Hearing Research, 1 (1979) 155–182

©Elsevier/North Holland Biomedical Press

CALCULATING VIRTUAL PITCH

ERNST TERHARDT

Institut für Elektroakustik der Technischen Universität München, Arcisstr. 21, D-8000 Munich 2, F.R.G.

(Received 16 October 1978; accepted 5 December 1978)

A procedure for the schematic and automatic extraction of 'fundamental pitch' from complex tonal signals, such as voiced speech and music, has been developed. While the auditively relevant 'fundamental' of a complex signal cannot be defined in purely mathematical terms, an existent model of virtual-pitch perception turns out to provide a suitable basis. The procedure comprises the formation of determinant spectral pitches (which correspond to the frequencies of certain signal components), and the deduction of virtual pitch (or 'fundamental frequency') from those spectral pitches. The latter deduction is accomplished by a principle of subharmonic matching, for whose realization a simple, universal and efficient algorithm was found. While the calculation may be confined to the determination of 'nominal' virtual pitch, certain typical auditory phenomena, such as the influence of SPL, partial masking and interval stretch, may be accounted for as well, in which case 'true' virtual pitch is obtained. The procedure operates on the frequencies and amplitudes of the signal's spectral components, is suitable for implementation on readily available programmable calculators and other arithmetic computers, and may be used in real-time 'fundamental-pitch' extraction as well. The procedure's performance and its applicability to the research and engineering of auditory communication are illustrated by some examples.

Keywords: virtual pitch; pitch calculation; fundamental-frequency extraction; pitch shifts; interval stretch.

INTRODUCTION

In the research and engineering on auditory communication it is desirable to have at hand appropriate means for the evaluation of the pitch which is produced by any given auditory signal. The signals which normally are used in human auditory communication are speech and music, and the pitch of this type of signal is in a complex way dependent on their physical parameters. A procedure is required by means of which the extraction of pitch (or equivalent frequency) from the signal parameters can be accomplished as quickly and conveniently as possible. Such a procedure may be helpful for example in the design of experiments on pitch perception and the evaluation of their results, in the evaluation of musical sounds, particularly those whose spectral components are not perfect harmonics (e.g. bells), and in automatic, real-time pitch extraction from speech or music. The aim of the present study is to yield such a procedure, following the principles of virtual-pitch perception, which have been described and whose psychoacoustic foundations have been discussed previously [18,22–24,28].

Virtual pitch is the type of pitch which is produced by complex signals, in contrast to *spectral pitch*, which is the type of pitch evoked by pure tones. While spectral pitch may

be considered as a product of relatively peripheral auditory analysis, the perception of virtual pitch probably is dependent on higher (i.e., more central) stages of auditory processing. In the perception of complex tones, e.g. voiced speech sounds, both types of pitch play a role. While some lower harmonics (including the fundamental) may produce simultaneous, corresponding spectral pitches, virtual pitch is the prevalent pitch mode of those signals and corresponds to the 'fundamental frequency'. 'Extraction of fundamental frequency' is in some respect equivalent to extraction of virtual pitch. In a strict sense, however, the frequency which corresponds to virtual pitch and the fundamental frequency are in general not identical. For example, the fundamental frequency of a complex signal consisting of three partials with the frequencies 520, 620 and 720 Hz is 20 Hz, while the perceived virtual pitch corresponds to about 104 Hz. Hence in the analysis of auditory signals such as speech and music actually 'extraction of fundamental frequency' is not the real aim but rather 'extraction of the frequency which corresponds to virtual pitch'.

In the following sections it will be demonstrated how that extraction may be accomplished by a systems approach to auditory processing of complex tonal stimuli, based on relatively simple and established psychoacoustic facts such as quantitative pitch data and masking patterns. A particular aim was to obtain sufficiently simple formulae and algorithms by means of which virtual pitch can be calculated schematically without requiring any psychoacoustic background knowledge.

THE PRINCIPLE OF DEDUCING VIRTUAL PITCH FROM THE PHYSICAL SOUND PARAMETERS

A relatively concise and yet precise specification of the transformation of sound parameters into virtual pitch is enabled by the established fact that virtual pitch can be deduced exclusively from the spectral pitches which are evoked by the complex stimulus [22–24]. Hence, for the present purpose the physical and physiological mechanisms which link the physical stimulus parameters to virtual pitch are of secondary interest. In particular, the frequently discussed question of whether pitch is essentially a product of temporal or spectral auditory analysis is without significance here. The problem is reduced to two questions:

(a) What is the functional relationship between physical stimulus parameters and spectral pitch?

(b) Which is the functional relationship between simultaneous spectral pitches and virtual pitch?

The answer to the first question is simple: To a particular spectral pitch there corresponds one particular spectral component (partial), and the spectral-pitch magnitude is essentially dependent on the component's frequency. A preliminary answer to the second question is: Virtual pitch is obtained by a specific matching process of subharmonics pertinent to certain determinant spectral pitches. This process is dependent on a previous perceptual mode of learning, in which the knowledge of harmonic pitch intervals has been acquired.

The principle is illustrated in Fig. 1, which is nothing but a replication of the virtual-pitch model [23,24] in the form of a flow-chart. In a pitch-extraction system such as the

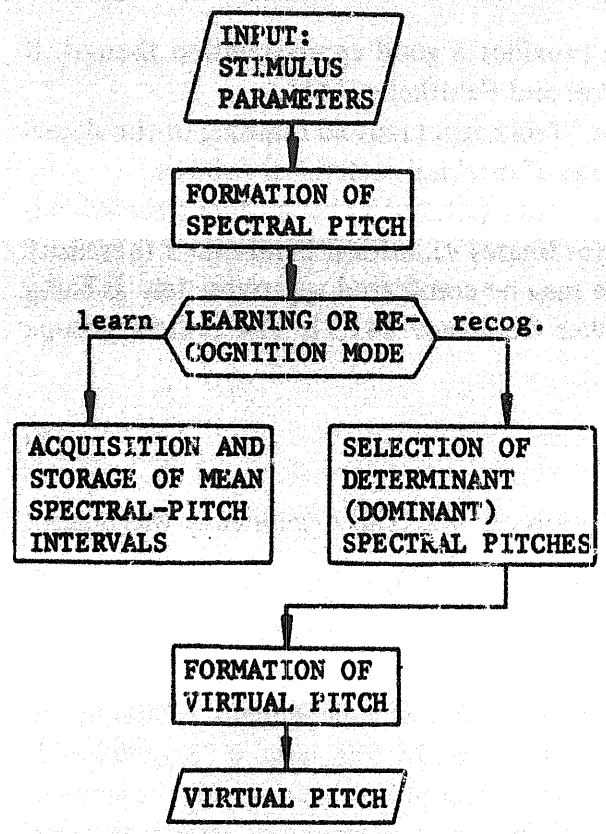


Fig. 1. Flow-chart representation of the virtual-pitch model [23,24]. It is supposed, that the extraction of virtual pitch depends on a previous learning process by which the series of basic harmonic pitch intervals has been acquired.

one to be developed, finally only the recognition-path is required. The acquired harmonic intervals are then established parameters of the system. However, sufficient understanding of the entire pitch-extraction process and appropriate evaluation of its essential parameters cannot be obtained without considering the acquisition process as well. In the following sections the separate steps of processing depicted in Fig. 1 will be specified quantitatively. It goes without saying that this can be accomplished only by introducing certain simplifications and idealizations. Their validity must be evaluated in terms of the final system's performance.

FORMATION OF SPECTRAL PITCH

The problem of quantitatively describing the formation of spectral pitch implies two parts, namely:

(a) evaluation of the *existence* of spectral pitch, and

(b) precise quantitative specification of the *stimulus-to-sensation relationship.*

Existence of spectral pitch

The most elementary restriction is the absolute threshold of hearing. The formula

$$L_{HS} = \{3.64x^{-0.8} - 6.5\exp[-0.6(x-3.3)^2] + 10^{-3}x^4\}\ \text{dB} \qquad (1)$$

(L_{HS} = SPL at hearing threshold; $x = f/\text{kHz}$) provides a good approximation thereof. It has been fitted to the data published by Zwicker and Feldtkeller [41].

The second and most significant restriction comes from mutual masking of simultaneous partials. This effect is quantified on the basis of masking patterns as follows.

In Fig. 2a, b the partial masking of one pure tone (labeled μ) by another (labeled ν) is illustrated. When displayed in terms of SPL (ordinate) vs. critical band-rate z (abscissa), the masking pattern produced by a pure tone may be considered approximately as being triangular in shape and independent of masker frequency [41,7]. The pattern's slope towards lower z values is

$$S_1 = 27 \text{ dB/Bark} \tag{2}$$

while the slope towards higher z values, S_2, depends on SPL. This dependency was quantified by the formula

$$S_2 = [24 + 0.23(f_\nu/\text{kHz})^{-1} - 0.2L_\nu/\text{dB}] \text{ dB/Bark} \tag{3}$$

At masker frequencies beyond about 100 Hz, S_2 is almost independent of frequency while at very low frequencies the steepness is increased by the term $0.23\ (f_\nu/\text{kHz})^{-1}$, taking into account the influence of the absolute threshold of hearing. Fig. 2c demonstrates that the calculated S_2 is in good agreement with corresponding values obtained from psychoacoustic masking patterns [41].

The partial labeled ν produces at the place of partial μ a certain masker level $L_{\nu\mu}$, which is

$$L'_{\nu\mu} = L_\nu - S_2(z_\mu - z_\nu) \qquad \text{in the case of } f_\nu < f_\mu \tag{4a}$$

and

$$L''_{\nu\mu} = L_\nu - S_1(z_\nu - z_\mu) \qquad \text{in the case of } f_\nu > f_\mu \tag{4b}$$

The difference $\Delta L_\mu = L_\mu - L_{\nu\mu}$ is called the *sound-pressure-level excess* of the partially masked tone μ and is considered as a suitable criterion of whether or not that tone produces a significant spectral pitch. When $\Delta L_\mu = 0$, the partial μ is about 3–6 dB above its masked threshold. As a tone being just a few dB above its masked threshold may already produce already a well-pronounced pitch, it appears to be justified to consider an SPL excess of 0–3 dB as the minimum typically required for the existence of an individual spectral pitch.

The transformation of frequency f into critical band-rate z is obtained by

$$z = 13.3 \arctan(0.75 f/\text{kHz}) \text{ Bark} \tag{5}$$

($0 \leqslant f \leqslant 4$ kHz), which closely approximates the *Tonheit* function [41,39] for frequencies below 4 kHz. As frequencies beyond that limit are not involved in the formation of virtual pitch, this restriction is acceptable for the present purpose.

