Journal of Moscow Physical Society, v. **3**, pp. 331–353 (1993). Printed in the UK.

A Mathematical Model of Scale Perception

L. V. Avdeev† and P. B. Ivanov‡§

†Bogolyubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna, Moscow Region, 141980, Russian Federation

‡Computational and Theoretical Division, Troitsk Institute for Innovation and Fusion Research, Troitsk, Moscow Region, 142092, Russian Federation

Received 2 April 1994

Abstract. A hierarchical approach to scale formation in human perception is applied to musical scales. The model provides an adequate mathematical description of the already known scales and reveals some other new possibilities, in particular, a universal 19-tone musical system.

A formula for the information difference between two probability distributions is employed to construct a numerical estimate of a contradiction between two compound tones. The *discordance* function obtained in this way possesses a number of minima which correspond to the degrees of a musical scale. A *dissonance* function is introduced, which reveals the scale as a set of zones. *Stationarity* under nonlinear transformations and maximum *regularity* provide the numerical criteria for selecting the preferable scales. The historical development from simple to ever more complex scales is thus traced. The *formant* structure of the internal timbre pattern characterizes the stability of local hierarchical structures. Harmonic, modal, and chromatic types of scale *lability* are described. Musical scale as a movable hierarchy of zone structures unfolds itself in various ways, forming the musical *context*. The analysis of the discordance functions indicates the ways of releasing the tensions and shows the musical consequences of any melodic move.

1. Introduction and general outline

A hierarchical model of scale formation in the European musical tradition developed hereafter is based on the following facts.

1. There is no direct correlation between physical properties of sound and the perceived intonation [1, 2]. Man rather constructs an internal model of external sound and then tries to fit all what he hears into the present pattern. That is why we speak about sound *perception*, which assumes sound sensation as well as its internal representation.

2. Man can never exactly define the pitch of a sound, and real toning bears *zone* character [3, 4]. Accordingly, tone perception should reflect this diffusiveness of perceived sounds.

3. Historically, simple distinguishing of differing sounds precedes, and only after a time the notion of interval is established [5]. The fifth and the octave take their place in music rather late, so one hardly can seriously speak about combinatorial origin of musical modes.

4. Globally, musical hearing evolves in the direction of distinguishing higher and higher overtones, and always more detailed perception of timbre [6].

5. Any human activity is *hierarchically* organized [7–9], including the perceptual activity.

6. Such musical phenomena as consonance and dissonance, tension, instability and steadiness arise only in a specific *context* [6]. This is why it is important for a theory to have some formalized concept of musical context.

---

§ E-mail: *unism@ya.ru*

We follow here D. Marr's approach [10], assuming that a detailed investigation of the mechanisms of perception gives little for the comprehension of perception itself. Rather, construction of higher-level models permits one to bind physiological data together. One can notice the ever growing penetration of *information* theory into psychophysics of perception [7, 8, 11]. To be sure, we have always been taking into account the available physiological and psychological data on hearing. Rather comprehensive overviews can be found in [1, 2].

In general features, a model incorporating all these considerations is developed like this.

The lowest level of scale perception comprises musical *tones* (abstracted from noises, phase effects, etc, which are unessential for the theory of scaling). Any tone is represented by its harmonic series, that is, the set of partials, any one of which is supplied with an amplitude $t_n (n = 1 \ldots N)$. The set of amplitudes will be called the internal *timbre*. A partial is described by a density *distribution* with the dispersion σ. Since N, t_n, and σ give the internal representation of tone, they can be assumed the same for all tones and considered as the parameters of the current perceptional tuning.

In the process of hearing, man establishes relations between tones. As our model deals with probability distributions, we derive in section 2 the formula for the quantity of information in one probability distribution relative to another. This rule is applied in section 3 to a pair of isolated partials of musical tones, to obtain a 'discordance value', characterizing the contradiction of the two sounds, the difficulty of binding them in a common system of tones. Assuming that compound tones 'interact' through the pair relations between their partials and proportionally to the amplitudes of the partials, we define the discordance for two complex tones dependent on the difference of the pitches of their ground harmonics (section 4). When a reference tone is fixed, one obtains a *discordance function*, the discordance distribution related to the reference tone.

On the next level, various tone structures formed of the minimally discordant tones are perceived. Since not all timbres are of equal worth for the distinct structure perception, we seek the parameters indicating the best scales. The demand of the minimal timbre distortion (*stationarity*) under nonlinear transformations accompanying any image processing in human brain leads to the τ criterion (section 5). The tuning of perception to a leading 'rhythm' in the discordance function, arising from *regularity* requirements, forms internal timbre patterns which we call *optimal* timbres (section 6). Most suitable for scale formation are both optimal and stationary timbres.

Section 7 shows how the relation of a discordance function to some locally defined average level generates a set of *zones*, which we associate with the scale. The zone character of musical scaling is thus accounted for.

Computer-aided calculations permit us to find a series of optimal stationary timbres. The properties of some corresponding scales (for example, with 2, 3, 5, 7, and 12 tones to the octave) are well known in musicology. We can therefore understand how a timbre pattern correlates with the scale properties, and with a good certainty describe some new scales, hardly yet established in musical practice. In particular, the concept of scale *stability* is introduced, and the modal, harmonic, and chromatic types of lability are described (section 8). Stable stationary regular scales are most fit for universal application. The well-known example is the 12-tone scale. Another universal scale, the 19-tone one, often appears in theoretical considerations [12–16], though only our model suggests adequate notions for studying basic features of 19-tone music.

Even in the environment of a fixed optimal scale, one can perceive its various *substructures*, tuning perception to other 'rhythms' (quasiperiodic components) present in the discordance function by correspondingly modifying the dispersion σ. The hierarchical system of scale imbeddings thus arising (usually including the local mode, and the local harmony) represents the scale *context* and permits one to naturally pass from the scale level to higher levels of musical perception, such as melodics, tonality, harmony (section 9).

Thus, our model opens a new view upon all the variety of musical phenomena related to musical scales, clarifies some musical regularities, and reveals new prospects of evolving the sound base of music.

2. The informational measure of a change in the distribution

Suppose we have a random variable x taking on values in an interval $a \le x < b$ with a smooth distribution $f_0(x)$. We are interested in the quantity of information which is obtained if, as a result of measurements, the distribution is found to become another function $f_1(x)$. According to the classical definition, information I communicated by an event is expressed through the *a priory* probability p of the event:

$$I = -\log p \qquad 0 < p \le 1 \tag{2.1}$$

To know a distribution, we should perform a large number n of independent measurements of our random quantity. We divide the interval $[a, b)$ into m small subintervals

$$x_j - d_j/2 \le x < x_j + d_j/2 \qquad (j = 1\ldots m) \tag{2.2}$$

$$a = x_1 - d_1/2 \qquad b = x_m + d_m/2 \ .$$

Let the number of the x values that occurred in the jth subinterval (2.2) after n tests be n_j. This situation corresponds to observing the *a posteriori* distribution f_1 to be

$$f_1(x_j)d_j = n_j/n \tag{2.3}$$

up to $O(d_j^2)$ corrections, with the natural normalization condition

$$\sum_{j=1}^{m} f_1(x_j)d_j = 1 \ . \tag{2.4}$$

We now compute the probability of such a set of elementary events $\{n_j\}^m$, using the *a priory* distribution $f_0(x)$. It is

$$p = n!\prod_{j=1}^{m}\left\{\left[f_0(x_j)d_j\right]^{n_j}\big/ n_j!\right\} . \tag{2.5}$$

Substituting equation (2.5) into equation (2.1) permits us to compute the information

$$I = \sum_{j=1}^{m}\left\{\log(n_j!) - n_j\log\left[f_0(x_j)d_j\right]\right\} - \log(n!) \ . \tag{2.6}$$

We are going to take a continuum limit $m \to \infty$, $d_j \to 0$. However, to get a sensible result for the distribution, we should tend n to infinity much faster so as to ensure that each $n_j \to \infty$. Therefore, we set

$$m = O(n^{1/2}) \qquad d_j = O(n^{-1/2}) \qquad n_j = O(n^{1/2}) \qquad n \to \infty . \tag{2.7}$$

