton, the leader's bow and the 'body language' of section leaders. Musicians hear the sound around them, and read the gestures and 'body language' of the conductor. This affects their processing of the written information. The interpretation of a dynamic instruction such as *forte* depends on the ensemble loudness at the time. Fine pitch adjustments depend on the prevailing pitch and harmonic context. Players also receive feedback from the interaction with the instrument of their hands, arms and mouths – but this is getting ahead of the logical order, in which the obvious next question is: how much information does the musician put out?

Some instruments have a binary digital component. In keyboard instruments, and in some percussion, the individual pitches are effectively a finite number of parallel pitch channels. In harpsichords and organs, the keys are strictly digital: a key is either depressed or not, and the player's control of the loudness of that note is binary. Bach reportedly said, disingenuously, of his organ playing: "There is nothing remarkable about it. All you have to do is hit the right notes at the right time, and the instrument plays itself" [22]. Bach, who played the viola too, would of course have known that playing a single, beautiful note on such an instrument requires much more than simply starting and stopping at the right time. The exact timing of the depressing and release of keys are analogue parameters of great importance in musical expression. In the piano, another analogue parameter is the momentum with which the hammer strikes the string. In percussion instruments, there are the complications of the position, speed and angle of the strike. Most woodwind and brass instruments have keys and valves used almost always in a binary way: either depressed or not. This does not however restrict the pitch to discrete values because pitch is also controlled by the player's lips and air pressure. In orches-

[6] The example cited is from rehearsal mark 11 in [21]. Demisemiquavers with triplets, quintuplets and septuplets at crotchet = 66 require sampling at 924 Hz. With 6 bits for pitch, the 31 parts require 172 kBaud. Using a code like Fig 2b, but with several more bits of articulation and expression marking, 200-300 notes per bar require several kBaud.

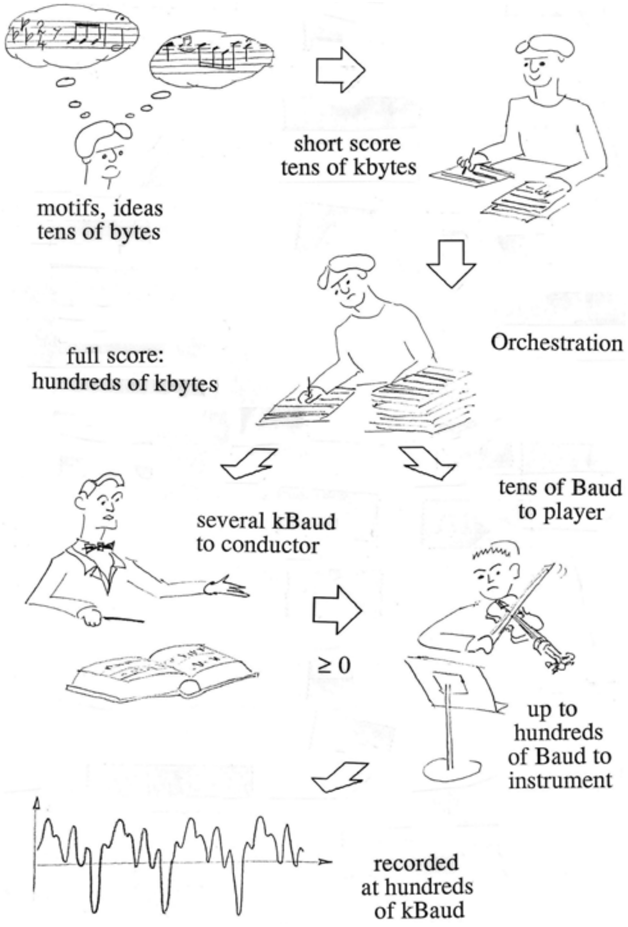


Figure 3. One information chain, from composer's original ideas to performed music. The approximate data content is given in bits and kilobytes (1 kbyte ≅ 8000 bits) and the rate of data transfer is given in Baud (1 Baud = 1 bit per second) and kiloBaud.

tral string instruments, the pitch is controlled by a continuous parameter (position of the finger stopping the string) plus choice of string.