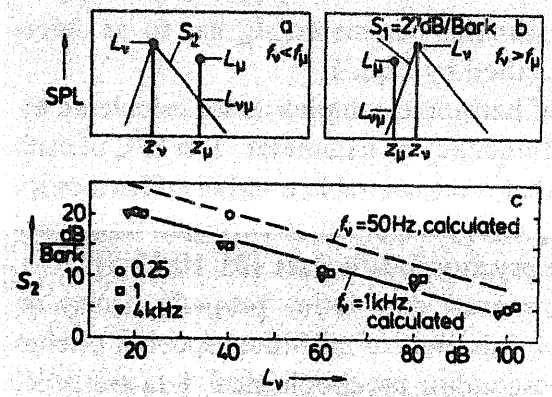


Fig. 2. Quantification of partial masking of one spectral component (frequency f_μ, corresponding to critical band-rate z_μ), by another (f_ν; z_ν). If $f_\nu < f_\mu$, the steepness S_2 of the higher slope of the simplified masking pattern determines the SPL excess ($L_\mu - L_{\nu\mu}$) of component μ (a). If $f_\nu > f_\mu$, S_1 is relevant (b). While S_1 is constant, S_2 depends on L_μ and (slightly) on f_μ, as depicted in c. The symbols in c represent steepness magnitudes of experimental masking patterns of narrow-band noise [41].

As a more general case, the situation must now be considered where one spectral component of a complex tonal stimulus is partially masked by *several* other components. Thus the question arises how the masking effect of different maskers at the place of the maskee is summated. Corresponding psychoacoustic experiments show clearly that in general there is not a simple addition of sound intensities [41,42]. The cooperation of several maskers cannot yet be described with full precision by a simple law, and the relative phase between the involved components plays a considerable role as well [40]. Thus a formalism was chosen which appears to be a fair compromise and may be sufficient for the present purpose: The combined masking effect, produced by several spectral components, is represented by the *sum of their amplitudes.*

The relative amplitude pertinent to a particular value of $L_{\nu\mu}$ is

$$A_{\nu\mu} = 10^{L_{\nu\mu}/20\text{dB}} \tag{6}$$

If there are N partials simultaneously present, the resulting relative masker amplitude A_μ at the place of the μth partial is

$$A_\mu = A'_\mu + A''_\mu = \sum_{\nu=1}^{\mu-1} 10^{L'_{\nu\mu}/20\text{dB}} + \sum_{\nu=\mu+1}^{N} 10^{L''_{\nu\mu}/20\text{dB}} \tag{7}$$

($f_1 < f_2 < f_3 < \ldots < f_N$). $L'_{\nu\mu}$ is specified by Eqn. 4a, $L''_{\nu\mu}$ by Eqn. 4b. The first sum in Eqn. 7 is called A'_μ, the second A''_μ; they will be used once again separately for calculating the precise pitch of partial μ.

In order to determine the ultimate SPL excess of the μth partial, the absolute threshold of hearing L_{HS} has also to be taken into account. This is accomplished by intensity addition such that finally the SPL excess is specified by

$$\Delta L_\mu = L_\mu - 10 \log^{*} (A_\mu^2 + 10^{L_{\text{HS}}/10\text{dB}}) \text{ dB} \tag{8}$$

* Logarithm to base 10.

L_μ is the SPL of the considered partial, i.e., a stimulus parameter. A_μ has to be determined by Eqn. 7, using also Eqns. 2–5. L_{HS} is specified by Eqn. 1.

The SPL excess of the individual harmonics of harmonic complex tones calculated by this formula is shown in Fig. 3 with fundamental frequency as parameter. The SPL of each harmonic is assumed to be 60 dB. It is evident that a considerable number of harmonics (up to about 10, depending on fundamental frequency) can produce corresponding spectral pitches. This result agrees with corresponding psychoacoustic data [31,10,11,5].

As the procedure reflects the available experimental data rather properly it may be considered as an appropriate means to evaluate the existence of individual spectral pitches in case of stimuli, about which there are no corresponding psychophysical data available. As an example, the spectral envelopes of six different speech vowels were calculated from suitably chosen formant frequencies F_n by means of the formula

$$L = (2.16x^2 + 0.023x^4)\ \text{dB} + 20\log\{x/(1 + 100x^2)\}\ \text{dB}$$

$$- \sum_{n=1}^{4} 10\log\{(1 - x^2/X_n^2)^2 + 2.5 \cdot 10^{-3}(x + 0.1x^4)^2/X_n^4\}\ \text{dB} \qquad (9)$$

(L = SPL at frequency f; arbitrary reference; $x = f/\text{kHz}$; $X_n = F_n/\text{kHz}$), which is a slightly modified version of that provided by Fant [2], and is based on Fant's acoustic theory of

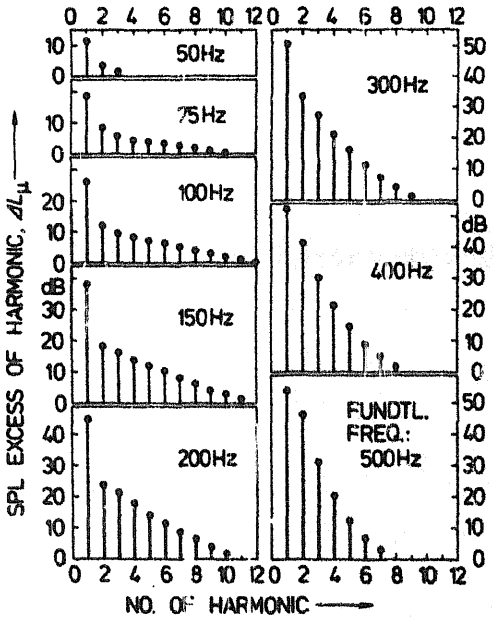


Fig. 3. SPL excess ΔL_μ of the harmonics of a harmonic complex tone, of which each harmonic has an SPL value of 60 dB; calculated from Eqn. 8. Parameter is the complex tone's fundamental frequency.

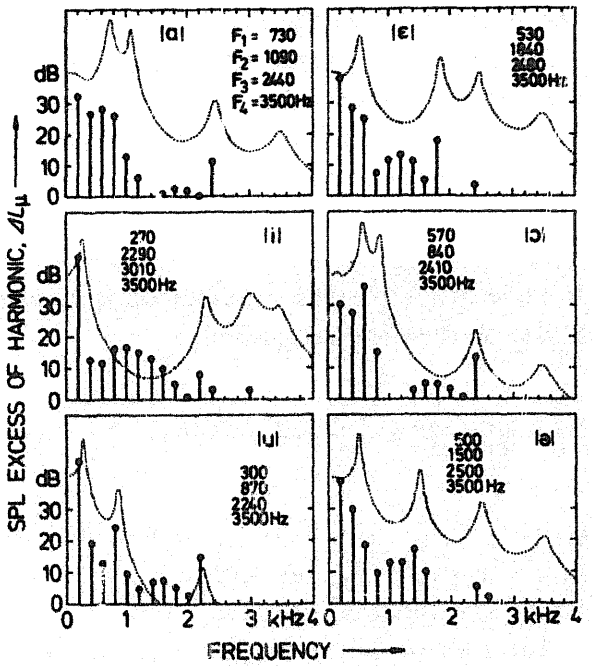


Fig. 4. SPL excess ΔL_μ of the harmonics of speech vowels; calculated from Eqn. 8. The vowel spectra were calculated from Eqn. 9 on the basis of four formant frequencies F_n, assuming an SPL value of 60 dB. The vowel spectra are shown in terms of the dB-scale at the ordinate, but with an arbitrary reference level.

vowel production. The overall SPL of each vowel spectrum, extending from the fundamental frequency to 4 kHz, was assumed to be 60 dB. The fundamental frequency was 200 Hz. The calculated SPL excess of the harmonics of these vowel spectra is depicted in Fig. 4, and the calculated spectral envelopes are also displayed. The resulting SPL-excess patterns reveal that a considerable number of harmonics produce individual spectral pitches in speech vowels and thus may be considered as being perceptually relevant.

The stimulus-to-sensation relationship of spectral pitch

As spectral pitch essentially and monotonically depends on the frequency of the corresponding partial, the most convenient and unbiased approach to its quantification is to specify it in terms of a tone's frequency. However, as spectral pitch depends somewhat on SPL, partial masking and some other parameters as well, one has to prescribe these additional parameters in order to get an unambiguous definition of pitch magnitude. This is accomplished by defining a *standard tone*, which is an *isolated, unaffected pure tone with 60 dB SPL.* (For the present purpose, this definition is more convenient than that suggested by Fletcher [3] and the present author previously [22]; there a standard loudness level of 40 phon was suggested.) The spectral pitch of the standard tone is specified by

$$H = f_S/\text{Hz pu} \tag{10}$$

where H is the pitch magnitude (a perceptual entity), expressed in *pitch units* pu, and f_S is the standard tone's frequency.

The standard tone is used to describe the pitch of any arbitrary auditory signal. This is accomplished by the definition Eqn. 10 and, in addition, a listening experiment in which the standard tone's pitch is matched to that of the test signal. It should be noticed that this is but an analogy of measuring the sensation of loudness by means of the loudness level L_N. The equation

$$L_N = L_{1\text{kHz}}/\text{dB phon} \tag{11}$$

with $L_{1\text{kHz}}$ being the SPL of a pure 1-kHz tone which has been matched in loudness to the test stimulus, is equivalent to Eqn. 10.

Following this concept, the spectral pitch of an arbitrary pure tone with the frequency f_T is described as follows. In a listening experiment, the test tone and the standard tone are matched in pitch. As a result, a certain systematic departure of f_S from f_T will usually be observed. The normalized value of this departure is called *pitch shift* v:

$$v = (f_S - f_T)/f_T|_{\text{equal pitch}} \tag{12}$$

v hardly exceeds a few percent. $v > 0$ indicates that the test-tone pitch is higher than that of the standard tone with the same frequency. By introducing $f_S = f_T(1 + v)$ from Eqn. 12 into Eqn. 10, one obtains

$$H = f_T/\text{Hz}\,(1 + v)\,\text{pu} \tag{13}$$

This equation specifies completely the relation between stimulus parameters (frequency, SPL, etc.) and spectral pitch provided that v, as a function of SPL, partial masking, etc., is known. In many cases v may be ignored because of being so small. It is advisable to make a clear distinction between that pitch which accounts for the pitch shift effects, i.e. *true* pitch, and the pitch which is numerically identical with frequency, as v is ignored; the latter will be termed *nominal* pitch.