Now we can apply Stirling's formula for the factorials

$$n! = \sqrt{2\pi n}\; n^n \,\exp(-n)\left[1 + O(1/n)\right] \; . \tag{2.8}$$

Using equations (2.8), (2.3), and (2.4), we rewrite equation (2.6) with the $O(1)$ accuracy as

$$I = n\sum_{j=1}^{m} f_1(x_j) d_j \log\left[f_1(x)/f_0(x)\right] - \frac{1}{2}\log(2\pi n) + \frac{1}{2}\sum_{j=1}^{m}\log(2\pi n_j) \; . \tag{2.9}$$

In the limit (2.7) the last sum in equation (2.9) behaves at most as $O(\sqrt{n}\log n)$, hence, the last two terms become negligible as compared to the first sum. Now, replacing the sum with the integral it approaches the limit of, we arrive at the final formula

$$I/n \to \Delta(f_1|f_0) = \int_a^b dx\; f_1(x)\log\left[f_1(x)/f_0(x)\right]. \tag{2.10}$$

In a somewhat less rigorous way, expression (2.10) has earlier been obtained independently in [17].

It is not unexpected that the total information is infinite; however, the specific information per measurement has a finite limit (2.10). We see that $\Delta(f_0|f_0) = 0$: if the distribution remains unchanged, then, to the leading order, $I = 0$, that is, the probability to obtain $f_1 = f_0$ is relatively not very far from 1. In any other case, as it follows from equation (2.1), $I > 0$, hence, $\Delta(f_1|f_0) \geq 0$. It can also be directly checked that the specific-information functional (2.10) has a local minimum at $f_1 = f_0$ under the restrictions

$$\int_a^b dx\; f(x) = 1 \qquad f_j(x)>0 \qquad (j=0,1)\, . \tag{2.11}$$

If we deal with the particular case of the Gaussian distributions

$$f_j(x) = (2\pi)^{-1/2}\sigma_j^{-1}\exp\left[-\frac{1}{2}(x-h_j)^2/\sigma_j^2\right] \tag{2.12}$$

then the specific information is given by

$$\Delta(f_1|f_0) = \frac{1}{2}(h_1-h_0)^2/\sigma_0^2 - \log(\sigma_1/\sigma_0) + \frac{1}{2}(\sigma_1^2/\sigma_0^2 - 1) \tag{2.13}$$

which corrects Golitsyn's formula [11] where the last term was absent. Our result (2.13) is always positive as it should be unless $h_1 = h_0$ and $\sigma_1 = \sigma_0$.

Formulas (2.10) and (2.13) may be taken as a rigorous basis for wide general speculations concerning the philosophy of esthetic perception [7, 11, 18]. We, however, leave the subject, to proceed further with our model of musical scales.

3. The discordance value for a pair of partials

In this section we consider an interaction between two elementary stimuli, which results in an informational measure of their mutual influence and contradiction. In our case, elementary stimuli are the partials of musical tones. They are represented by probability distributions in a perceptual space of pitch. This presentation is formed on lower levels; we are not going to enter into details here.

The pitch is measured by the logarithm of the frequency of a partial: logarithmic scales appear already on the sensation level. It is convenient to choose the *binary* logarithm so that the octave interval corresponded to the unit pitch difference.

We describe the distributions by the Gaussian curves (2.12) because, according to a theorem of the probability theory, this is the limit distribution for a sum of a sufficiently large number of any small fluctuations. Thus, it is the distribution that is most likely formed on the lower levels.

The center of the distribution h_j — the position of its maximum — specifies the pitch localization of a partial, and the *dispersion* σ determines diffusiveness of the perceived image. One should not mix up this diffusiveness with the physical dispersion of the external sound or with the physiological limit of distinguishing different pitches δ. Depending on the purpose, the perception may be tuned to any rougher scale $\sigma > \delta$. The dispersion σ is determined by a particular level of perception (Section 6 and further), and we assume it to be the same on that level for all the partials.

In describing the interaction of elementary stimuli we follow Golitsyn's ideas [11]. The two stimuli play different rôles: one — $f_0\,(x)$ with the center h_0, 'old' — is a reference point; the other — $f_1\,(x)$ with the centre $h_1 = h_0 + R$, 'new' — is compared to it. At first, a mixture of the stimuli is formed

$$F(x) = (1-\mu) f_0(x) + \mu f_1(x) \tag{3.1}$$

with a small weight $\mu \ll 1$. As a result of the mixing, the maximum of the distribution (3.1) shifts with respect to the old center by a distance

$$r = \mu R \exp\left(-\frac{1}{2} R^2/\sigma^2\right) + O(\mu^2) \tag{3.2}$$

to the point $h = h_0 + r$. Man tries to represent the complex signal (3.1) in a uniform manner (2.12). The simplest way to achieve this is to follow the shift (3.2) of the center, obtaining as a result the standard Gaussian distribution $f(x)$ with a new center at h and the same dispersion σ as for all the partials. The informational measure of the novelty of f as compared to f_0 is calculated through formula (2.13):

$$\Delta(f|f_0) = \frac{1}{2} r^2/\sigma^2 = \frac{1}{2}\mu^2 (R^2/\sigma^2) \exp\left(-R^2/\sigma^2\right) + O(\mu^3)\,. \tag{3.3}$$

Thus, the information that corresponds to the change of the old image due to the appearance of a new stimulus is determined by

$$d(R) = (R^2/\sigma^2) \exp\left(-R^2/\sigma^2\right) \tag{3.4}$$

the mixing parameter μ giving only an unessential overall factor. Quantity (3.4) is called the *discordance value* for the two partials. It has a maximum at $R = \sigma$, when the distance between the stimuli is equal to their dispersion. Trivial ($R \ll \sigma$) or too far-away signals ($R \gg \sigma$) convey no information. Function (3.4) is even, $d(-R) = d(R)$, therefore, it makes no difference which of the partials has been selected as old and which as new: the discordance effect is mutual.

4. The discordance function for compound tones

The strongest assumption we make is that of *additivity*: the discordance for a complex signal with a discrete spectrum is given by a sum of the discordance values for all the pairs of its distinct harmonics proportionally to their amplitudes. Such a linearity is connected with a certain selection of the perception level [1]. It is supported by an analogy with quantum mechanics: the discordance can be considered as an expectation value (on a superposition of partials) of an operator defined by its matrix elements between elementary tones.

We should point out a restriction which is necessary for consistency of the additivity hypothesis: either individual partials must be physiologically well-discernible $R \gg \delta$, or their mutual discordance $d(R)$ ought to be negligible, so that the effect would be equivalent to one harmonic of the sum amplitude. This leads to the condition $\delta \ll \sigma$: the dispersion that we deal with can never approach the physiological limit.

Unlike noises, musical tones possess a natural harmonic series of partials with frequencies $n\eta$, $(n = 1 \ldots N)$, the frequency of the ground harmonic being η. The internal image of a tone includes a set of amplitudes $\{t_n\}^N$ for N perceived partials. We call this set the internal *timbre*. It should by no means be directly identified with the physical spectrum of the external sound, which is drastically transformed in the sensory activity [1, 19]. Particular values of N and $t_n \geq 0$ (controlled by higher levels of perception) are fixed for all tones on the chosen level. We ignore boundary effects of extremely high or low pitch and do not involve phases of the partials. Any inharmonic overtones can be treated in the same way; however, the proportional partials are most important to study.

Consider an interval h comprising two compound tones which have different ground frequencies $\eta_1 = 2^h \eta_0$. According to the additivity hypothesis, the discordance for this complex signal consists of two contributions. One is due to the mutual discordance of the partials of different tones

$$\Delta(h) = \sum_{m,n=1}^{N} t_m t_n d[h + \log_2(m/n)]\,. \tag{4.1}$$

Another relates to the discordance of each tone's partials between themselves. The latter contribution does not depend on the size of the interval and equals $\Delta(0)$. Thus, quantity (4.1), called the *discordance function*, accumulates the total information about the discordance of compound tones.