Phrasing and expression are largely supplied by performers. Consciously or unconsciously, musicians decide how to 'shape' the phrase. This includes varying the loudness and amount of vibrato of individual notes, and making slight adjustments to indicated durations. A note judged to be important might be given emphasis by increasing the loudness and vibrato, and by increasing its duration slightly beyond the indicated value. This is one notable – and valuable! – difference between a performance by a musician and one by a primitive music sequencer. To some extent these elements of interpretation are similar among musicians [23] and so they may, to that extent, be codified. Acousticians Friberg, Sundberg and colleagues, in consultation with prominent musicians, have induced and formalised performance rules that add such elements of interpretation to a sequence representing written music [5,24,25]. Their software produces a 'performance' that is much more idiomatic and "musical" than that produced by an ordinary sequencer. These ideas have influenced modern commercial music sequencers[7].

The instrument: input and output

Written music is an incomplete instruction set. To oversimplify, the individual musician reads at typically 100 Baud or less, and outputs time-varying control signals, which may have several times this rate. The instrument outputs an analogue signal. For most monophonic instruments the output spectrum is dominated by approximately harmonic components whose fundamental frequency (which equals the harmonic spacing) determines the pitch. The pitch varies in time (with vibrato and with successive notes) and the amplitudes of the spectral components vary in time. The information required to encode this output depends on the fidelity and dynamic range required. It is at this stage that there is a great increase in the data required for encoding. If the performance is recorded uncompressed on a CD, then it results in the same enormous data transfer rate whether it be the intricate orchestration of 'The Rite of Spring' or one of the much simpler examples given above.

On many instruments, players control several interdependent analogue parameters connected with phrasing, such as vibrato, loudness, and variations in timing and intonation. Performers may also control several parameters that contribute to the timbre. In string instruments these include bow position, speed and force. In wind instruments, they include blowing pressure, several aspects of embouchure (e.g. lip tension, jaw position, position of lips on reed) and the shape of the vocal tract. These parameters may be adjusted several times per second, and each may have several bits of precision. Together they may contribute up to a few hundred Baud.

The instrument, then, is where the data rate increases dramatically. But surely the instrument is not creating information? Rather, we could say that the instrument increases the redundancy – creates redundant data – by a large factor: one period of the note is very similar to the preceding one. This oversimplifies a little: two similar hypothetically identical performances by a player – or even by a music sequencer and synthesiser – will not

[7] These typically have a range of settings for 'expressive' performance, from *meccanico* to *molto espressivo* and *molto rubato*, with varying interpretations including *straight*, *swing*, *Viennese waltz* and *funk*.

produce the same waveform, but the differences are not information to be transmitted from composer and player to listener.

TRANSMISSION AND RADIATION

In performance, instruments radiate sound into the air. These signals, plus background noise, are convoluted by the delays and multiple reflections of the performance venue. This extra information is recognised by listeners who can discern some details about the venue from listening to a recording – the difference between a cathedral and open air is an extreme example. This information contributes feedback to the conductor and players, who in general adapt their performance to the acoustic environment. For instance, they might play more quietly in a room with a low background noise and more slowly and more *marcato* in a room with a long reverberation time.

A performance creates a sound pressure field: the sound pressure p varies with position vector (**r**) and time (t). It would take a prodigious number of data to record such a field with a resolution in space and time corresponding to the half-wavelength and half-period of the highest audible frequencies (say 30 μs and 1 cm). Of course, the whole field is not sampled by a single listener, who receives just the sound pressure at each ear ($p(\mathbf{r}_1,t)$ and $p(\mathbf{r}_2,t)$), although the positions of the ears may vary in time as the listener moves his/her head. So each ear receives an analogue signal which, if the level of background noise is sufficiently low, may have the same dynamic and frequency range as the sum of the signals from the instruments.

Our imaginary composer, orchestrator, musicians, conductor and performance venue have now delivered to the ear the great data rate mentioned in the introduction. Because of the high signal redundancy, the information rate or algorithmic entropy rate is considerably lower, but still perhaps impressive. The information has been generated by mental processes of the composer and performers, which we may consider as algorithms – subtle and in many cases not understood – processing inputs from memory, education and culture. The instrument has turned this information into the radiated signal, which has been filtered and convolved by the acoustic environment. It's now time to follow the signal into the listener's head.