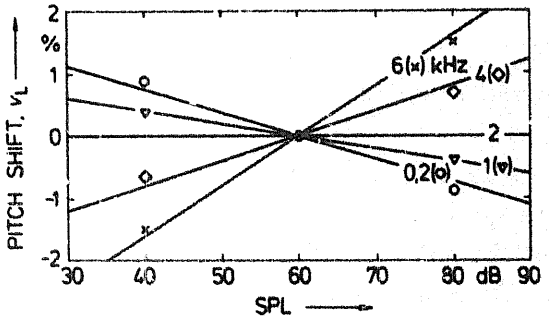
For the present purpose it is sufficient to quantify the dependence of v on SPL and mutual partial masking of simultaneous spectral components. As v is small, it is sufficient to investigate these two effects separately and superimpose the results:

$$v = v_L + v_M \tag{14}$$

where v_L represents the influence of SPL, and v_M that of partial masking. A simple representation of the SPL influence is

$$v_L = 2 \cdot 10^{-4}(L_\mu/\text{dB} - 60)(f_\mu/\text{kHz} - 2) \tag{15}$$

where L_μ and f_μ are SPL and frequency of the considered spectral component, respectively. Fig. 5 shows the relationship between the calculated pitch shift v_L of a single pure tone and corresponding experimental data (from 15 subjects) [25]. The approximation of the psychoacoustically observed mean tendency is sufficient; however, an individual subject's pitch may deviate considerably [30,9].



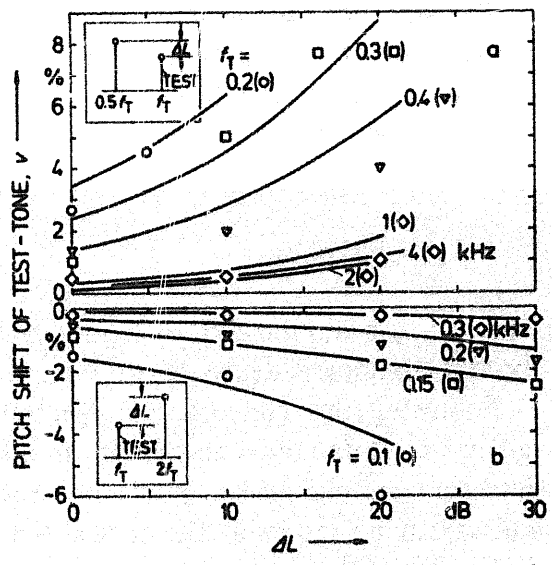


Fig. 5. Pitch shift v_L of a single pure tone as a function of SPL, with frequency as parameter; calculated from Eqn. 15 (solid lines). The symbols indicate corresponding experimental results (means of data from 15 subjects [25]).

Fig. 6. Pitch shift of a pure test-tone with frequency f_T, which is partially masked by another tone with frequency $0.5f_T$ (a) and $2f_T$ (b), respectively; calculated from Eqns. 14, 15 and 17, with a constant loudness levels of the test tone, $L_N = 50$ phon (solid curves). Parameter is the test-tone frequency f_T. Symbols represent corresponding experimental data [29].

The pitch shift caused by mutual partial masking, v_M, is described as follows. The shifting influence of those components which are lower in frequency than the shifted partial (labeled μ) is considered to be dependent of the corresponding SPL excess $\Delta L'_\mu$. Likewise, the effect of the higher components is dependent on $\Delta L''_\mu$. Explicitly, these two SPL-excess values are obtained by

$$\Delta L'_\mu = L_\mu - 20 \log A'_\mu \text{ dB} \tag{16a}$$

$$\Delta L''_\mu = L_\mu - 20 \log A''_\mu \text{ dB} \tag{16b}$$

where A'_μ and A''_μ are the two sums in Eqn. 7.

In addition a dependence on frequency is introduced as observed in psychoacoustic results. After some iterative fitting in which several sets of related experimental data were consulted [1,37,36,32,19,29], the formula

$$v_M = 1.5 \cdot 10^{-2} \exp(-\Delta L'_\mu/20\text{dB})(3 - \ln f_\mu/\text{kHz}) + 3 \cdot 10^{-2} \exp(-\Delta L''_\mu/20\text{dB})(0.36 + \ln f_\mu/\text{kHz}) \tag{17}$$

was established which sufficiently approximates the available experimental data.

Reconsidering Eqns. 13 and 14, the spectral pitch of any spectral component μ with the frequency f_μ is completely specified by

$$H_\mu = f_\mu/\text{Hz}(1 + v_L + v_M) \text{ pu} \tag{18}$$

where v_L and v_M are given by Eqns. 15 and 17. It should be noticed that the pitch H_μ may readily be 'retransformed' into 'equivalent frequency', i.e., frequency of standard tone with the same pitch by Eqn. 10. Formally this means that in Eqn. 18 the unit 'pu' has simply to be replaced with 'Hz'.

The validity of Eqn. 17 is verified by Figs. 6 and 7. In Fig. 6a, the pitch shift of a pure tone with frequency f_T caused by the simultaneous presence of another tone with frequency 0.5 f_T is depicted as a function of the SPL difference between the two tones, ΔL,

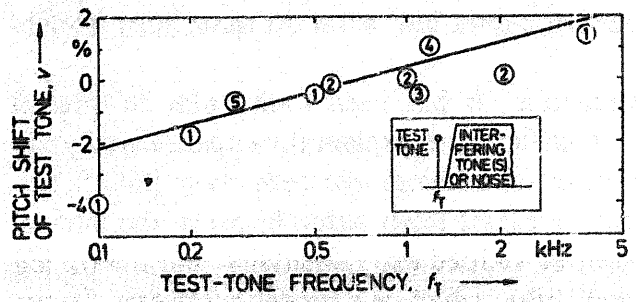


Fig. 7. Pitch shift v of a test tone with frequency f_T which is partially masked by another tone or random noise higher in frequency, but adjacent to the test tone (see insert). The solid line is obtained from Eqns. 14, 15 and 17, with $\Delta L'_\mu \to \infty$; $\Delta L''_\mu = 20$ dB. Related experimental data are from: (1) ref. 29; (2) ref. 36; (3) ref. 37; (4) ref. 32; (5) ref. 1.

while f_T is the parameter. In Fig. 6b, the frequency relation of the two tones is reversed (experimental data from ref. 29). Evidently, the calculated data (solid curves) fit the experimental results reasonably well. In Fig. 7, the specific and remarkable dependence of the pitch shift of a pure tone (frequency f_T) which is partially masked by adjacent *higher* tones or random noise (see insert) on frequency is illustrated. The straight line is obtained from Eqn. 17 by setting $\Delta L'_\mu \to \infty$; $\Delta L''_\mu = 20$ dB.

ACQUISITION OF HARMONIC PITCH INTERVALS

As outlined already, a basic feature of the virtual-pitch model is, that virtual pitch is deduced from spectral pitches of signal components by application of a series of sub-harmonic pitch intervals. In terms of frequency, harmonic (and hence also sub-harmonic) intervals are specified by ratios of integers, i.e., 1 : 2, 1 : 3, 1 : 4, etc. In terms of pitch, however, the corresponding ratios may differ significantly therefrom, even if the present definition of pitch magnitude (pitch proportional to frequency) is used. As an important example, consider the simultaneous harmonics of a harmonic complex tone. The frequency ratio of the mth harmonic to the fundamental is m, but the corresponding spectral-pitch ratio is

$$H_m/H_1 = f_m(1 + \upsilon_m)/\{f_1(1 + \upsilon_1)\} \tag{19a}$$

which with sufficient approximation can be replaced by

$$H_m/H_1 = m(1 + \upsilon_m - \upsilon_1)\,. \tag{19b}$$

H_m, H_1 are the spectral pitches of the mth and first harmonic, respectively, f_m and f_1 the corresponding frequencies, and υ_m and υ_1 the corresponding pitch shifts. These pitch shifts were calculated by Eqns. 14, 15 and 17 assuming that the complex tone consists of many harmonics, each of which has an SPL of 60 dB. Fig. 8 shows the results as a function of fundamental frequency (solid curves). Evidently, the fundamental is shifted down at low frequencies while the higher harmonics are shifted upward in the whole frequency range. As a consequence, $(\upsilon_m - \upsilon_1)$ is a significantly positive value with any m and fundamental frequency. The harmonic *intervals*, in terms of pitch, are *stretched* by the amount $(\upsilon_m - \upsilon_1)$. In Fig. 8 some corresponding experimental results are shown [21]. In detail, these data do not fit perfectly to the calculated curves but reveal the same basic tendencies.

In previous studies of virtual-pitch perception, it has been found that an interval stretch of exactly this type is required and is sufficient to explain the experimentally observed virtual-pitch magnitude on the principle of subharmonic deduction [34,18,22–24]. Thus it was concluded that there exists a causal relationship between the interval stretch of simultaneous harmonics, which can be verified experimentally, and the stretch of subharmonic intervals in virtual-pitch deduction, which is a model parameter. It was assumed that the knowledge of harmonic pitch intervals is acquired in an early learning phase by repeated perception and processing of speech [19,18,22–24,28]. The fact that considerable portions of natural speech are truly harmonic complex tones (i.e., the voiced

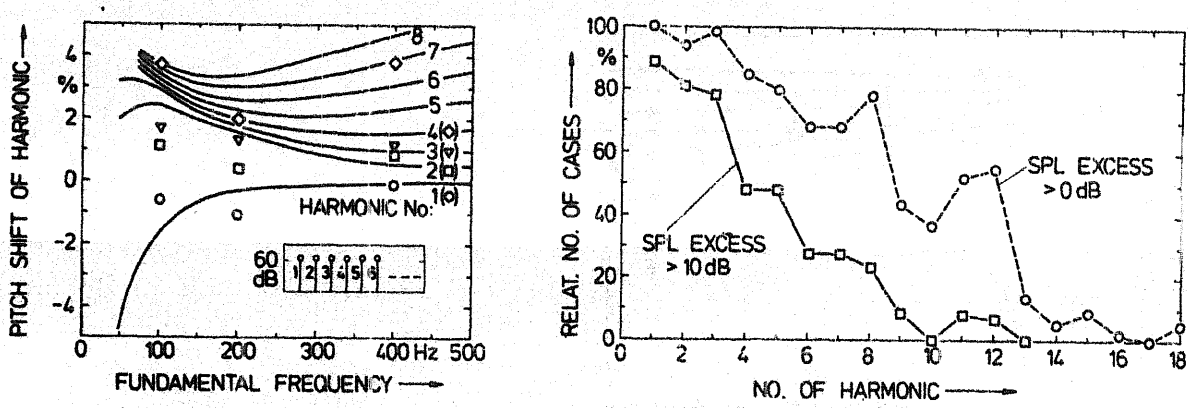


Fig. 8. Calculated pitch shifts of lower eight harmonics of a complex tone, comprising all harmonics up to 4 kHz with individual SPL values of 60 dB (solid curves; Eqn. 17). Symbols represent corresponding experimental data [21].

Fig. 9. Relative number of cases in which the calculated SPL excess (Eqn. 8) of individual harmonics of the six vowels |a|, |e|, |i|, |ɔ|, |u|, |ə| (cf. Fig. 4) exceeds 0 and 10 dB, respectively. The data are based on three fundamental frequencies (100, 200, 300 Hz) and three SPL values (40, 60, 80 dB), i.e., 54 vowel spectra in all.

speech segments) and that speech (or the human voice as such) is of highest biological significance in the development of each individual provides considerable plausibility to that assumption. Consequently, the process of interval acquisition by repeated processing of speech was included in the present study.