Function (4.1) gives a primary description of a musical scale. As a timbre $\{t_n\}^N$, a dispersion σ, and a reference point (key) are fixed, the $\Delta(h)$ minima determine a set of tones — degrees of a scale — which form a congruous unity with respect to the key. On the other hand, occurrence of tones near the maxima of the discordance function may indicate that the unity becomes inadequate. This may give rise to higher-level processes of resetting the scale — modifying its origin, dispersion, or timbre — to adapt it to the situation.

In principle, on the level considered here, parameters $\{t_n\}^N$ and $\sigma \gg \delta$ may be arbitrary. We could try various combinations, selecting the most suitable ones to describe the known musical phenomena. In fact, we have done this. However, there are theoretical higher-level criteria to pick out more preferable, stationary values which reflect natural tendencies of the perceptional tuning. The following exposition may be considered as introducing higher-layer parameters in the sense of [20].

5. A criterion of timbre stationarity

Human brain is known to be a *nonlinear* system. In data processing, the perceived signal is subject to multiple nonlinear transformations [1, 2]. As a result, for instance, combination frequencies are heard, which are sums and differences of the physical-spectrum lines. These facts force us to assume that only those timbres may be fundamental which are not easily destroyed by nonlinear transformations, are robust enough, insensitive, immune to nonlinear attacks.

With an internal timbre $\{t_n\}^N$ we associate a periodic (in time T) signal

$$F(T) = \sum_{n=1}^{N} t_n \cos(2\pi n\eta T) . \tag{5.1}$$

The ground frequency η may be arbitrary here. The essential in formula (5.1) of setting the phases for all harmonics to zero is discussed below. Consider the simplest kind of a nonlinear transformation, the quadratic one Q , which creates a new signal by the rule

$$F(T) \xrightarrow{Q} F^2(T) . \tag{5.2}$$

The decomposition of the squared signal into partials

$$F^2(T) = \sum_{n=0}^{2N} q_n \cos(2\pi n\eta T) \tag{5.3}$$

involves combination frequencies with the amplitudes

$$q_n = \frac{1}{2}\sum_{m=1}^{n-1} t_m t_{n-m} + \sum_{m=1}^{N-n} t_m t_{n+m} \qquad (n = 1\ldots 2N) \tag{5.4}$$

the constant partial of equation (5.3) $q_0 = \frac{1}{2}\sum_{m=1}^{N} t_m^2$ measuring the intensity of the primary signal (5.1).

The ‘closer’ is $\{q_n\}$ to $\{t_n\}$, the more stationary is the timbre. To compare timbres, we use the following quantity

$$\tau = \left(\sum_{n=1}^{N} q_n t_n\right)^2 \left(\sum_{n=1}^{2N} q_n^2\right)^{-1} \left(\sum_{n=1}^{N} t_n^2\right)^{-1} . \tag{5.5}$$

It has the meaning of the squared cosine of the angle between q and t in an (infinite-dimensional) vector space with the scalar product

$$(q,t) = \sum_{n=1}^{\infty} q_n t_n . \tag{5.6}$$

In other words, the τ criterion (5.5) measures the relative intensity of that part of vector q, which is proportional to t, the other part — the distortion — being orthogonal. In the theory of pitch-class sets an analogous criterion was used to compare set classes [21].

For any $\{t_n\}^N$, always $0 \le \tau < 1$. If $\tau = 0$, the transformed timbre $\{q_n\}$ has nothing in common with the original pattern $\{t_n\}$. In contrast, when τ approaches 1, the quadratic transformation (5.2) creates a timbre almost identical to the original. Moreover, if there were no distortion at all, τ=1, the timbre would be immune to an arbitrary-degree nonlinearity rather than the quadratic alone. Timbres with high τ are called *stationary*.

Table 1. The τ^N timbres.

τ^7	τ^9	τ^{12}	τ^{15}	τ^{19}	τ^{25}	τ^{30}	
0.9037	0.9162	0.9259	0.9313	0.9355	0.9391	0.9409	τ
0.5478	0.4884	0.4267	0.3836	0.3422	0.2995	0.274	t_1
0.5042	0.4629	0.4131	0.3754	0.3375	0.297	0.2724	t_2
0.4395	0.4238	0.3919	0.3624	0.3299	0.293	0.2698	t_3
0.3603	0.374	0.364	0.3451	0.3196	0.2875	0.2662	t_4
0.2741	0.3166	0.3306	0.3239	0.3069	0.2806	0.2616	t_5
0.1884	0.2554	0.2931	0.2995	0.2919	0.2724	0.2562	t_6
0.11	0.1937	0.2528	0.2725	0.275	0.263	0.2499	t_7
	0.1348	0.2112	0.2436	0.2565	0.2524	0.2428	t_8
	0.0815	0.1697	0.2136	0.2367	0.2409	0.235	t_9
		0.1296	0.183	0.2158	0.2286	0.2265	t_{10}
		0.0921	0.1526	0.1943	0.2155	0.2174	t_{11}
		0.0581	0.1231	0.1724	0.2017	0.2078	t_{12}
			0.095	0.1506	0.1875	0.1977	t_{13}
			0.0688	0.1289	0.173	0.1872	t_{14}
			0.0449	0.1079	0.1582	0.1763	t_{15}
				0.0877	0.1434	0.1652	t_{16}
				0.0686	0.1286	0.1539	t_{17}
				0.0507	0.114	0.1426	t_{18}
				0.0343	0.0997	0.1311	t_{19}
					0.0857	0.1197	t_{20}
					0.0723	0.1084	t_{21}
					0.0594	0.0973	t_{22}
					0.0473	0.0864	t_{23}
					0.0358	0.0758	t_{24}
					0.0252	0.0655	t_{25}
						0.0556	t_{26}
						0.0461	t_{27}
						0.0371	t_{28}
						0.0286	t_{29}
						0.0206	t_{30}

The choice of zero phases in formula (5.1), in a sense, corresponds to a worst-case situation. With the present choice, the intensity of the combination signal (5.3) is maximal: all the contributions to equation (5.4) are nonnegative, and there are no cancellations which would occur under a special phase selection.

Once the stationarity criterion has been introduced, we may set out a task of searching for the most stationary timbres — called τ^{N} timbres — which have the maximal τ, possible for a given number of partials N. The problem may be solved numerically. The results are shown in Table 1, where timbres are normalized to the unit intensity,

$$(t,t)=\sum_{n=1}^{N}t_n^2=1 \tag{5.7}$$

according to the form of the scalar product (5.6). The amplitudes of the partials t_n decrease smoothly with increasing n. The dependence resembles very much the Gaussian law (2.12).

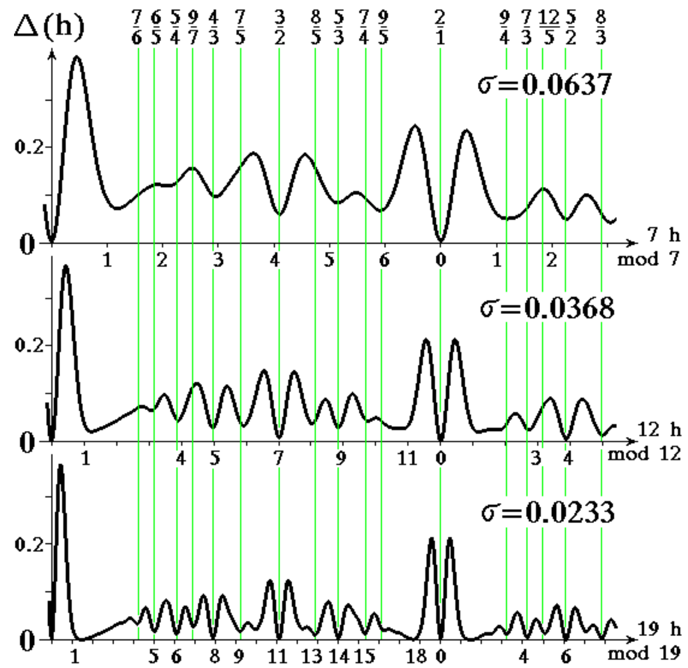

Figure 1. Discordance function plots for the τ^{12} timbre with various values of the dispersion σ. The pitch h is measured logarithmically in fractions of the octave which corresponds to the frequency ratio of 2:1. Positions of the natural intervals are marked by vertical lines.