THE ANALYSIS OF INFORMATION

The outer and middle ear are, for our purposes, primarily acoustic and mechanical impedance transformers that overcome the mismatch between the air of the radiation field and the cochlear fluid in the inner ear. (They are also filters, transmitting some frequencies more effectively than others.) The qualitative change occurs in the cochlea of the inner ear in which the input signal – single channel analog – is actively filtered, compressed and converted to parallel digital electrical signals in the auditory nerve.

Because of the position-dependent mechanical properties of the basilar membrane, pitch is in part coded by channel: only low frequency waves reach the apical end of the membrane, so nerve fibres from this region carry information about low frequencies. It is also partly coded in rate of firing, at low frequencies at least, because the hair cells are stimulated at the frequency of the motion[8]. Signal amplitude is also partly coded by channel (some fibres only respond to large signals) and partly by (analog) signal firing rate: overall, larger stimuli produce higher firing rates. The minimum firing rate is not however zero: most neurones have a 'background firing rate' – a rate at which they fire in the absence of any signal. This makes a neuron capable of carrying a "negative" signal: if the cell is inhibited by a neighbour, its firing rate falls below the background rate. Lateral inhibition among neighbouring cells is useful in amplifying small simultaneous differences. Nerves also become less sensitive with continued stimulation, so a changing signal usually has a greater effect than a steady one. For more detail, the reader is referred to reviews of perception and neurobiology [27,28,29,30,31,32].

Coding in the auditory nerve

The pulses in the nerve fibres, called action potentials[9], are binary – either the stimulus is strong enough produce an action potential, which travels along the nerve fibre, or else nothing happens. As in electronics, the advantage of digital signals is their immunity to noise and distortion. Nerve fibres are very lossy coaxial cables, so an unamplified signal is substantially lost after transmission of a few millimetres. Many stages of amplification and pulse shaping are conducted by the nerve membrane where it is exposed at the nodes of Ranvier.

What is the data transfer rate at this stage? There are about 30,000 nerve fibres or channels, each capable of

[8] Experiments with implanted electrodes show that, at low stimulation rates, perceived pitch depends approximately logarithmically on the stimulation rate but also linearly on the electrode position [26].

[9] The voltage inside biological cells is usually tens of mV negative. When nerve cells are stimulated (by briefly making their membrane "insulation" leaky), the internal voltage rises ~100 mV before returning to the resting value.

transmitting a few hundred action potentials per second. If the coding were strictly digital, the data transfer rate would surpass that of a CD. The practical rate is much less, because of redundancy: in part because nearby fibres carry highly correlated signals. What happens to this signal in the brain is difficult to follow directly. The experimental observations of psychophysics include integration, sampling and signal treatment at higher levels.

Effects including the active filtering in the basilar membrane give rise to the masking of weak signals by strong signals in nearby frequency bands. There are only roughly 30 critical bands so, instead of 30,000 parallel frequency channels, perception effectively involves only of the order of 30. For an unmasked tone, the just noticeable difference (JND) in sound level is roughly 1 dB. Over a short term dynamic range of 60 dB, this gives about 60 perceptible loudness levels (requiring 6 bits). The JND for frequency may be as small as tenths of a percent for sustained signals, but in our calculation the maximum frequency resolution is limited over most of the range by the temporal sampling rate. The greatest perceptual resolution in time is a few tens of milliseconds. At this rate, the number of different frequency percepts is about 1000 (10 bits). So there are about 16 bits, sampled at up to 30 times per second, in 30 channels. The product gives data transmission rate of 16 kBaud: a considerable overestimate because the JNDs increase towards the ends of the frequency range and as sampling rate and number of simultaneous stimuli increases[10]. Whatever the actual maximum rate, to achieve it would require a signal that, at the perceptual level, had no redundancy or order: a signal that sounded random. Not music.