The question of whether the harmonics of speech vowels may actually produce individual spectral pitches has been already investigated quantitatively (cf. Fig. 4). In order to obtain some additional statistical data, the SPL excess of each harmonic of the six vowels described by Fig. 4 was calculated with fundamental frequencies of 100, 200 and 300 Hz (thus representing the most typical range of voice pitch) and SPL values of 40, 60, and 80 dB, 54 vowel spectra in all. The number of cases in which the SPL excess was greater than 0 and 10 dB, respectively, was determined for each harmonic. The result (Fig. 9) suggests that, in human speech the chance of perceiving even relatively high harmonics as individual spectral pitches is very great.

Another study was devoted to the *mean* width of the spectral-pitch intervals existing between the harmonics of vowels. Taking Eqn. 19 into consideration, this requires that for each harmonic the mean pitch shift $\bar{v}_m$ was to be determined as a function of fundamental frequency. This has been done for eight harmonics with the six vowels depicted in Fig. 4, assuming an overall SPL of 60 dB. The result is shown in Fig. 10. The position of the harmonic numbers in the diagram indicates the calculated mean pitch shifts (arithmetic means of six individual values). The basic tendencies which were observed in Fig. 8 are well confirmed; however, there is some unsystematic variance in the data of higher harmonics. This is caused by the great difference between the spectral envelopes of the vowels. The ensemble of only six vowels obviously is not sufficient to yield statistically reliable data. However, taking into consideration the results pertinent to a harmonic complex tone with 'horizontal envelope' (Fig. 8), these data appear as a sufficient basis of an approximate quantitative representation. The solid lines in Fig. 10 depict the final

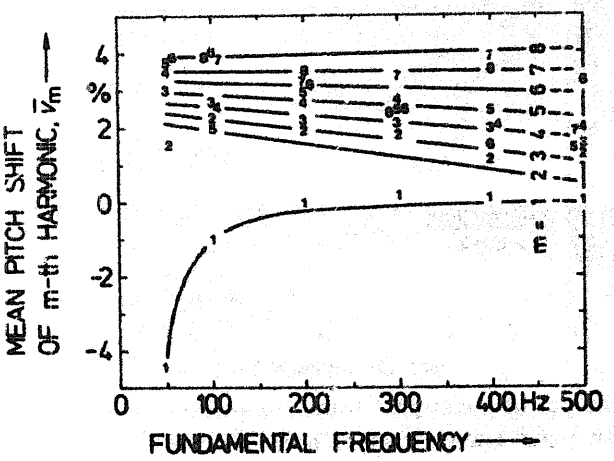


Fig. 10. Calculated mean pitch shifts $\bar{v}_m$ of the lower eight harmonics of the same vowels as in Figs. 4 and 9 as a function of fundamental frequency. The means of six vowels with 60 dB SPL are indicated by the position of the corresponding harmonic numbers in the diagram. The solid curves, labeled with harmonic number m, represent an approximation to these data which is specified by Eqn. 20a, b.

representation which was chosen. They are specified by:

$$\bar{v}_1 = -10^{-4}(f_1/\text{kHz})^{-2} \tag{20a}$$

$$\bar{v}_m = 10^{-3}\{18 + 2.5m - (50 - 7m)\, f_1/\text{kHz}\} \tag{20b}$$

($m = 2, 3, 4, \ldots$; $50\ \text{Hz} \leqslant f_1 \leqslant 500\ \text{Hz}$). In analogy to Eqn. 19, the series of harmonic pitch intervals, which is assumed to have been acquired, is specified by the corresponding ratio α_m of mean spectral pitches, $\bar{H}_m$ and $\bar{H}_1$:

$$\alpha_m = \bar{H}_m/\bar{H}_1 = m(1 + \bar{v}_m - \bar{v}_1) \tag{21}$$

It is supposed that α_m governs the relationships between harmonically related *pitches* in the same vein as m describes the corresponding relations between harmonic *frequencies*. α_m is an established parameter of the system. Hence, when a particular pitch H_a is given and another pitch H_b is required to be in that particular harmonic relation which is labeled by m (e.g., $m = 2$ corresponds to the octave relation), then H_b is obtained by

$$H_b = \alpha_m H_a = mH_a(1 + \bar{v}_m - \bar{v}_1) \tag{22a}$$

if H_b is meant to be a 'harmonic', i.e., $H_b > H_a$. Correspondingly,

$$H_b = \alpha_m^{-1} H_a \approx m^{-1} H_a(1 - \bar{v}_m + \bar{v}_1) \tag{22b}$$

if H_b is meant to be a 'sub-harmonic', i.e., $H_b < H_a$. $\bar{v}_1$ and $\bar{v}_m$ are specified by Eqn. 20. In the case where H_b is a sub-harmonic of H_a, it is advisable to replace f_1 in Eqn. 20 with f_a/m, where f_a is the frequency of the component which evokes H_a. As f_a and H_a may numerically differ by only a few percent (namely, the pitch shift), one may in Eqn. 20b

also replace f_1/kHz with $m^{-1}H_a$/kpu without introducing a significant error.

While the acquisition of harmonic intervals as well as their application in the formation of virtual pitch are considered as corresponding to unconscious perceptual processes, the basic harmonic intervals (i.e., octave, fifth, etc.) play a role in conscious pitch evaluation as well. In particular, the conscious matching of two *successive* tones to establish a perceptually optimal octave interval provides a convincing experimental demonstration of the interval stretch phenomenon and thus enables another experimental check of the calculated data. The dashed curve in Fig. 11 depicts the stretch which is observed in the frequency relation between successive pure tones which have been matched in a listening experiment to establish an optimal octave leap (mean data from refs. 35, 33, 20). As the successive tones in that experiment are isolated, unaffected tones with an SPL value of approximately 60 dB, they may virtually be considered as standard tones such that the dashed curve actually demonstrates the existence of an octave stretch *in terms of pitch*. This result is compared in Fig. 11 with those octave intervals which are implied in the series of acquired pitch intervals (solid curves, calculated from Eqn. 20). The existence region of each curve is restricted by the principle that they are based on analysis of voiced-speech spectra, i.e., fundamental frequencies of essentially 50–500 Hz. For example, the octave established by ($m = 1$ vs. $m = 2$) is existent in the higher-tone frequency region (abscissa of Fig. 11) of 100–1000 Hz; that established by ($m = 2$ vs. $m = 4$) in the region 200–2000 Hz, etc. There is a pronounced similarity between the *mean* value of the acquired octave stretches (solid curves) and the experimentally observed octave stretch (dashed). This suggests that the stretch of the acquired octave intervals is involved in the conscious evaluation of the octave as a musical interval.

In the final system of pitch extraction, the acquisition process as such will no longer be of interest. Rather, the system's essential parameter, the harmonic pitch ratio α_m (Eqns. 20 and 21), will be used as an established system parameter. Following the path of virtual-pitch recognition (Fig. 1), the selection of determinant spectral pitches now has to be considered.

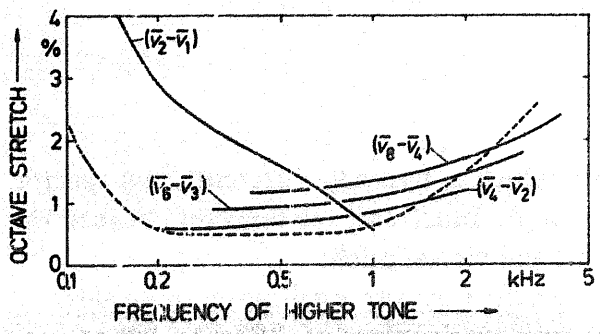


Fig. 11. Octave stretch, i.e. relative departure of the pitch ratio from 2 : 1. Dashed curve: Mean representation of results which are obtained by auditory octave matches of successive pure tones (from refs. 35, 33, 20). Solid lines: Stretch of the octave intervals which are constituted by the series of acquired harmonic pitch intervals; calculated from Eqn. 20.

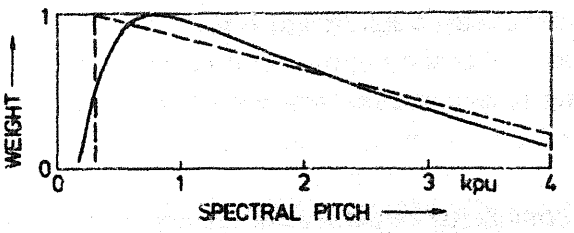


Fig. 12. Schematic illustration of the weight by which a particular spectral pitch may contribute to virtual pitch (dominance principle; solid curve). The dashed function represents an approximation, which illustrates the principle of selection realized in the present system.

SELECTION OF DETERMINANT SPECTRAL PITCHES

From the spectral components of a complex stimulus, only a few actually contribute to the formation of virtual pitch. The most essential part of the selection process is implied in the previous formation of spectral pitches, as only those components which produce corresponding spectral pitches may eventually contribute to the formation of virtual pitch.

Another selective effect is caused by the existence of a *dominant region* of spectral pitch. The existence of that region may be considered as being caused by the tendency of the 'central processor' (i.e., the subsequent formation of virtual pitch) to optimize its own performance by operating in a frequency region where relevant peripheral parameters such as threshold of hearing and frequency resolution are optimal. This process of optimization seems to be rather complex and, in particular, stimulus-dependent such that a precise and general description as yet appears to be hardly possible. However, its essential effect may be easily described by a pitch-dependent weight which specifies the extent to which a particular spectral pitch contributes to the formation of virtual pitch. A schematic and somewhat speculative illustration of that weight is given in Fig. 12 (solid curve) indicating that very low and very high spectral pitches do not contribute to virtual pitch.

The principle, after which the determinant spectral pitches are selected in the present system, is outlined by the dashed curve in Fig. 12 which is a simplification of the first:

(a) Spectral pitches below 300 pu and beyond 4 kpu are considered to be irrelevant;

(b) From the remaining spectral pitches, the lowest one is most determinant while the following contribute with descending weight.

FORMATION OF VIRTUAL PITCH

Though virtual pitch is a perceptual entity which is basically different from spectral pitch, there is a well-established and rather simple functional relationship between the determinant spectral pitches of a complex signal and virtual pitch:

(a) In first approximation, virtual pitch is specified by any small pitch interval (small region) which comprises at least two subharmonic pitch values of different determinant spectral pitches.

(b) The precise virtual-pitch magnitude is specified by that subharmonic pitch magnitude within the crucial interval which pertains to the most determinant spectral pitch.