The obtained timbres permit us to evaluate the discordance function with various dispersion values. As an example, Figure 1 demonstrates $\Delta(h)$ curves with $\sigma = 0.0233$, 0.0368, 0.0637 for the τ^{12} timbre. Distinct minima are observed for 'harmonic' intervals with simple frequency ratios, the 'sharper' σ leading to a more detailed picture.

As we became aware much later than our discordance functions have been computed for the first time, very similar curves had been obtained yet by H. Helmholtz for describing a subjective feeling of dissonance [22]. Moreover, he has found out that the perceived effect depends on the timbre of sound and that higher harmonics generate a more complicated picture of minima and maxima. H. Helmholtz, however, believed that all this can be derived from physiology alone, he did not take into account the possibility of various perceptional tuning and the corresponding higher-level regularities.

As concerns the τ^N timbres constructed by merely requiring the maximum stationarity, an important observation is that narrow intervals, like the diatonic second, cannot properly be described with their aid, even at a larger number of partials N. This is one of the reasons for seeking other higher-level principles of timbre selection we proceed to, bearing in mind the necessity of always sticking to stationary-enough timbres.

6. The discordance Fourier spectrum. Optimal timbres

Consider now the discordance function (formed on the ground of a timbre $\{t_n\}^N$ and a dispersion σ) as a self-contained object of perception. It is reasonable to assume that perceiving the $\Delta(h)$ structure implies distinguishing its different '*rhythmic*' components (periodic in the pitch space). Mathematically, this is performed by the Fourier transformation

$$\tilde{\Delta}(\nu) = \int_{-\infty}^{+\infty} dh \exp(-2\pi i\nu h)\Delta(h)$$
$$\Delta(h) = \int_{-\infty}^{+\infty} d\nu \exp(2\pi ih\nu)\tilde{\Delta}(\nu)\,. \tag{6.1}$$

A maximum in the $\tilde{\Delta}$ magnitude (6.1) at a point ν corresponds to a 'rhythm' in $\Delta(h)$ with the rate of ν degrees to the octave.

The additivity hypothesis and the overtone-series structure lead to a particular form of the Fourier spectrum (6.1) for the discordance function (4.1): it splits in two factors $\tilde{\Delta} = \tilde{\Delta}_\sigma\,\tilde{\Delta}_t$,

$$\tilde{\Delta}_\sigma(\nu) = \sigma\sqrt{\pi}\left(1/2 - \pi^2\sigma^2\nu^2\right) \tag{6.2}$$

$$\tilde{\Delta}_t(\nu) = \left|\sum_{n=1}^{N} t_n \exp(2\pi i\nu \log_2 n)\right|^2 . \tag{6.3}$$

Surprisingly enough, the σ factor (6.2) proves to coincide exactly with the $\nabla^2 G$ filter introduced by D. Marr [10] for general reasons. We see that the 'rhythmic'-structure essence of a scale is contained in the *timbre factor* (6.3). Applying to it filters (6.2) of various measure, we may emphasize one or another specific 'rhythm'. With the Fourier-spectrum tool we can analyze timbres for their inherent 'rhythms', worth emphasizing. Such an analysis of the τ^N timbres immediately exhibits the distinctly preferable scales well known in musicology. However, to reveal these scales in their full development and to get a systematic detailed description of them, we introduce a concept of an *optimal* timbre.

The fundamental principle is *regularity*: a good scale should have a manifest leading periodic component. Thus, we maximize the discordance Fourier-transform amplitude with respect to the parameters of our model, maintaining the normalization (5.7) of the timbre. Regularity is not just a demand of the strictly equal temperament; other, nonleading 'rhythms' are not supposed to be cleared out or suppressed unless they diminish the main Fourier maximum. Neither do we restrict beforehand the rate to be an integer (preconceiving the pure octave): ν is rather a tuning parameter for a local maximum, too.

The regularity of a scale may be quantitatively estimated by the leading-'rhythm' amplitude $\left|\tilde{\Delta}(\nu)\right|$. The dispersion parameter σ enters only into the filter factor (6.2). Maximizing its amplitude, we find

$$\sigma = \pi^{-1}\sqrt{(2+\sqrt{3})/2}\Big/\nu\ . \tag{6.4}$$

With this choice the whole σ factor (6.2) is proportional to the ν reciprocal. Thus, $\Re(\nu) = \tilde{\Delta}_t(\nu)\big/\nu$ is the regularity measure, to be maximized in ν.

The mathematical problem of maximizing an expression of the form (6.3)

$$\varphi = \left|\sum_{n=1}^{N} t_n \exp(i\alpha_n)\right|^2 \tag{6.5}$$

in parameters $t_n \geq 0$ restricted by the normalization condition (5.7), with N and a set of real α_n fixed, has an exquisite self-consistent solution. The nonzero (active) amplitudes are given by

$$t_n = A\cos(\alpha_n - \alpha) \qquad (n \in \{n\})\,. \tag{6.6}$$

normalization (5.7) determines A; and α should be the average phase on the subset of active partials $\{n\}$,

$$\sqrt{\varphi}\,\exp(i\alpha)=\sum_{\{n\}} t_n \exp(i\alpha_n)\,. \tag{6.7}$$

Under these assumptions, one can check that the variation of ϕ (6.5) in $\{t_n\}$ is proportional to the variation of the normalization condition (5.7); this ensures a conditional extremum. The solution to equations (6.6), (6.7), and (5.7) for α, ϕ, and A is

$$\begin{aligned}\tan\alpha &= s\Big/\left(c+\sqrt{c^2+s^2}\right)\\ \varphi &= A^{-2}=\frac{1}{2}\left(N_a+\sqrt{c^2+s^2}\right)\end{aligned} \tag{6.8}$$

where $\begin{matrix}c\\s\end{matrix}=\sum_{\{n\}}\begin{matrix}\cos\\ \sin\end{matrix}(2\alpha_n)$, and N_a is the number of active partials. All one has to do, to find the maximum ϕ, is examine possible subsets $\{n\}$ with the aid of equations (6.8) and (6.6) for consistency: $\cos(\alpha_n-\alpha)>0 \Leftrightarrow n\in\{n\}$.

Now we can start constructing optimal timbres. For any selected rate ν and a number of partials N, we compute $\alpha_n = 2\pi\nu\log_2 n$ and ϕ (6.8). A fine tuning of ν is done to achieve the highest $\Re=\varphi/\nu$. The historical evolution of musical scales is globally directed to sharpening σ — that is, increasing ν, equation (6.4), — and simultaneously to mastering higher overtones — greater N. Starting with $\nu = 1$, $N = 2$, we enable two additional partials for every next scale, cutting the trailing new harmonics that themselves naturally remain inactive. Thus, a sequence of optimal timbres ν^N is generated (Table 2: only stationary-enough timbres with $\tau \geq 0.5$ are listed). Before discussing these results in detail, we ascend one more level of the scale-structure perception.

7. Formation of scale zones

So far, we have not given a precise definition of a scale degree. It is not quite clear whether a minimum of the discordance function is pronounced enough to represent an essential part of the scale. Here we give an answer to a more correct question of defining *zones*, for a given $\Delta(h)$, inside which any tone represents the same entity, agreeable with the key, while tones outside the zones are regarded as dissonant to the scale. The proposed mechanism of comparison with a local average level describes the result of adapting the perception to a repeatedly reproduced image.