Processing – sorting into notes

It is easier to perceive notes (which usually include several or many separate frequency components) than to perceive the individual frequency components of its spectrum. With practice and careful listening, one can distinguish some spectral components in notes in some circumstances[11]. That naïve listeners rarely do so suggests that we have either a very well-learned or an inbuilt mechanism for combining the various frequency components of a note together and perceiving it as a whole. This capacity is partly explained in terms of two general properties attributed to the nervous system: that change is more noticeable than lack of change, and that things that change in the same way are often grouped together. Consider a note comprising several harmonics: if the pitch of the note changes (either melodically or due to vibrato), then the pitches of all its components change in exact proportion; if the loudness changes, then the loudness of the harmonics also changes. Evidently we possess signal processors that group these separate, but similarly changing elements together and identify them as a single note. Instrumental and operatic soloists make use of vibrato to make their notes identifiable against the sound of the orchestra[12].

The system works especially well for notes whose spectral components are approximately harmonic, which we identify as having a definite pitch. This capacity may have been important in the evolution of human audition. Many human vocal sounds (the vowels in speech, but also inarticulate cries and screams, whether sung or spoken) have at any instant a definite pitch and spectral components which fall in the harmonic series. It is likely that we have evolved hard- and soft-ware capable of identifying vocalised sounds among other sounds that do not have harmonic structure, such as wind noise. The system works so well that we hear missing fundamentals and Tartini tones.

Analysis in time

The shortest time scale of interest in music is the period of the vibration. This ranges from about 50 μs to 50 ms. For low pitches, the auditory nerve carries some information about pressure variation on this time scale, but while we are aware of pitch, we are rarely aware of the variation in pressure that gives rise to that pitch[13].

The next time scale is that of transients. When an instrument begins to play a note, there is a short time (tens of milliseconds) over which the amplitudes of the various components vary considerably before ‘settling down’ to establish a relatively unvarying spectrum. These transients are so important to the timbre of a note that different wind instruments are readily confused if the initial and final transients are removed [34]. Transients in musical notes are analogous to plosive conso-

[10] There are further complications such as feedback loops and other control signals which come “downwards” from the brain to the ear, and these affect the “upwards” signals to the brain [33].

[11] Or, conversely, a small number of harmonics may be made sufficiently louder than the rest that they can be identified as separate notes, as in harmonic singing.

[12] This effect is especially useful if some of the harmonics of the soloist occur in a frequency range where the accompanying sounds have relatively low level – if we can hear one component clearly, it seems that we can track other components which have the same vibrato and phrasing.

[13] A contrabassoon can play Bb_0 at 29 Hz. When this note is played loudly, we can just detect a periodic variation as the reed opens and closes 29 times per second. Most of the sound we hear, however, is in the higher harmonics rather than the fundamental.

nants (d, t, g, k, b, p) in speech or singing. In both cases we are capable of concentrating and hearing them with some clarity, but under most circumstances these details are analysed subconsciously.

The third time scale (several tens of milliseconds and longer) is that of notes [35,36]. It is at this level that we sense pitch and timing: the basic elements of melody. With little concentration, we can readily be conscious of the rhythm and the pitch, and also of the timbre of the instrument playing it. It is, however, difficult to introspect much beyond this: although our ears and their associated low-level processing have coded the various component frequencies and how they vary on the scale of tens of milliseconds, we are usually aware of the signal at a higher level: that of pitches, rhythms and timbres.

A changing signal is less redundant than a constant one, and our senses reflect this. After a while we no longer notice the sound of the wind, the weight of our clothes, the strange colour of artificial lighting; but we do notice sudden changes in them – changes over time. Similarly, we notice sharp boundaries in a visual image rather than a gradual change between two colours or shades – changes in space or channel. Changes in time are enhanced by the property of nerves to fire more rapidly when first excited than they do during a steady stimulus. Differences in space or channel number are enhanced by neural circuits that effectively subtract the signals from adjacent nerves using lateral inhibition [37].

Pitch sensitivity provides a good example. A single note without vibrato is a steady signal, which is probably carried at all times by the same nerve fibres. A note with vibrato is a varying signal, which is probably carried at different times by different nerve fibres. Vibrato makes notes more noticeable, and also makes it easier to identify a single instrument in an ensemble. Timing sensitivity provides another example. We are usually less conscious of the duration and end of the note than the beginning: a variation in the timing of the end of each note is noticed as a change in articulation – some notes more staccato than others; a variation in the timing of the beginning is noticed as a variation in the rhythm, and is more noticeable.