The implications of these rules may be illustrated by the following example which represents a simple pitch-extracting procedure. It is assumed that the stimulus consists of three components with frequencies of 520, 620 and 720 Hz, respectively, which produce corresponding spectral pitches. It is also assumed that pitch shifts and interval stretch may be ignored because of being relatively small, i.e., it is considered as sufficient to extract the *nominal* virtual pitch instead of the true one. The determinant spectral pitches may then be represented numerically by the component frequencies. A table of potential virtual pitches is obtained by dividing each component frequency by m ($m = 1, 2, \ldots M$). In Table I, the result is shown for $M = 8$. By visual inspection one finds three subharmonic values in the small interval 102.9–104.0 Hz, indicating the approximate nominal virtual pitch. As the first component (520 Hz) is considered as being most determinant, the final virtual pitch is specified by the fifth subharmonic of that component, i.e., 104 Hz.

Depending on the width of the 'integrating interval', one may find *several* near-coincidences of at least two subharmonics. These actually are perceptually relevant as virtual pitch is somewhat ambiguous. However, one of the potential virtual pitches can usually be identified as being the most pronounced and thus most significant one. One may consider as being most significant that virtual-pitch value which

(a) is indicated by the integrating interval comprising the greatest number of near coincidences;

(b) corresponds to the smallest subharmonic number m; and

(c) can be obtained by the smallest integrating interval.

There are various ways in which these principles may be transformed into an algorithm for the schematic extraction of virtual pitch. A basic version of the procedure which finally was chosen in the present study, is depicted in Fig. 13. Input parameters are R determinant spectral pitches H_j ($j = 1, 2, 3, \ldots R$), arranged such that $H_1 < H_2 < H_3 <$

TABLE I

SUBHARMONIC FREQUENCIES OF THE COMPONENTS 520, 620 AND 720 Hz

m	Component frequency (Hz)		
	520	620	720
1	520.0	620.0	720.0
2	260.0	310.0	360.0
3	173.3	206.7	240.0
4	130.0	155.0	180.0
5	*104.0*	124.0	144.0
6	86.7	*103.3*	120.0
7	74.3	88.6	*102.9*
8	65.0	77.5	90.0

Near coincidence indicates an approximate nominal virtual pitch (italicized values). Nominal virtual pitch is specified by that of the nearly coincident values which pertains to the lowest component, i.e. 104.0.

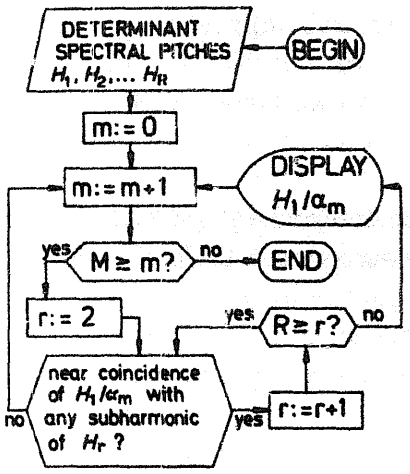


Fig. 13. Flow-chart which specifies the principle of formation of virtual pitch from a number R of determinant spectral pitches (H_1 to H_R). m is the subharmonic number pertinent to H_1. In this version a virtual pitch is displayed if R subharmonic near-coincidences have been detected.

$\ldots < H_R$. The system will detect near-coincidences of R subharmonics and display the true virtual pitch. Near-coincidences of less than R subharmonics are not detected by this basic version. The system begins with the subharmonic number $m = 1$ and examines whether there is near-coincidence of H_1/α_m with *any* subharmonic of H_2. If the result is negative, there is no chance for R coincidences with this particular m, and the procedure is repeated with $m = 2$, etc. If the result of the examination is positive, it is repeated with the subharmonics of H_3, and so on. After detection of that subharmonic number m of H_1, at which there exist R near-coincidences, the subharmonic pitch value of the most determinant spectral pitch, i.e. H_1/α_m, is displayed as the most significant virtual pitch. Thereafter the procedure may be continued with the next m value to find eventually another secondary virtual pitch, etc. When m exceeds a certain maximal number M, the process is finished.

The parameter M (highest subharmonic number of lowest determinant component) is determined by the extent to which harmonic intervals have been acquired in the learning mode. That extent is limited by the finite resolution of spectral components; however, as Fig. 9 illustrates, there does not exist a sharp boundary toward higher harmonic numbers but rather a decreasing probability of the harmonics to be relevant which approaches zero at about the 20th harmonic. The present algorithm accounts for these characteristics when M is chosen to be 8–10. The highest harmonic number which actually may be involved in specifying the precise virtual pitch, i.e. M, is confined to those acquired harmonic intervals which can be considered as being pronounced sufficiently, while in the detection of near-coincidences (i.e. determination of approximate virtual pitch) higher harmonic numbers (corresponding to possibly less pronounced harmonic intervals) are used as well.

The examination of near-coincidences can be carried out rather efficiently by the following approach. As the 'integrating pitch interval' (i.e., the interval in which near-coincidences occur) does not specify precisely the ultimate virtual pitch but indicates only an approximate value, the examination can be carried out in terms of the corresponding fre-

quencies rather than spectral pitches. This yields considerable simplifications as the criterion of near-coincidence may be specified as follows.

The integrating interval may be considered as a representation of that interval width in which slightly different pitches perceptually fuse. This interval may be approximately described by a constant percentage (a few percent) of its mean frequency. Hence an appropriate definition of the coincidence criterion is

$$-\delta \leqslant (f_q/m) \cdot (n/f_r) - 1 \leqslant +\delta \tag{23a}$$

where f_q and f_r are the frequencies of the components whose subharmonic pitches are examined for coincidence ($f_q < f_r$), m is the subharmonic number pertinent to f_q ($m = 1, 2, 3, \ldots M$), and n that pertinent to f_r. δ is a small value which represents the width of the integrating interval ($\delta = 0.01$–0.05). Another, equivalent representation of Eqn. 23a is

$$(1 - \delta)\, mf_r/f_q \leqslant n \leqslant (1 + \delta)\, mf_r/f_q\,. \tag{23b}$$

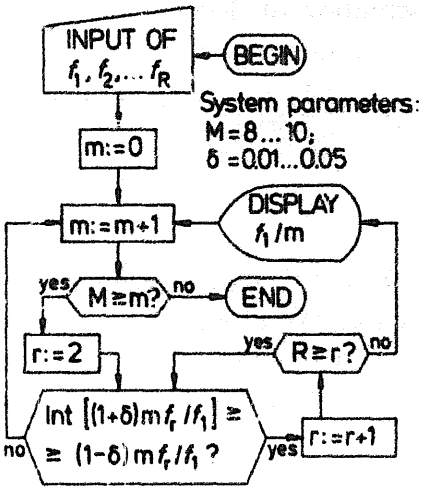
Taking advantage of the fact that m and n are integer numbers, Eqn. 23b can be used to examine the existence of near-coincidence *without checking every combination of* m *and* n. When a particular value of m is chosen, the examination is reduced to the question of whether there exists any integer number n in the interval $(1 - \delta)m\, f_r/f_q \ldots (1 + \delta)m\, f_r/f_q$. This can be accomplished for example by examining whether the condition

$$\mathrm{Int}\{(1 + \delta)\, mf_r/f_q\} \geqslant (1 - \delta)\, mf_r/f_q \tag{24}$$

is fulfilled. (The operation Int(x) means 'integer part of decimal number x'.) If Condition 24 is fulfilled, this indicates near-coincidence of f_q/m with any subharmonic of f_r. The resulting virtual pitch is not affected by this particular method of coincidence detection because it still is specified by H_1/α_m.

In many potential applications, only *nominal* virtual pitch is of interest, and pitch shifts and interval stretches may be ignored. In that case, the input parameters of the algorithm are the frequencies f_1 to f_R of the determinant components, and nominal virtual pitch is numerically equal to f_1/m. Fig. 14 shows the flow-chart of the procedure appropriate to this case. With $R \leqslant 3$ it can be implemented even on small programmable electronic calculators such as TI 57 (Texas Instruments) and HP 25 (Hewlett Packard). The pitch-extraction problem which was approached in TableI is automatically solved on these calculators within 10–15 s. Taking into account that in many (possibly even the majority of) cases the number of really determinant components hardly exceeds $R = 3$ and that the previous selection of determinant components from the original signal parameters may to some extent be carried out by 'guessing' (consulting the criteria described), this simple algorithm turns out to be a rather powerful tool. It may also be effectively used in real-time pitch extraction by any digital system.

The choice of the parameter δ is not critical but may be optimized with respect to the particular purpose and type of input signal. When it is required to obtain quickly and unambiguously *one* result which is representative of the most pronounced pitch of a



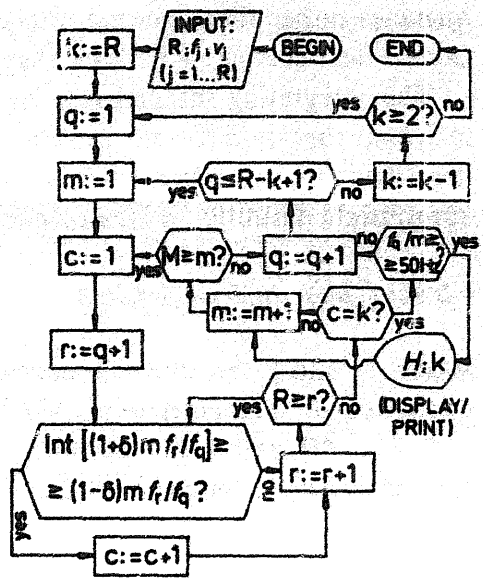


Fig. 14. Flow-chart of an extractor of nominal virtual pitch (f_1/m) from the frequencies (f_1 to f_R) of determinant components.

Fig. 15. Generalized version of the virtual-pitch extractor. The system also starts with detecting R coincidences, but then proceeds with detecting $(R-1)$, $(R-2)$, etc., coincidences. Nominal virtual pitch (Eqn. 25) and/or true virtual pitch (Eqn. 28) are displayed, in addition to the number of coincidences, k. Extraction of true virtual pitch requires previous determination of the pitch shifts v_j.

harmonic signal (as, for example, in voice-pitch extraction), it is advisable to choose $\delta = 0.01-0.015$. When it is required to obtain a survey of less pronounced pitch values as well, in particular in case of inharmonic signals, it is advisable to choose $\delta = 0.02-0.05$.

In the analysis of the pitch qualities of highly complex signals such as mixtures of complex tones and considerably inharmonic sounds, it may be desirable to determine not only those virtual pitches which correspond to R near-coincidences (such that each of the R determinant spectral pitches contributes to the extracted virtual pitch) but also those which correspond to coincidences of minor order because these are perceptually relevant in principle as well. In Fig. 15 the flow-chart of a procedure is shown by which this can be accomplished. The value k specifies the number of near-coincidences to be detected. In a first run, the procedure is carried out with $k = R$ and operates essentially in the same way as that previously described. When the first run is finished ($m > M$), k is reduced by one and detection of $(R-1)$ coincidences takes place, and so on. The procedure is finished when $k < 2$. When a particular virtual pitch pertinent to k coincidences has been displayed already, the same (or nearly the same) pitch is obtained by $(k-1)$ coincidences as well. These redundant results are to a considerable extent eliminated by this algorithm. Complete automatic elimination would increase the complexity of the procedure and was not considered as necessary *.