The belonging of a tone h to a zone is defined through the *dissonance function* derived from the primary discordance:

$$\square(h)=\Delta(h)-\int_{-\infty}^{+\infty} dx\,\Delta(x) f^{(h,\lambda)}(x)\,. \tag{7.1}$$

The integral represents a local average with the Gaussian weight (2.12), the center and dispersion being h and λ. Whenever the discordance is lower than the local average level, $\square(h) < 0$, the tone does belong to a scale zone. Formula (7.1) looks very simple in terms of the Fourier spectrum (6.1):

$$\tilde{\square}(\nu)=\left[1-\exp(-2\pi^2\lambda^2\nu^2)\right]\,\tilde{\Delta}(\nu)\,. \tag{7.2}$$

Table 2. The optimal timbres with $\tau \geq 0.5$

4	6	7	9	11	12	14	15	N
1.9622	3.0407	3.8809	5.0138	6.0246	6.9428	8.1129	8.9324	ν
0.7312	0.7275	0.6617	0.6718	0.536	0.7248	0.5234	0.611	τ
1.701	1.366	1.163	1.423	1.228	1.249	0.9123	0.9336	$\mathfrak{R}$
0.5466	0.4897	0.4648	0.3743	0.3666	0.3353	0.3625	0.3457	t_1
0.5382	0.4814	0.3915	0.3722	0.3664	0.3328	0.2354	0.3235	t_2
0.4023	0.1796	0.2098	0.3559	0	0.3338	0.2759	0.1725	t_3
0.4996	0.4418	0.109	0.3674	0.3575	0.2879	0	0.2437	t_4
	0.4661	0.4584	0	0.3638	0.2063	0.2415	0	t_5
	0.2893	0.4405	0.3646	0	0.3345	0.3675	0.281	t_6
		0.4132	0.3296	0.2994	0	0.1186	0.2971	t_7
			0.3598	0.34	0.2062	0	0.1206	t_8
			0.2979	0.3155	0.332	0	0	t_9
				0.3676	0.2879	0.3637	0	t_{10}
				0.177	0.3269	0.3071	0.2929	t_{11}
					0.2925	0.2818	0.3394	t_{12}
						0.3511	0.3193	t_{13}
						0.3167	0.344	t_{14}
							0.2892	t_{15}

Table 2. (continued)

17	19	20	22	23	25	26	N
9.9943	11.0209	12.0063	12.9647	13.8934	15.0397	15.9196	ν
0.6306	0.5421	0.6479	0.5616	0.5708	0.6056	0.5426	τ
1.007	0.9628	1.139	0.9227	0.9095	0.8827	0.7899	$\mathfrak{R}$
0.3115	0.3062	0.2704	0.2881	0.2747	0.2739	0.2701	t_1
0.313	0.3064	0.2704	0.2864	0.2529	0.2612	0.2756	t_2
0.209	0	0.2667	0	0.2647	0.1577	0	t_3
0.3142	0.3014	0.2699	0.2707	0.1219	0.2323	0.2121	t_4
0.038	0	0.1909	0.2151	0	0.2492	0.2814	t_5
0.2004	0	0.2647	0	0.2667	0.2082	0.0907	t_6
0.2741	0.2764	0	0	0.2732	0.0313	0	t_7
0.3149	0.2911	0.2691	0.2416	0	0.189	0.0957	t_8
0	0.2728	0.2539	0.2222	0.2501	0	0	t_9
0.0492	0	0.1983	0.2523	0.1077	0.2699	0.2552	t_{10}
0	0.2305	0	0.1897	0.2298	0.2663	0.2067	t_{11}
0.1915	0	0.2623	0	0.1536	0.2459	0.2086	t_{12}
0.3149	0.0403	0	0.2884	0	0	0.2712	t_{13}
0.2795	0.2915	0	0	0.2558	0	0	t_{14}
0.2841	0.2941	0.2229	0	0	0.0318	0.0134	t_{15}
0.3152	0.2758	0.2678	0.2008	0	0.134	0	t_{16}
0.2244	0.299	0.243	0.2889	0.1252	0	0.2091	t_{17}
	0.2889	0.25	0.2575	0.276	0	0	t_{18}
	0.1036	0.2704	0.2475	0.266	0.2195	0	t_{19}
		0.2054	0.2772	0.2457	0.2739	0.1653	t_{20}
			0.2793	0.2623	0.249	0.2771	t_{21}
			0.137	0.2809	0.2417	0.2737	t_{22}
				0.2078	0.2645	0.2625	t_{23}
					0.2684	0.2744	t_{24}
					0.1644	0.2784	t_{25}
						0.1999	t_{26}

Table 2. (continued)

28	30	31	36		50		N
17.0311	18.1164	18.939	22.0135	...	30.9714	...	ν
0.5814	0.5065	0.6692	0.6391		0.646		τ
0.993	0.7509	0.9636	0.8441		0.8457		$\Re$
0.2421	0.2621	0.2296	0.2319		0.1945		t_1
0.2419	0.2415	0.2301	0.2308		0.1947		t_2
0.2411	0	0.2231	0.1818		0.155		t_3
0.2324	0.0976	0.1972	0.2279		0.1888		t_4
0	0.2682	0.2339	0.1726		0.1763		t_5
0.2427	0.0655	0.2334	0.1933		0.1738		t_6
0.0717	0.112	0.0723	0.0751		0.1902		t_7
0.2141	0	0.1358	0.2235		0.1767		t_8
0.2396	0	0.2138	0.049		0.0689		t_9
0	0.1729	0.2205	0.1588		0.1584		t_{10}
0.1996	0	0	0.128		0.1056		t_{11}
0.2351	0.2244	0.21	0.2035		0.1869		t_{12}
0.2429	0.2711	0.1764	0		0		t_{13}
0.1154	0.2482	0.1503	0.0934		0.1792		t_{14}
0	0	0.2314	0.2318		0.1942		t_{15}
0.1877	0	0.0547	0.2174		0.159		t_{16}
0	0.2707	0	0.2305		0		t_{17}
0.2431	0	0.2339	0.0681		0.1004		t_{18}
0	0.2339	0	0		0		t_{19}
0	0	0.1752	0.1439		0.1354		t_{20}
0.0623	0	0.0472	0		0.1851		t_{21}
0.2227	0	0	0.1112		0.1333		t_{22}
0.2398	0.2281	0	0		0.1452		t_{23}
0.2186	0.2687	0.1561	0.2122		0.1941		t_{24}
0.2165	0.2303	0.2325	0.0287		0.1072		t_{25}
0.236	0.205	0.2211	0		0		t_{26}
0.2377	0.2194	0.2019	0		0		t_{27}
0.1547	0.2576	0.2065	0.1111		0.1624		t_{28}
	0.2655	0.2278	0.2177		0		t_{29}
	0.1644	0.2279	0.23		0.1949		t_{30}
		0.1482	0.2146		0		t_{31}
			0.2098		0.1361		t_{32}
			0.2217		0.0029		t_{33}
			0.2319		0		t_{34}
			0.2006		0.1397		t_{35}
			0.0866		0.1288		t_{36}
					0		t_{37}
					0		t_{38}
					0		t_{39}
					0.108		t_{40}
					0.1845		t_{41}
					0.1933		t_{42}
					0.1743		t_{43}
					0.1567		t_{44}
					0.1536		t_{45}
					0.1662		t_{46}
					0.1862		t_{47}
					0.195		t_{48}
					0.1657		t_{49}
					0.0763		t_{50}

In principle, the local-average volume λ may be an independent parameter. However, we set it equal to the dispersion of all the stimuli $\lambda = \sigma$. Choosing $\lambda \gg \sigma$ would only subtract a constant partial from Δ, retaining the h dependence unchanged. In contrast, $\lambda \ll \sigma$ would almost cancel Δ, leaving a very small dissonance function $\square(h)$, proportional to the second derivative of $\Delta(h)$, so that any bend would form a zone. The choice of $\lambda = \sigma$ corresponds to the critical *lowest* value for which the local average of the elementary discordance $d(h)$ has no minimum at $h = 0$ and thus really represents an *average*. Thus, our σ-universality choice seems reasonable. By construction, zones of one fixed level never overlap. Changes in zone widths and their possible overlap [3] should be attributed to a mixing between levels with different σ values.