Symmetries: the ear and the instrument

In this sense, our ears and their associated low-level processing perform a role that is the reverse of that of the instrument: the player controls the note's pitch, duration and often the timbre; the instrument converts the player's partly digital, partly analogue parallel signal into a complicated vibration, or equivalently a set of simple (usually harmonic) vibrations in a mechanical oscillator (string or air column). These vibrations, often via an impedance transformer (bridge and body of string instruments, bells of brass instruments) cause a pressure wave that is a single analogue signal: p(t).

The ear receives a wave p'(t) and, via impedance transformers (the outer and middle ear) this causes a complicated vibration, or equivalently a set of simple, often harmonic, vibrations in a mechanical oscillator (the basilar membrane). These vibrations are sensed and processed, and we perceive the note's timing, pitch, duration and timbre.

The perception of notes is subject to categorisation (*i.e.* digitisation): when fine differences in pitch are presented, listeners, especially those with musical training, tend to sort them into the discrete notes in a scale [38]. Thus the perception of pitch is partly digital and partly analogue – we perceive a note, but may remark that it was a little sharp or flat.

More symmetries: the listener and the composer

On time-scales larger than those discussed above, listeners are capable of perceiving structures and features in music: we may identify (whether consciously or otherwise) themes, harmonies, orchestration etc. This article gives no more than some pointers to research in this area. Sloboda [3] compares the analysis of linguistic structure by Chomsky with the analysis of musical structure by Schenker, which uses hierarchies of note groupings and their functions. Some seem general, while others are specific to certain cultures. One way of studying this level of structure is by proposing plausible models and comparing their performance with that of human subjects [39-41].

These processes complete the communication symmetry. To the extent that the listener hears melodic patterns, repeats and transformations of thematic material, s/he reverses the process of composition and may leave the concert hall humming the themes or ideas that began the whole process.

The information transmitted between the minds of composer and listener may differ in detail, but the coding is physiologically similar in the two minds, in that it involves many parallel digital signals in neurones. Between the two, however, the information passes through a coding totally foreign to the operation of the brain – a data-rich, serial, analogue signal. The interpreters for this foreign signal are the musical instrument in one direction and the ear in the other, whose symmetry is discussed above. The performing musicians direct and supervise translation at one end. The listener has an interpretive role that may be the reverse of those of player and composer, depending on training and attitude. A discussion of this is beyond our current aim.

MUSICAL COMMUNICATION

To a communications engineer, music might seem inefficient and unreliable. Different listeners may extract different messages from the same signal. Listeners may differ with the composer over the question "what is it about?" This does not mean, of course, that it is without meaning or value: the signal is rich in information often input by different people (composer, performers, conductor) so it is not surprising that different people extract different subsets of that information, or interpret it differently. To quote Aaron Copland: "'Is there a meaning to music?' My answer to that would be, 'Yes'. And 'Can you state in so many words what the meaning is?' My answer to that would be, 'No'." [42]. Researchers are however quantifying aspects of the meaning. Schubert [43], for instance, measures emotional responses to music in a two-parameter space and finds reasonably consistent responses, with a resolution of a few bits in each direction and a time resolution of seconds. This gives a Baud rate not far below that of text being read.

In the context of musical enjoyment, the processes of encoding and decoding may be at least as important as any part of the communication. But why do we so enjoy this encoding and decoding? Why have we evolved the capacity for this sophisticated, complicated but imprecise method of communication of abstract ideas? Does musical ability confer survival advantages on individuals possessing it? Why can such abstract communication have powerful emotional effects? These questions are invitations for speculation, but it is interesting to look at them with regard to information coding.

Music and speech: similarities and complementarities

The physiological hardware used for listening to music and speech is the same, and some of the software may be shared too. Most speech sounds involve vibration of the vocal folds. The time scale of these vibrations is shorter than that of nerve or muscle response, so any given vibration is very similar to its predecessor, so the sound produced is usually quasi-periodic. These periodic speech sounds (as well as screams, cries, and moans) have harmonic spectra. The ability to discern a set of harmonic frequency components as an entity, and to track simultaneous changes in that set, is an ability to discern one voice or cry from background sound. It is also much of the ability to follow a melody.