As outlined already, nominal virtual pitch is obtained by

$$\underline{H}_{\text{nom}} = m^{-1} f_q/\text{Hz pu} \qquad (25)$$

* The variety of displayed results was further reduced by eliminating values below 50 pu.

where f_q is the first of the determinant component frequencies. When the system depicted in Fig. 15 is confined to extraction of $\underline{H}_{nom}$ (which is sufficient in many cases), it may be implemented on medium-size programmable electronic calculators such as HP 29C, HP 67/97 and TI 58.

When it is required to account for the effects of pitch shifts and interval stretch as well, i.e., 'true' virtual pitch is to be detected, the pitch shifts v_i of the determinant components have to be calculated first and entered into the system in addition to the corresponding frequencies (Fig. 15). The true virtual pitch $\underline{H}$ is explicitly specified as follows. Using Eqns. 14, 18 and 21, there is

$$\underline{H} = \{f_q/\text{Hz}(1 + v_q)\}/\{m(1 + \bar{v}_m - \bar{v}_1)\} \text{ pu} \tag{26a}$$

which with sufficient approximation can be replaced with

$$\underline{H} = m^{-1} f_q/\text{Hz}\{1 + v_q - (\bar{v}_m - \bar{v}_1)\} \tag{26b}$$

In the case of $m = 1$, $(\bar{v}_m - \bar{v}_1)$ is zero and Eqn. 26b is sufficient to specify $\underline{H}$ while v_q is obtained from Eqns. 14, 15 and 17. In the case of $m \geqslant 2$, $\bar{v}_m$ is provided by Eqn. 20b. By replacing in Eqn. 20a, b the fundamental frequency f_1 with f_q/m, and introducing Eqn. 20a, b into Eqn. 26b, one obtains

$$\underline{H} = m^{-1} f_q/\text{Hz}[1 + v_q - 10^{-3}\{18 + 2.5m - (50 - 7m)\, m^{-1} f_q/\text{kHz} + 0.1(m^{-1} f_q/\text{kHz})^{-2}\}] \text{ pu} \tag{27}$$

$m = 2, 3, 4, \ldots M$; $50 \text{ Hz} \leqslant f_q/m \leqslant 500 \text{ Hz}$. When $\underline{H}$ is determined by a calculator program, it may be convenient to specify it by only one equation which takes account of the cases $m = 1$ and $m \geqslant 2$ as well. This can be accomplished by using the sign(x) function:

$$\underline{H} = m^{-1} f_q/\text{Hz}[1 + v_q - \text{sign}(m - 1)\, 10^{-3}\{18 + 2.5m - (50 - 7m)\, m^{-1} f_q/\text{kHz} + 0.1(m^{-1} f_q/\text{kHz})^{-2}\}] \text{ pu} \tag{28}$$

$m = 1, 2, 3, \ldots M$; $50 \text{ Hz} \leqslant f_q/m \leqslant 500 \text{ Hz}$. Eqn. 28 completes the stock of formulae and algorithms which are required for the schematic and automatic calculation of virtual pitch from the frequencies and SPL values of any tonal complex, including many subtle but significant auditory phenomena. In Fig. 16 the flow-chart of the complete system is shown. It mainly consists of the procedures required for 'formation of spectral pitch' and 'selection of determinant spectral pitches' while the box labeled 'virtual pitch extraction' represents the part shown in Fig. 15. This part may as well be replaced by the procedure shown in Fig. 14, if greater simplicity is preferred at the expense of less flexibility. The system operates essentially as follows. In addition to the number N of components, the frequencies f_i and SPL's L_i have to be entered paying attention to $f_1 < f_2 < f_3 < \ldots < f_N$. The first component whose frequency f_μ is beyond 300 Hz is selected. The relative amplitude A'_μ which is produced by the lower components at the frequency f_μ is then calculated as specified by the first sum in Eqn. 7. Thereafter the corresponding procedure

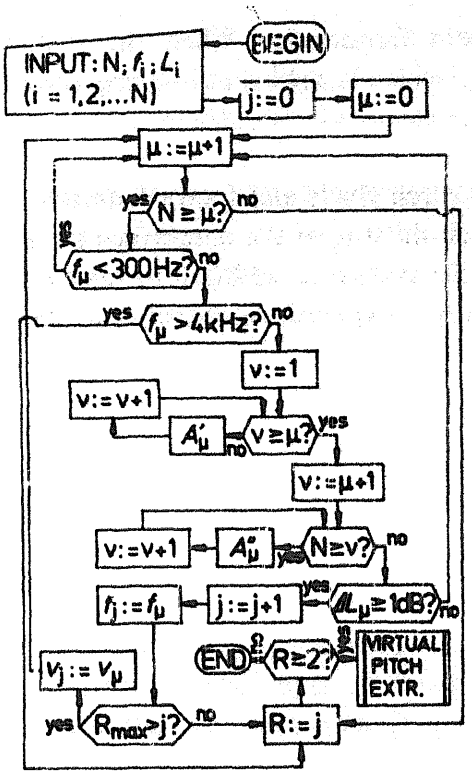


Fig. 16. Flow-chart of the entire system for calculating virtual pitch. The frequencies and SPL values of a number N of stimulus components have to be entered. The system extracts the determinant spectral pitches, calculates the individual pitch shifts v_j thereof, and finally extracts nominal and/or true virtual pitch. The box labeled 'virtual pitch extractor' corresponds to Fig. 15. It may be replaced with the simpler system of Fig. 14 as well.

takes place with A''_μ, the relative amplitude of the higher components (second sum in Eqn. 7). Then the SPL excess ΔL_μ is calculated by Eqn. 8. When ΔL_μ exceeds 1 dB, the μth component is considered as a determinant one. (The decisive minimal value of ΔL_μ is not very critical; reasonable results are obtained with 0–3 dB.) In this case, $f_j = f_\mu$ is stored and represents one input data of the final extractor (Fig. 15). Also, the pitch shift v_μ pertinent to that component is calculated (Eqns. 14–17), and stored ($v_j = v_\mu$). The procedure is repeated with the next component ($\mu + 1$), until either

(a) the highest component has been examined ($\mu > N$),

(b) the maximally relevant frequency of 4 kHz is exceeded, or

(c) a certain maximal number R_{max} of determinant components is attained.

When one of these conditions has been fulfilled, ($R \leqslant R_{max}$) pairs of component data (f_j; v_j) are available and the final virtual-pitch extractor is entered, which detects and displays H, as described (actually, v_j with $j = R_{max}$ is not required).

In addition to the parameters M and δ, R_{max} in this system is another parameter which may be chosen freely within reasonable limits. It was found that in most cases $R_{max} = 3$ is not only sufficient but optimal. It is only for the purpose of carefully and extensively studying the pitch qualities of highly complex stimuli, that it may be advisable to chose $R_{max} > 3$.

The complete system, including the flexible virtual-pitch extractor shown in Fig. 15,

was implemented on the programmable calculator TI 59. This program provides extraction of true and nominal virtual pitch from maximally $N = 11$ components, while R_{max} may be chosen within $2 \leqslant R_{max} \leqslant 5$. As a minor simplification, the formula Eqn. 1, which specifies the absolute threshold of hearing, was reduced in this program to the first term, $3.64\ (f/\text{kHz})^{-0.8}$. This term accounts sufficiently well for the influence of the hearing threshold in the frequency region in which the system operates ($f \leqslant 4$ kHz).

EXAMPLES

It may be helpful to demonstrate the system's performance and to suggest its potential applications by some calculated examples. In the calculations, the system parameters usually were $M = 10$; $\delta = 0.04$; $R_{max} = 3$, and from the series of displayed potential virtual pitches only the first is taken into consideration unless otherwise indicated.

Extraction of virtual pitch from harmonic complex tones is the most elementary and yet very important case, being identical with 'fundamental-frequency extraction'. As an example, consider a speech vowel. From Figs. 4 and 9 it is evident that the system will extract with virtually 100% probability at least three spectral pitches in the region beyond 300 pu. Hence the performance in this case essentially depends on that of the final virtual-pitch extractor of which the simple version shown in Fig. 14 is sufficient. If one implements this algorithm for example on a programmable calculator, one readily finds that it perfectly extracts the correct fundamental frequency from three almost arbitrarily chosen harmonic frequencies provided that:

(a) the frequencies of the three selected harmonics are available with an error of less than δ;

(b) the lowest selected harmonic does not exceed M; and

(c) not all of the three harmonic numbers are even (in which case the extracted fundamental frequency is an octave higher).

It is not required that the selected components are *successive* harmonics. As mentioned already, δ should be chosen $\leqslant 0.015$ when straightforward 'extraction of fundamental frequency' is desired because the first display may otherwise be an inharmonic one (being perceptually relevant as well, but secondary), while the fundamental frequency occurs in one of the next displays.

The departure of true virtual pitch from nominal virtual pitch is depicted in Fig. 17a for a full harmonic complex tone (comprising all harmonics up to 4 kHz with equal amplitudes), a residue tone, obtained from a harmonic complex tone by high-pass filtering with 1 kHz cut-off frequency, and another residue with 2 kHz cut-off frequency. Evidently the calculated points (filled symbols) represent the experimentally observed effect as a function of fundamental frequency reasonably well (open symbols; from ref. 20). In most cases the departure is negative as the true virtual pitch tends to be lower than nominal virtual pitch. The amount of the departure increases with ascending cut-off frequency and decreases with ascending fundamental frequency. Fig. 17b depicts the calculated pitch departure of a full complex tone, a 1.5 kHz residue and a 3 kHz residue as a function of SPL. The full complex tone with 300 Hz fundamental frequency depends essentially in the same way on SPL as a corresponding pure tone, i.e., its pitch descends slightly with increasing SPL. The same tendency, but distinctly smaller, occurs with a 1.5 kHz residue