From formula (7.2) with $\lambda = \sigma$ we see that only the filter $\tilde{\Delta}_\sigma$ is changed in the dissonance Fourier spectrum to

$$\tilde{\square}_\sigma(\nu) = \sigma\sqrt{\pi}\left[1-\exp(-2\pi^2\sigma^2\nu^2)\right]\left(1/2-\pi^2\sigma^2\nu^2\right)\exp\left(-\pi^2\sigma^2\nu^2\right), \qquad (7.3)$$

the timbre factor remaining untouched. The optimal dispersion for ζ degrees to the octave is now determined by maximizing in σ formula (7.3) at $\nu = \zeta$:

$$\begin{aligned} &\sigma = \pi^{-1}\sqrt{C/2}\big/\zeta \\ &C^2 - 4C + 1 = (3C^2 - 6C + 1)\exp(-C) \end{aligned} \qquad (7.4)$$

where the greater root of the equation $C = 3.863\,973\,234\,654\,32$ replaces $2+\sqrt{3}$ as compared to formula (6.4).

To tune to a regular scale, we choose an optimal timbre (which specifies the rate ν) and a dispersion (7.4) according to $\zeta = \nu$. Then the dissonance function $\square(h)$ realizes the most pronounced scale with ν zones to the octave.

However, the perception may as well be not so sharply concentrated as to stick to the basic scale all the time. It may relax, allowing a wider dispersion σ, that is, filtering a slower *effective rate* $\zeta < \nu$. In this way, various-level substructures of the scale may be revealed. The necessary condition is, of course, the presence of the ζ 'rhythm' in the timbre factor (6.3) — an example of the regularity spectrum for the optimal 12^{20} timbre is presented in Figure 2.

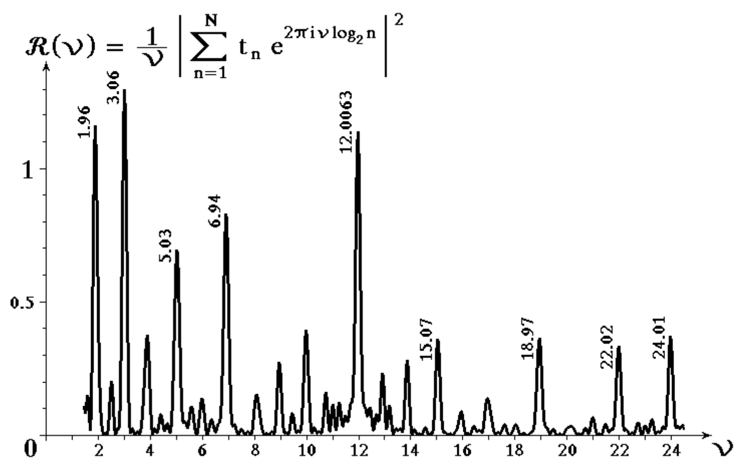

Figure 2. The regularity spectrum of the 12^{20} optimal timbre. The curve shows which substructures with effective rate of ν degrees to the octave can apparently be revealed in the scale.

Simultaneously with widening the dispersion, the redundant higher partials may be ‘switched off’, to economize the efforts of distinguishing them on lower levels of perception. The likeliest for a rate-ζ substructure is a reduction to that number of partials Z which corresponds to an existing optimal timbre ζ^Z. This also improves the regularity $\Re(\zeta) = \tilde{\Delta}_t(\zeta)/\zeta/(t,t)$ and suppresses the noise background $\Delta(0)$ for the substructure. So, ζ^Z reductions form a system of *imbeddings* in the basic scale ν^Z.

As an example, in Figure 3 we show the dissonance-function plot for scale 12^{20} and its reductions to 7^{10}, 5^9, 3^6, and 2^4 (ζ and ν rounded to integers). A further discussion of the curves, in connection with the notion of scale context, follows in section 9.

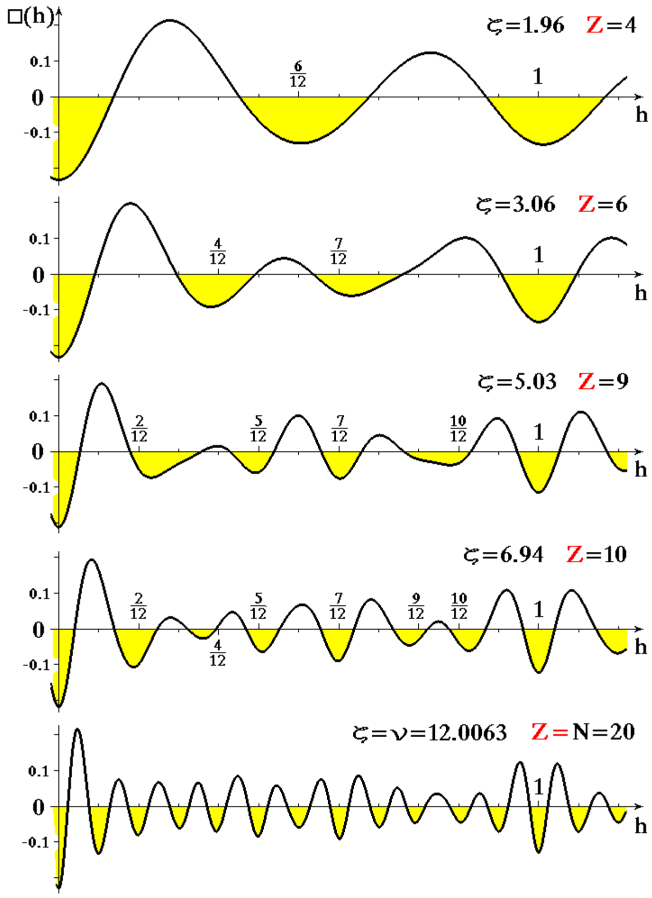

Figure 3. The dissonance function for various imbeddings in the commonly used 12-tone scale. The marked regions where the function is negative define the zones of the degrees of the imbedded sub-scales: duotonic (harmony scale for pentatonic) $\zeta \approx 2$, triad (harmony scale in the usual tonal music) $\zeta \approx 3$, pentatonic $\zeta \approx 5$, diatonic (common mode scale) $\zeta \approx 7$, and chromatic scale $\zeta \approx 12$. Positive values indicate dissonances of the corresponding levels.

8. Timbres and scales

We have computed optimal timbres for $\nu \leq 75$. Table 2 contains all the optimal timbres with $\tau \geq 0.5$ up to $\nu = 22$ and the only timbre with $\tau \geq 0.6$ at higher ν. A remarkable correlation is observed between the stationarity τ and regularity $\Re = \varphi/\nu$ for optimal timbres. We also see that only ‘octaval’ timbres, with an approximately integer number of degrees to the octave, are stationary enough and therefore noticeable in music perception. Nevertheless, a small octavity defect is always present, but still it falls inside the zone width, though coming sometimes up to 40% of $1/\nu$. The least octavity defects are found at $\nu = 10$ and $\nu = 12$. Tetrachords and hexachords cannot appear in this way, because they are rather fragments of a scale than perfect scales, miniature functional ‘models’ of scales with less degrees, and their ‘rhythmic’ structure still belongs to some octaval scale.

Optimal timbres generally feature a number of ‘blocks’ of partials with approximately equal amplitudes, separated by ‘holes’ or, maybe, some isolated harmonics. Holding in mind the resemblance to speech perception [1], we call such blocks ‘*formants*’. Then we find that the musical quality of a scale depends on the formant pattern of its optimal timbre. Like in speech perception, the first and the second formants are most significant. The intervals between the partials of the first formant indicate which intervals in the scale can sound in accord simultaneously, so we refer to the first formant as ‘harmonic’. The length of the harmonic formant correlates with the richness of harmony in the scale, and musical hearing evolves in the direction of ever more complex chords. The poorest are *anomalous* scales, with only the octave possible as a harmony. Four partials in the harmonic formant give *quintal* scales which admit also the fifth (and the fourth) as a harmonic interval; they are still rather poor for modern musical thought. *Tertial* scales, with six partials in the first formant, introduce the third in harmony, which makes it rather good in most cases. Though the most interesting harmony can be achieved in *harmonic* scales, with more than six partials in the first formant. Only in such scales, harmony can freely use the seventh and the second.