On the other hand, the signal codings of speech and music are different. Oversimplifying for the sake of the argument, we could say that they are almost complementary, especially with regard to digitisation. Speech coding is digital in that it uses a discrete set of speech sounds (phonemes). In alphabetic languages, (a subset of) these are all that is recorded, as letters, in the text or transcribed form. Further, they are digitised in perception (*i.e.* they are perceived categorically [44]). Phonemes are encoded by features of the sound spectrum (formants and formant trajectories) and by transients. But in music, transients (especially the way notes start) and features of the spectrum are together what we call timbre. Most of the 'text' of music is notes: digital representation of pitch and timing. These are also perceived digitally (categorically) in music [38]. In speech, however, these features are prosody and (except in tonal languages such as Mandarin and Thai) they are analog variables, which are not notated. So the texts of music and speech use the acoustical features and digitisation in almost complementary ways, as the table shows. I discuss this in greater detail elsewhere [45].

Why music?

The capacity to communicate using sound, whether by speech or more primitive articulations, may have been sufficiently important to select for a suitable capacity for sound analysis. This explains (at the evolutionary level) why we have the mechanisms that we use for analysing music. But why do we so use those mechanisms? Why do parents sing to infants? Why do we like and make music? Perhaps signal processing can provide part of the answer.

Those who write or use automatic speech recognition software know that it is non-trivial to extract the spectral features, envelope and pitch that carry information in both speech and music, especially in the presence of background noise. In some cases, however, it may be easier in music. Consider an unaccompanied melody, sung or played by a single instrument, which might be an example of music from our early pre-history. This signal has frequencies that are usually stable during a note, compared with the rapid, continuous (*i.e.* analogue) pitch changes in speech. Rhythms in music are also more regular in music than in speech. In instrumental music and in *vocalise* (singing without words), the spectral features change less, and in a more regular way, than they do speech. When we sing to babies [46], is it possible that we are using the reduction-

Table 1. Acoustical features of music and speech signals show complementary coding. (Reproduced from [45]).

Acoustical feature	Music without words	Speech
Fundamental frequency (when quasi periodic)	*pitch component of melody*	*pitch component of prosody*
	categorised	not categorised
	notated	not notated
	precision possible	variability common
Temporal regularities and quantisation on a longer time scale	*rhythmic component of melody*	*rhythmic component of prosody*
	categorised	not categorised
	notated	not notated
	precision possible	variability common
Short silences	*articulation*	*parts of plosive phonemes*
	sometimes notated	implicitly notated
Steady formants	*components of instrumental timbre*	*components of sustained phonemes*
	not notated	notated
	not categorised *per se*	categorised
Varying formants	*not widely used*	*components of plosive phonemes*
	—	categorised
		notated
Transient spectral details	*components of timbre*	*components of consonants*
	not categorised	categorised
	sometimes notated	notated

ist method to teach them how to listen, developing the skills necessary to understand speech?

Could music be a game for the ear? Games are often described as models of social behaviour, that develop useful mental and physical skills. Games develop reflexes, co-ordination and muscular strength that may confer evolutionary advantages. Intellectual and socialising games develop skills that could also confer survival or mating advantages. If speech and signal processing skills enhanced our ancestors' chances of survival or mating, the game of music may have been selected, whether it were transferred between generations by genetics or culture.

The basic skills of sound analysis are subtle and beyond introspection, but that is true of many games: we are no more conscious of how we analyse sounds than we are of the muscular control we used to catch a ball. What we do with these skills is sometimes elaborate, but that is also true of games such as cricket and chess. In games and in music, our enjoyment of neurological exercise and challenges seems to require successively more complicated games as our capacities develop.

Speech carries the meaning of the words spoken, but it also carries information in the way in which the words are spoken. The rhythms and tempi, subtle pauses and variations in articulation and loudness, the overall register and the changing pitch – all carry information. Information of this latter type gives subtle shades to the meaning conveyed by the words, and it often tells of the speaker's emotional state. The ability to convey this information distinguishes a good actor from someone who just reads the words. Music also carries expressive information in subtle variations in rhythm and phrasing [24,47], coded in a comparable way [48].