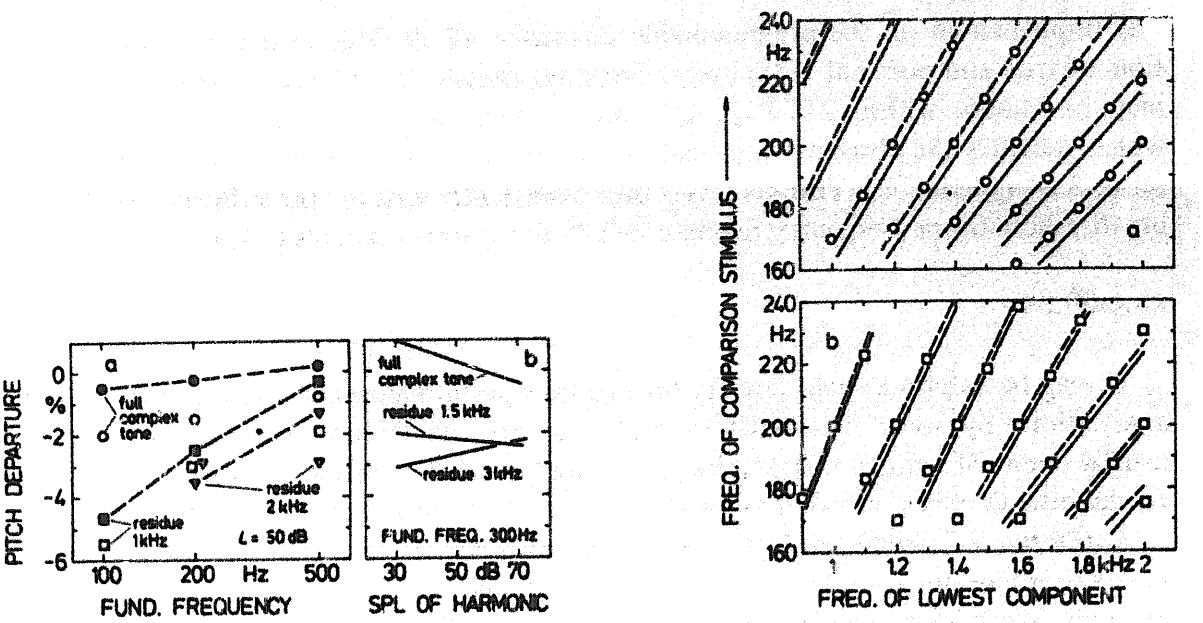


Fig. 17. Departure of true from nominal virtual pitch (Eqns. 28 and 25). (a) Influence of fundamental frequency (abscissa) and high-pass filtering (parameter) of harmonic complex tones (full complex tone; 1 kHz residue; 2 kHz residue) with 50 dB SPL. Filled symbols: calculated; open symbols: experimental data (mean values of 7 subjects; from ref. 20). (b) Influence of SPL, calculated.

Fig. 18. Calculated true virtual pitch (solid lines) and nominal virtual pitch (dashed) of three-component stimuli (a), and two-component stimuli (b), with a constant frequency distance $\Delta f = 200$ Hz between adjacent components, as a function of the lowest component's frequency. Virtual pitch is expressed in terms of equivalent frequency (ordinate, see text). In case of two-component stimuli, the additional presence of third-order and fifth-order combination tones was supposed. Circles represent experimental data of Schouten et al. [13], and squares those of Smoorenburg [17].

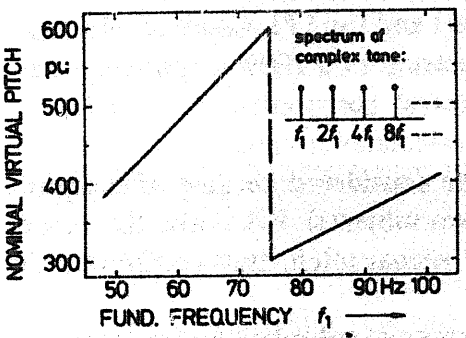
while the tendency is reversed with the 3 kHz residue. These results correspond well to experimental data [34,26]. When the fundamental frequency is below 300 Hz, the true virtual pitch of a full complex tone is virtually independent of SPL, a result which also is in agreement with experimental observations.

When all the components of a harmonic residue are shifted in frequency by the same amount, a particular type of *inharmonic residue tones* is obtained, and virtual pitch changes in a typical way. Schouten et al. [13] determined this effect by means of a three-component residue which was matched in pitch with another truly harmonic residue whose components were in the same frequency region as the inharmonic ones. As the pitch departure (the difference between true and nominal virtual pitch) of this particular test stimulus is virtually identical with that of the comparison stimulus, the fundamental frequency of the harmonic comparison stimulus, which is obtained by pitch matching, is numerically identical with the *nominal* virtual pitch of the test stimulus. If a standard tone would be used as a comparison stimulus, its frequency would by definition be numerically equal to the true virtual pitch of the test stimulus, and thus would tend to be slightly lower. These relationships have to be taken into consideration when the experimental results are compared with calculated virtual pitch, as shown in Fig. 18a. The solid lines represent the true virtual pitch (expressed in frequency of standard tone) of a three-

component residue whose components are 200 Hz apart and the SPL values of which are 54, 60 and 54 dB, respectively (i.e. the amplitude spectrum of a 100% amplitude modulated tone), as a function of the frequency of the lowest component (i.e., carrier frequency minus 200 Hz). The dashed lines represent the calculated nominal virtual pitch. In these calculations, several successive displays had to be considered because of the pitch ambiguity. The circles are median results of ref. 13 (two subjects). Evidently, the experimental data agree excellently with calculated *nominal* virtual pitch, thus confirming the procedure's validity.

In Fig. 18b experimental data on *two-component stimuli*, published by Smoorenburg [17], are compared with calculated results. Following Smoorenburg's observations and conclusions, it was presumed that in case of two-component stimuli (frequencies f_1 and f_2) combination tones with the frequencies $(f_1 - \Delta f)$ and $(f_1 - 2\Delta f)$ are additionally involved ($\Delta f = f_2 - f_1$). Taking experimental results on the strength of combination tones into consideration [17,38,4], the third-order combination tone (frequency $f_1 - \Delta f$) was represented by an additional spectral component with 35 dB SPL, and the fifth-order combination tone $(f_1 - 2\Delta f)$ by another component with 20 dB SPL, while the 'external' stimulus consisted of two components with $\Delta f = 200$ Hz and individual SPL values of 50 dB. The system actually extracts the combination tones (provided that their frequency exceeds 300 Hz) such that they play a determinant role. Fig. 18b depicts the calculated true and nominal virtual pitch, respectively, as a function of the frequency f_1 of the lower 'external' component. There is good agreement with Smoorenburg's data (squares), giving support to his suggestion concerning the role of combination tones in the particular case of two-component stimuli.

When a harmonic complex tone consists only of components which are an octave apart, i.e., $f_1, 2f_1, 4f_1$, etc., the *'Shepard pitch phenomenon'* is observed, which recently has been reconsidered by Pollack [12] (cf. ref. 16). Essentially, the phenomenon is characterized by the observation that under certain conditions the complex tone may be perceived as descending in pitch although the fundamental frequency actually is increased, and vice versa. In Fig. 19 the calculated nominal virtual pitch is depicted as a function of f_1. First of all, it is apparent that the calculated virtual pitch corresponds to frequencies which are considerably higher than the fundamental frequency. As a function of f_1, virtual pitch at first ascends monotonically but jumps down by an octave when f_1 approaches 75 Hz. Thereafter virtual pitch ascends again. Hence, when for example a complex tone with 60 Hz fundamental frequency is compared with a 70 Hz complex tone, the calculated virtual pitch ascends; but when the 60 Hz tone is compared with an 80 Hz tone, virtual pitch descends. This essentially is the 'Shepard-pitch effect'. It is caused by the principle that the low components of the stimulus are ignored in virtual-pitch formation while virtual pitch is determined by the lowest component which exceeds 300 Hz. This component is followed by the system when frequency is continuously increased, but when the next lower component exceeds 300 Hz, virtual pitch 'locks in' to that one, etc. The phenomenon generally can be considered as being caused by the existence of the dominant frequency region (cf. Fig. 12). It should be noticed that the simplified weighting function which has been chosen to represent the most essential consequences of the dominance principle (dashed curve in Fig. 12) is not sufficient to account for every detail of the Shepard-pitch effect, especially its stochastic aspects. It is obvious,



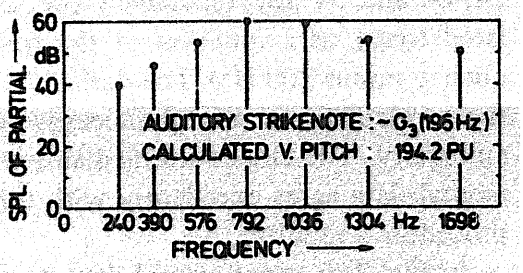


Fig. 19. Calculated nominal virtual pitch of a harmonic complex tone comprising only harmonics which are an octave apart (see insert), as a function of fundamental frequency. The downward pitch jump at 75 Hz fundamental frequency reflects in a simplified way the 'Shepard-pitch phenomenon'.

Fig. 20. Amplitude spectrum of a tubular bell (from ref. 6). The strongest partial has been arbitrarily assigned an SPL value of 60 dB. Calculated nominal virtual pitch, 195.0 pu; true virtual pitch, 194.2 pu.

however, that the present system, even in its simplified form, reflects the essential features of the phenomenon.

As an example of *inharmonic, but musical, sounds,* in Fig. 20 the amplitude spectrum of a tubular bell is shown [6]. The perceived strike note is reported to be approximately G_3 which corresponds to 196 Hz in standard intonation. There is obviously no component with this frequency present in the spectrum. The system selects the second, third and fourth components as being determinant, and extracts $\underline{H} = 194.2$ pu and $\underline{H}_{nom} = 195.0$ pu as true and nominal virtual pitch, respectively.

Finally, an application of the calculation procedure to the perceptual *analysis of a musical cord* is considered. The chord is a C-major triad which was produced on an electronic organ and consists of three harmonic complex tones with the fundamental fre-

TABLE II

FREQUENCIES AND SPL VALUES OF A C-MAJOR CHORD CONSTITUTED BY THREE HARMONIC COMPLEX TONES IN EQUALLY TEMPERED INTONATION

i	f_i (Hz)	L_i (dB)
1	392.0	59
2	523.2	59
3	659.2	60
4	784.0	56
5	1046.4	56
6	1176.0	50
7	1318.4	56
8	1568.0	52
9	1569.6	52
10	1960.0	54

Calculated true virtual pitch is 126.6 pu; nominal virtual pitch 130.7 pu, corresponding to the fundamental note C_3.

quencies 392, 532.2 and 659.2 Hz, respectively. According to intonation in standard equal temperament, these fundamental frequencies are in the relation $1 : s^5 : s^9$, where $s = \sqrt[12]{2}$. Table II displays the frequencies and SPL values of the lower 10 components which have been obtained by spectral analysis (the SPL of the strongest component was arbitrarily assumed to be 60 dB). When $R_{max} = 5$ is chosen (as an example), the system extracts the first five components as spectral pitches, indicating that in particular the musical tones which constitute the chord are actually perceived. In addition, a pronounced virtual pitch of 126.6 and 130.7 pu (true and nominal virtual pitch) is extracted, which corresponds to the musical note C_3. This note plays actually a significant role in muscial perception, as it represents the 'fundamental note' or 'root' of the chord [28,27]. Hence, in addition to the aspect of pitch, the present calculation system provides a means for the evaluation of musical sounds in terms of their *harmonic* aspects, which are related to the phenomenon of musical consonance.