We refer to an interval that is generated *inside* the second formant as a ‘move’. In turn, an interval between a partial of the second formant and a partial of the first formant is called a ‘jump’. Both moves and jumps define characteristic intervals of a *mode*, so we call the second formant ‘modal’. When harmonic intervals prevail, the scale is *harmonically labile*, so that any chord may play a centralizing rôle, and music cannot be tonally organized in a wider range. The opposite case, when modal intervals prevail, leads to *modal lability*, weak fixation of tonality in any melodic sequence. In both cases, higher levels of musical hierarchy can hardly be built. This property is usually exploited to produce some coloristic or stylistic effects, but generally it restricts the practical use of the scale [23].

Noting that the numbers of harmonic and modal intervals are equal to $H(H-1)/2$ and $M(M-1)/2+MH$, respectively, where H and M are the lengths of harmonic and modal formants, we can estimate the scale lability by

$$L = \log \frac{M(M-1)+2MH}{H(H-1)} . \tag{8.1}$$

Now, $L < 0$ indicates harmonic, $L > 0$ modal lability. Well balanced timbres with $|L| < 1$ lead to *stable* scales, which can be used most universally, provided their optimal timbre admits a sufficiently developed hierarchy of scale imbeddings. Note that τ timbres, as well as the $\nu = 2, 3, 4$ optimal timbres, possess only one formant and are absolutely harmonically labile, which is related to the absence of narrow intervals in such scales. Of the lower-ν scales, we point to the $\nu = 5$ and $\nu = 7$ modally labile scales which correspond to the well-known pentatonic and diatonic. As it should have occurred, the pentatonic is quintal and admits only the 2^4 harmony of the fourths and the fifths. The diatonic is tertial and therefore adopts the (major and minor) triads.

The first universal scale appears at $\nu = 12$ ($L = 0.49$). This is a tertial scale, so there are some nuisances with the seventh and the second in diatonic chords. But the rich hierarchy of imbeddings (Figure 3) makes it a rather versatile musical tool. Then follow the scales with $\nu = 14, 15, 19, 22$. The ν=14 scale is quintal, and it has a somewhat weaker stationarity. The $\nu = 15$ scale, which corresponds to the full major-minor system, can occasionally be found in musical literature, but the weak seventh partial makes it almost tertial, so, practically nothing new appears as compared to the 12-tone system.

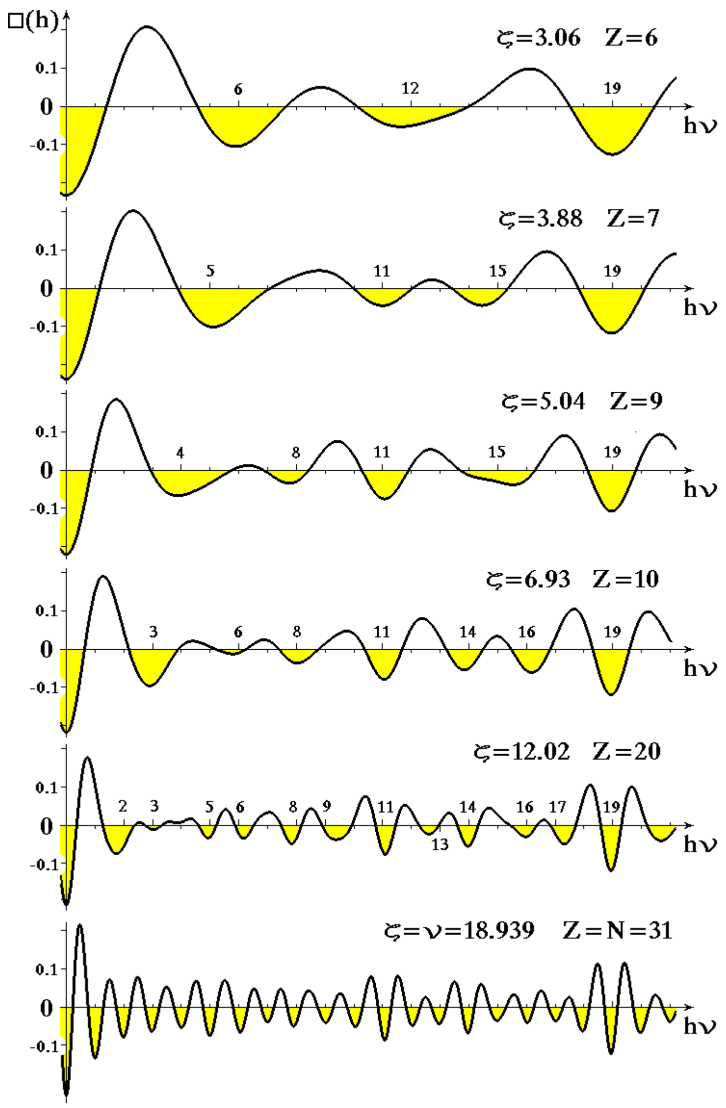

Figure 4. The system of imbeddings in the 19-tone scale. A sub-scale of the seventh chord $\zeta \approx 4$ becomes well pronounced, and non-equally tempered $\zeta \approx 12$ hypermodes with their natural $\zeta \approx 5$ harmony appear as possible imbeddings in the $\nu \approx 19$ basic scale.

Rather stationary and stable ($L = 0.42$) is the $\nu = 19$ scale which can be considered as a harmonic generalization of the tertial 12-tone scale. The more developed hierarchy of imbeddings

(Figure 4) makes it preferable to the $\nu = 22$ scale. At higher ν, only the 31^{50} timbre possesses the stationarity $\tau \geq 0.6$, comparable with 12^{20} and 19^{31}. A greater number of partials than 50 can hardly be comprehended, just because the frequency range is limited; thus, $\nu = 31$ may be considered a higher limit for the number of scale degrees. Still, the $\nu = 31$ scale (proposed back in the 17th century by Christiaan Huijgens) is rather interesting from the musical point of view and is worth a more thorough examination.

Higher partials usually concentrate in a vast 'chromatic' formant which is responsible for the distinction of narrow intervals. The basic scale becomes thus more pronounced, quite like the higher 'vocalist's formant' makes vocal performance more articulate. Too long a chromatic formant may, however, lead to *chromatic lability* of the scale, that is, an easy drift from any musical structure by a narrow interval. On the other hand, if the maximum chromatic move is too short, then some modal moves will not have chromatic analogs. Also, for chromatic stability, the minimum chromatic move should be narrow enough, so that a chromatic alteration of a degree would not change its function. Chromatically stable scales are $\nu = 10, 12, 17, 19, 22, 31$.

9. Scale hierarchy and musical context

The dissonance function generated by an optimal timbre with a certain ν still incorporates 'rhythms' other than the main one (ν degrees to the octave). As the basic scale is fixed in mind, man begins to distinguish its finer features, tuning the perception to various 'rhythms' possible in the scale. The Fourier transform of the dissonance function (6.3)×(7.3) shows which scales can thus be 'imbedded' in the basic scale, with the optimal timbre $\{t_n\}$. Our model specifies the mechanism of such a perceptional tuning: leaving the same t_n , we choose the dispersion σ according to formula (7.4), with $\zeta \neq \nu$. The natural timbre truncation takes place, as described in section 7. The dissonance function that corresponds to the new ζ generates a new zone structure, still associated with the ν-optimal timbre.