However, an important vehicle for affective information in speech is prosody. These features, completely omitted in the text of speech, are the dominant features of music, whereas the features used to encode the explicit information in speech are used, in music, for timbre and are often varied little. I end by inviting the reader to wonder, as I do, whether this may one of the reasons for the attraction and emotional power of music, this peculiarly coded, abstract method of communication.

ACKNOWLEDGMENT

I thank Sten Ternström and Emery Schubert for helpful comments.

REFERENCES

1. Kolmogorov, A.N. "Three approaches to the definition of the concept "amount of information"." (1965) Problemy Peredachi Informatsii, **1**, 3-11 (Russian, cited by Chaitin, *ibid.*)

2. Chaitin, G.J. “Information, Randomness and Incompleteness”. (World Science, Singapore, 1987).
3. Sloboda, J. “The Musical Mind” Oxford: Clarendon Press, (1985).
4. Garnett, G. “Music, Signals, and Representations: a Survey” in “Representations of Musical Signals”, de Poli, Piccialii and Roads, eds., (MIT Press, Cambridge Mass, 1991)
5. Sundberg, J. “Musical performance: a synthesis-by-rule approach”. Computer Music J. **7**, 37-43 (1987).
6. Friberg, A. “Generative rules for music performance: a formal description of a rule system”. Computer Music J., **15**, 49-55 (1991).
7. Schönberg, A. “Fundamentals of Musical Composition” (Faber, London, 1967).
8. Schwanauer, S.M. and Levitt, D.A. eds. “Machine Models of Music”, (MIT Press, Cambridge MA, 1993).
9. Cope, D. “Experiments in musical intelligence (EMI): non-linear linguistic-based composition”, Interface, **18**, 117-139 (1989).
10. Smetanin, M. “Strange Attractions”, (Sounds Australian, Sydney, 1990).
11. Wesley-Smith, M. “White Knight and Beaver”. (Sounds Australian, Sydney, 1984).
12. Masson, C. “Nouveau Traité des Règles pour la Composition de la Musique” (1705). (Facsimile edition, Minkoff, Geneva, 1971).
13. Leibowitz, R. “Introduction à la musique de douze sons”. (L’Arche, Paris, 1949)
14. Helmholtz, H.L.F. “On the Sensations of Tone as a Physiological Basis for the Theory of Music”, (1877) English translation by A.J. Ellis, (Dover, N.Y. 1954).
15. Stravinsky, I “The Rite of Spring: sketches 1911-1913. Facsimile reproductions with commentary by R. Craft”. (Boosey & Hawkes, London, 1969).
16. Dirst, M. and Weigend, A.S. “Baroque forecasting: on completing J.S. Bach’s last fugue”, in “Time Series Prediction: Forecasting the Future and Understanding the Past” A.S. Weigend and N.A. Gershenfeld, eds. (Addison-Wesley, Reading MA 1993).
17. Dubnov, S. and Assayag, G. “Universal pediction applied to stylistic music generation”, in “Mathematics and Music” G.Assayag, H.G.Feichtinger and J.F.Rodrigues, eds. (Springer, Berlin, 2002).
18. Maxwell, H.J. “An expert system for harmonizing analysis of tonal music” in “Understanding Music with AI: Perspectives on Music Cognition” M. Balaban, K. Ebcioglu and O.Laske, eds.pp 335-353. (MIT Press, Cambridge, MA, 1992).
19. Hild, H., Feulner, J. and Menzel, W. “HARMONET: a neural net for harmonizing chorals in the style of J.S. Bach” in “Advances in Neural Information Proceessing Systems, J.E. Moody, S.J. Hanso and R.P. Lippman, eds, 4:267-274. Morgan Kauffman, San Mateao, CA (1992).
20. Stone, K. “Music Notation in the Twentieth Century”. (Norton, New York, 1980).
21. Stravinsky, I. “The Rite of Spring” (1921). The example cited is from rehearsal mark 11. Boosey & Hawkes, London (1967).
22. Köhler, J.F. “Historia Scholarum Lipsiensium” (1776), quoted by David, H.T. and Mendel, A. “The Bach Reader”, (Norton, NY, 1972)
23. Repp, B.H. “A constraint on the expressive timing of a melodic gesture: evidence from performance and aesthetic judgment”, Music Perception, **10**, 22-242 (1992).
24. Sundberg, J. Fribert, A. and Fryden, L. “Threshold and preference quantities of rules for music performance”, Music Perception, **9**, 71-92 (1991).
25. Juslin, P.N; Friberg, A., Bresin, R. “Toward a computational model of expression in music performance: The GERM model.” Musicae Scientiae. Spec Issue, 2001-2002, 63-122 (2002).
26. Fearn, R., Carter, P. and Wolfe, J. “The perception of pitch by users of cochlear implants: possible significance for rate and place theories of pitch” Acoustics Australia, 27, 41-43 (1999).
27. Barlow, H.B. in “Physics and mathematics of the nervous system” (Conrad, M, Güttinger, W. and Dal Cin, M., eds) (Springer-Verlag, Berlin, 1974).
28. Møller, A.R. “Auditory Physiology”. (Academic, NY, 1983).
29. Fletcher, N.H. “The physical bases of perception”, Interdisciplinary Sci. Rev., **9**, 6-13 (1984)
30. Kandel, E.R. and Schwartz, J.H. Principles of Neural Science, (Elsevier, 1985).
31. Altschuler, R.A., Bobbin, R.P., Clopton, B.M. and Hoffman, D.W. “Neurobiology of Hearing: the Central Auditory System”. (Raven, NY, 1991).
32. Yates, G.K. “The Ear as an Acoustical Transducer”, Acoustics Australia, **21**, 77-81 (1993).
33. Spangler, K.M. and Warr, W.B. “The descending auditory system” in “Neurobiology of Hearing” R.A. Altschuler et al, eds, pp 27-45, (Raven, NY, 1991).
34. Berger, K.W. Some factors in the recognition of timbre, J. Acoust. Soc. Am. **36**, 1888 (1963).
35. Warren, R.M., Gardner, D.A., Brubaker, B.S. and Bashford, J.A. “Melodic and nonmelodic sequences of tones: effects of duration on perception”, Music Perception, **8**, 277-290 (1991).
36. Warren, R.M. “La perception des séquences acoustiques: intégration globale ou résolution tempo-