DISCUSSION

When the 'heart' of the present pitch-extraction procedure (Fig. 14) is considered in isolation from its psychoacoustic background and origin, its similarity to previous 'harmonic pitch extractors, such as Schroeder's 'period histogram system' [14] and Miller's 'HIPEX system' [8], becomes apparent (cf. also ref. 15). Hence, that algorithm may be considered as an efficient and universal version of that type of system. It is extremely suitable for implementation on any digital computer.

The entire pitch-extraction procedure may also be considered as a somewhat simplified representation of present psychoacoustic knowledge on pitch perception. It is a remarkable, though not unexpected, result of the present study that the classical concept of the masking patterns as a representation of the auditory systems's spectral resolving power is efficiently applicable to the description of pitch perception of complex stimuli. Critical evaluation of the system against its psychophysical background will nevertheless produce several questions and thus may stimulate further research.

There are some restrictions in the present system which do not significantly reduce its applicability but are of scientific interest and may be eliminated by additional work. First of all, it should be noted that the system is devoted only to virtual pitch, while the complex perceptual interaction and cooperation of spectral and virtual pitch, which in general yields the final conscious pitch percept, has not been considered. The mutual masking of spectral components, which plays a decisive role in the extraction of determinant spectral pitches, evidently depends to a considerable degree on the phase relations between the components. Though this effect fortunately is not critical for the system's performance, it may be incorporated in an improved version. In this way the slight influence of phase which is experimentally observed in the perception of virtual pitch may be accounted for to a considerable extent without changing anything in the principle. Another feature which may be subject to further research is the dominance principle including the more-or-less random process of decision between ambiguous virtual pitches.

APPENDIX

In Tables A1 and A2 calculator programs are provided which are to some extent representative of the two presently existing calculator systems. They may be useful for

TABLE A1

EXTRACTOR OF NOMINAL VIRTUAL PITCH (FIG. 14); CALCULATOR PROGRAM FOR TI 57

Fix parameters: $M = 9$; $\delta = 0.04$. Data entry: R into register 4 ($R = 2; 3$); $f_1, f_2, (f_3)$ into reg. 1, 2, (3).

CLR STO 5 Lbl 0 1 SUM 5 RCL 5 x/t 9 INV GE R/S RCL 4 – 1 = STO 0 RCL 2 Lbl 1: RCL 1 × RCL 5 = STO 6 × . 9 6 = x/t RCL 6 × 1 . 0 4 = Int INV GE GTO 0 Dsz GTO 2 RCL 1 : RCL 5 = R/S GTO 0 Lbl 2 RCL 3 GTO 1

Special notation: $x/t = x \rightleftharpoons t$; GE = $x \geqslant t$.

TABLE A2

EXTRACTOR OF NOMINAL VIRTUAL PITCH (FIG. 14); CALCULATOR PROGRAM FOR HP 25, etc.

Fix parameter: $M = 10$. Enter δ into reg. 4; R into reg. 0 ($R = 2; 3$); $f_1, f_2, (f_3)$ into reg. 1, 2, (3).

CLX STO 5 1 STO + 5 RCL 5 1 0 $x \geqslant y$ GTO 1 3 CLX R/S GTO 0 0 RCL 0 1 – STO 6 RCL 2 RCL 1: RCL 5 × STO 7 1 RCL 4 – × 1 RCL 4 + RCL 7 × INT $x < y$ GTO 0 3 1 STO – 6 RCL 6 $x = 0$ GTO 4 2 RCL 3 GTO 1 8 RCL 1 RCL 5 : R/S GTO 0 3

convenient and quick evaluation of nominal virtual pitch by means of the algorithm shown in Fig. 14.

ACKNOWLEDGEMENTS

The author is indebted to E. Zwicker and H. Fastl for contributing valuable comments on the manuscript. This work was carried out in the Sonderforschungsbereich Kybernetik, Munich, and supported by the Deutsche Forschungsgemeinschaft.

REFERENCES

[1] Egan, J.P. and Meyer, D.R. (1950): Changes in pitch of tones of low frequency as a function of the pattern of excitation produced by a band of noise. J. Acoust. Soc. Am. 22, 827–833.
[2] Fant, G. (1960): Acoustic theory of speech production. Mouton, The Hague. (2nd edn, 1970).
[3] Fletcher, H. (1934): Loudness, pitch, and timbre of musical tones and their relations to the intensity, the frequency and the overtone structure. J. Acoust. Soc. Am. 6, 59–69.
[4] Helle, R. (1969): Amplitude und Phase des im Gehör gebildeten Differenztones dritter Ordnung. Acustica 22, 74–87.
[5] Houtgast, T. (1974): Lateral suppression in hearing. Rep. Inst. Percept. TNO, Soesterberg, The Netherlands.
[6] Hueber, K.A. (1972): Nachbildung des Glockenklanges mit Hilfe von Röhrenglocken und Klavierklängen. Acustica 26, 334–343.
[7] Maiwald, D. (1967): Ein Funktionsschema des Gehörs zur Beschreibung der Erkennbarkeit kleiner Frequenz- und Amplitudenänderungen, Acustica 18, 81–92.
[8] Miller, R.L. (1970): Performance characteristics of an experimental harmonic identification pitch extraction (HIPEX) system. J. Acoust. Soc. Am. 47, 1593–1601.
[9] Miyazaki, K. (1977): Pitch-intensity dependence and its implications for pitch perception. Tohoku Psychol. Folia 36, 75–88.
[10] Plomp, R. (1964): The ear as a frequency analyzer. J. Acoust. Soc. Am. 36, 1628–1636.

[11] Plomp, R. and Mimpen, A.M. (1968): The ear as a frequency analyzer. II. J. Acoust. Soc. Am. 43, 764–767.
[12] Pollack, I. (1978): Decoupling of auditory pitch and stimulus frequency: the Shepard demonstration revisted. J. Acoust. Soc. Am. 63, 202–206.
[13] Schouten, J.F., Ritsma, R.J. and Cardozo, B.L. (1962): Pitch of the residue. J. Acoust. Soc. Am. 34, 1418–1424.
[14] Schroeder, M.R. (1968): Period histogram and product spectrum: new methods for fundamental frequency measurement. J. Acoust. Soc. Am. 43, 829–834.
[15] Seneff, S. (1978): Real-time harmonic pitch detector. IEEE Trans. ASSP 26, 358–365.
[16] Shepard, R. (1964): Circularity in judgments of relative pitch. J. Acoust. Soc. Am. 36, 2346–2353.
[17] Smoorenburg, G.F. (1970): Pitch perception of two-frequency stimuli. J. Acoust. Soc. Am. 48, 924–942.
[18] Terhardt, E. (1970): Frequency analysis and periodicity detection in the sensations of roughness and periodicity pitch. In: Frequency Analysis and Periodicity Detection in Hearing, pp. 278–287. Editors: R. Plomp and G.F. Smoorenburg. Sijthoff, Leyden.
[19] Terhardt, E. (1970): Oktavspreizung und Tonhöhenverschiebung bei Sinustönen. Acustica 22, 345–351.
[20] Terhardt, E. (1971): Die Tonhöhe Harmonischer Klänge und das Oktavintervall. Acustica 24, 126–136.
[21] Terhardt, E. (1971): Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon. Proc. 7. ICA, Budapest, Vol. 3, pp. 621–624.
[22] Terhardt, E. (1972) Zur Tonhöhenwahrnehmung von Klängen. I. Psychoakustische Grundlagen. Acustica, 26, 173–186.
[23] Terhardt, E. (1972): Zur Tonhöhenwahrnehmung von Klängen. II. Ein Funktionsschema. Acustica 26, 187–199.
[24] Terhardt, E. (1974): Pitch, consonance, and harmony. J. Acoust. Soc. Am. 55, 1061–1069.
[25] Terhardt, E. (1974): Pitch of pure tones: its relation to intensity. In: Facts and Models in Hearing, pp. 353–360. Editors: E. Zwicker and E. Terhardt. Springer, Heidelberg.
[26] Terhardt, E. (1975) Influence of intensity on the pitch of complex tones. Acustica 33, 344–348.
[27] Terhardt, E. (1977): The two-component theory of musical consonance. In: Psychophysics and Physiology of Hearing, pp. 381–390. Editors: E.F. Evans and J.P. Wilson. Academic Press, London.
[28] Terhardt, E. (1978): Psychoacoustic evaluation of musical sounds. Percept. Psychophys. 23, 483–492.
[29] Terhardt, E. and Fastl. H. (1971): Zum Einfluss von Störtönen und Störgeräuschen auf die Tonhöhe von Sinustönen. Acustica 25, 53–61.
[30] Verschuure, J. and van Meeteren, A.A. (1975): The effect of intensity on pitch. Acustica 32, 33–44.
[31] Von Helmholtz, H. (1863): Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik. F. Vieweg, Braunschweig.
[32] Walliser, K. (1969): Über die Abhängigkeiten der Tonhöhenempfindung von Sinustönen vom Schallpegel, von übergelagertem drosselndem Störschall und von der Darbietungsdauer. Acustica 21, 211–221.
[33] Walliser, K. (1969): Über die Spreizung von empfundenen Intervallen gegenüber mathematisch harmonischen Intervallen bei Sinustönen. Frequenz 23, 139–143.
[34] Walliser, K. (1969): Zusammenhänge zwischen dem Schallreiz und der Periodentonhöhe. Acustic 21, 319–329.
[35] Ward, W.D. (1954): Subjective musical pitch. J. Acoust. Soc. Am. 26, 369–380.
[36] Webster, J.C. and Muerdter, D.R. (1965): Pitch shifts due to low-pass and high-pass noise bands. J. Acoust. Soc. Am. 37, 382–383.
[37] Webster, J.C., Miller, P.H., Thompson, P.O. and Davenport, E.W. (1952): The masking and pitch shifts of pure tones near abrupt changes in a thermal noise spectrum. J. Acoust. Soc. Am. 24, 147–152.

[38] Zwicker, E. (1955): Der ungewöhnliche Amplitudengang der nichtlinearen Verzerrungen des Ohres. Acustica 5, 67–74.
[39] Zwicker, E. (1961): Subdivision of the audible frequency range into critical bands (Frequenzgruppen). J. Acoust. Soc. Am. 33, 248.
[40] Zwicker, E. (1976): Influence of a complex masker's time structure on masking. Acustica 34, 138–146.
[41] Zwicker, E. and Feldkeller, R. (1967): Das Ohr als Nachrichtenempfänger. Hirzel, Stuttgart.
[42] Zwicker, E. and Herla, S. (1975): Über die Addition von Verdeckungseffekten. Acustica 34, 89–97.