Figure 3 shows the imbeddings possible in the 12-tone scale. They form the scale hierarchy for $\nu = 12$. This hierarchy may unfold itself in various ways. For example, the systems of imbeddings $12 \supset 7 \supset 3$ and $12 \supset 7 \supset 5$ can appear. If an imbedded scale is harmonically labile, it can play the rôle of harmony in the basic scale (as the imbedding $\zeta^Z = 3^6$ in the $\nu = 12$ scale). A modally labile or stable imbedded scale defines a possible mode or tonality ($\zeta^Z = 5^9, 7^{10}$). Naturally, the reference tone of any imbedding may differ from the key tone of the basic scale, which just specifies the pitch of a layer in a musical piece.

The movable hierarchical system of imbeddings represents in our model the current musical *context*, defining dissonances and tensions on various levels. In general, a dissonance on a certain level occurs whenever an actually sounding tone does not fit into the current imbedded scale of that level. The tensions of a sound according to the imbedding with the discordance function $\Delta(h)$ can be estimated by the integral

$$\rho(h_1, h_2) = \int_{h_1}^{h_2} \Delta(h)\, dh \,. \qquad (9.1)$$

For example, in the $12 \supset 7 \supset 3$ system of imbeddings, describing the usual tonal music, there exist harmonic dissonances ($\zeta = 3$ level) and modal dissonances ($\zeta = 7$ level), which can be resolved inside the 12-tone scale. There are two ways of releasing the tension: a move to the nearest consonance, distance defined by equation (9.1), and a change of the current imbedding system. Both of them take place in real music, and the detailed study of scale hierarchies would reveal possible solutions in any situation.

The $\nu = 19$ scale (Figure 4) admits modeling tonal music within the $19 \supset 7 \supset 3$ system of imbeddings (Figure 5 shows how the keyboard of a piano should be organized). Moreover, there exist 12-tone 'hypertonalities' predicted by A. Schönberg [6]. Some of them involve no diatonic, so that another system of imbeddings is needed: $19 \supset 12 \supset 5$. Here the $\zeta = 5$ level plays the rôle of harmony in the 12-tone 'hypermodes'. Note that the 5^9 imbedding in the 19^{31} scale is harmonically labile, though the pentatonic $\nu = 5$ is modally labile as an independent scale.

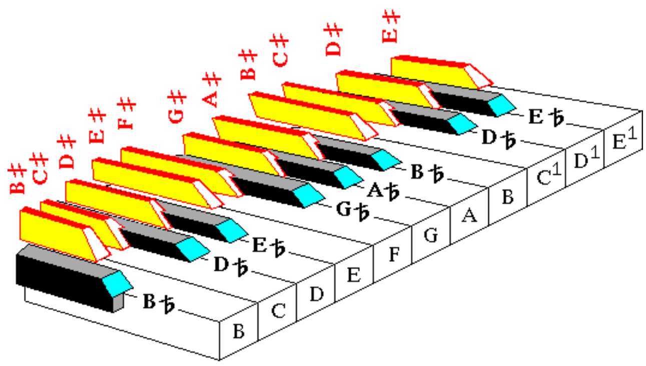

Figure 5. The keyboard of a 19-tone piano. The usual keyboard contains just white and black keys. The 19-tone keyboard has an additional row of *red* keys. Thus, instead of every black key, between the adjacent white keys there are now *two* intermediate degrees in case of a diatonic whole tone, while a diatonic semitone (*B–C*, *E–F*) is *divided* by a red key in half. With the equal temperament, all the intervals between the nearest keys are the same (1/19 of the tempered octave which should be a little stretched to the frequency ratio $2^{19/\nu} \approx 2.00447$). The white and red keys separately form diatonic 7-degree scales, while the black keys, as previously, assemble into a pentatonic. Similarly to diatonic modes, 12-degree *hypermodes* of various structure can be constructed.

Conclusions

We have outlined a theory which is capable of treating the scaling on any level of musical perception. In its static aspect, it describes all existing and possible in the future scales with regard to their musical value and expressive abilities. Also, it shows how ever more complex scales have been forming themselves with the growth in the number of partials comprehended in musical tones and the tightening of the perceptional tuning associated with the dispersion σ. The stages of scale development are embodied in the hierarchy of possible scale imbeddings, which is linked to modal and harmonic levels of music. The model has been elaborated in the framework of a general theory of hierarchies. The logic of hierarchical approach and applied musicological results are to be presented elsewhere.

Though our theory is based on the history of the European music, it can be applied, with minor changes, to musical systems that do not lean so much on scaling, such as oriental modes, ragas, or modern serial and aleatoric investigations. Any inharmonic overtones can as easily be taken into account without complicating the mathematical formalism. We have thus presented a novel outlook at the acoustic foundations of music, ascending from the sensory level of the Helmholtzian physiological acoustics to the psychophysics of perception.

Acknowledgements

We express our gratitude to Alexey V. Muzykantov for interesting conversations and for his kind permission to use a guitar with the 19-tone temperament in experiments at an early stage of this work. We also thank Mikhail P. Papush for useful discussions and a guide in the literature on music theory. Besides, L. V. Avdeev is grateful to Valentin A. Zagrebnov, Alexander L. Kuzemsky, and late Yakov A. Smorodinsky for their encouraging interest in the subject.

References

[1] Gelfand S A 1981 *HEARING: An Introduction to Psychological and Physiological Acoustics* (New York, Basel: Marcel Dekker) (Russ. transl. 1984 Moscow: Medicina)
[2] Pozin N V (ed.) 1978 *Elements of the Theory of Biological Analyzers* (Moscow: Nauka) (in Russian)
[3] Garbuzov N A 1948 *The Zone Nature of Pitch Hearing* (Moscow, Leningrad: USSR Academy of Sciences) (in Russian)
[4] Johnston B 1966 *Persp. of New Music* **5** 112
[5] Kvitka K 1971 Primitive Scales *Selected Works* ed K Kvitka (Moscow: Soviet Composer)
[6] Schönberg A 1973 *Persp. of New Music* **11** 3
[7] Moles A 1958 *Théorie de l'Information et Perception Esthétique* (Paris: Flammarion) (Russ. transl. 1966 Moscow: Mir)
[8] Pribram K H 1971 *Languages of the Brain* (Englewood Cliffs, New Jersey: Prentice-Hall) (Russ. transl. 1975 Moscow: Progress)
[9] Leontiev A N 1975 Activity. Consciouness. Personality (Moscow: Politizdat) (in Russian)
[10] Marr D 1982 *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (New York: Freeman) (Russ. transl. 1987 Moscow: Radio i sviaz)
[11] Golitsyn G A 1980 Information and the Laws of Aesthetic Perception *Numbers and Thought* (Moscow: Znanie) **3** (in Russian)
[12] Yasser J 1932 *A Theory of Evolving Tonality* (New York: American Library of Musicology)
[13] Sonneborn H 1983 *Popular Computing* **3** 218
[14] Stoney W 1970 Theoretical Possibilities for Equally Tempered Musical Systems *The Computer and Music* ed H B Lincoln (Ithaca, London: Cornell Univ. Press) p 163
[15] Hartmann G C 1987 *Amer. J. Phys.* **55** 223
[16] Gagliardo E 1980 *Enneadecaphonic Music. A New System of Harmonic Tones.* (Atti dell'Accademia Ligure) **37** 536
[17] Bernstein R B and Levine R D 1975 *Adv. At. Mol. Phys.* **11** 215
[18] Moles A 1962 *Sociodynamique de la Culture* (Paris: Flammarion) (Russ. transl. 1962 Moscow: Mir)
[19] Garbuzov N A 1956 *The Zone Nature of Timbre Hearing* (Moscow: Muzgiz) (in Russian)
[20] Regener E 1967 *Persp. of New Music* **6** 52
[21] Lewin D 1980 *Persp. of New Music* **18** 498
[22] Kuznetsov L A 1989 *Acoustics of musical instruments* (Moscow: Legprombytizdat) 66 (in Russian)
[23] Katunyan M 1983 Problems of Musical Science (Moscow: Sov. Composer) **5** 4 (in Russian)

http://unism.pjwb.org/arc

http://unism.pjwb.net/arc

http://unism.narod.ru/arc