relle?" *in* "Penser les Sons. Psychologie Cognitive de l'Audition" McAdams, S. and Bigand E., eds., Presses Universitaires de France (1994).

37. Shepard, G.M. "Neurobiology", (Oxford Uni. Press, 1988).
38. Locke, S. and Kellar, L. "Categorical perception in a non-linguistic mode" Cortex, **9**, 355-369 (1973).
39. Lischka, C. "Understanding Music Cognition: A Connectionist View" in "Representations of Musical Signals", de Poli, Piccialii and Roads, eds., (MIT, Cambridge Mass, 1991).
40. Longuet-Higgins, H.C. "Artificial intelligence and musical cognition", Phil. Trans. R. Soc. Lond. A **349**, 103-113 (1994).
41. Longuet-Higgins, H.C. and Lisle, E.R. "Modelling musical cognition", Contemporary Music Review, **3**, 15-27 (1989).
42. Copland, A., "What to listen for in music". (New American Library, NY, 1967).
43. Schubert, E. "Continuous measurement of self-report emotional response to music" in "Music and emotion: Theory and research. Series in affective science." Juslin, P.N. (Ed); Sloboda, J.A. (Ed), eds. pp 393-414. (Oxford University Press, London, 2000).
44. Clark, J. and Yallop, C. "An Introduction to Phonetics and Phonology" (Blackwell, Oxford, 1990).
45. Wolfe, J. "Speech and music, acoustics and coding, and what music might be 'for'". International Conference on Music Perception and Cognition, Sydney, 2002, K Stevens, D. Burnham, G. McPherson, E. Schubert, J. Renwick, eds. pp 10-13 (2002). www.phys.unsw.edu.au/~jw/ICMPC.pdf
46. Gérard, C and Auxiette, C. "The processing of musical prosody by musical and nonmusical children" Music perception, **10**, 93-126 (1992).
47. Mersenne, M. "Harmonie Universelle, contenant la Théorie et la Pratique de la Musique" (1636). (Facsimile edition, CNRS, Paris, 1975).
48. Banse, R. and Scherer, K.R. "Acoustic profiles in vocal emotion and expression" J. Personality and Social Psychology, **70**, 614-636 (1996).