An optimum processor theory for the central formation of the pitch of complex tones

Julius L. Goldstein*

Department of Electrical Engineering and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 12 March 1973; 18 June 1973)

A comprehensive theory is formulated for the central formation of the pitch of complex tones, i.e., periodicity pitch [Schouten, Ritsma, and Cardozo, J. Acoust. Soc. Amer. 34, 1418–1424 (1962)]. This theory is a logical deduction from statistical estimation theory of the optimal estimate for fundamental frequency, when this estimate is constrained in ways inferred from empirical phenomena. The basic constraints are (i) the estimator receives noisy information on the frequencies, but not amplitudes and phases, of aurally resolvable simple tones from the stimulus and its aural combination tones, and (ii) the estimator presumes all stimuli are periodic with spectra comprising successive harmonics. The stochastic signals representing the frequencies of resolved tones are characterized by independent Gaussian distributions with mean equal to the frequency represented and a variance that serves as free parameter. The theory is applicable whether frequency is coded by place or time. Optimum estimates of fundamental frequency and harmonic numbers are calculated upon each stimulus presentation. Multimodal probability distributions for the estimated fundamental are predicted in consequence of variability in the estimated harmonic numbers. Quantification of the variance parameter from musical intelligibility data in Houtsma and Goldstein [J. Acoust. Soc. Amer. 51, 520–529 (1972)] shows it to be dependent upon the frequency represented and not upon other stimulus frequencies. The quantified optimum processor theory consolidates known data on pitch of complex tones.

Subject Classification: 4.11, 4.9.

INTRODUCTION

Complex periodic sounds in a normal speech or music context have a pitch that covaries with fundamental frequency independently of the presence or absence of energy in the sound spectrum at this frequency (Seebeck, 1841; Fletcher, 1924; Schouten, 1938). Classical theories have explained this phenomenon in terms of simple spectral cues (Helmholtz, 1863) or temporal cues (Schouten, 1940a) that could be measured in the output of the cochlear spectrum analyzer at single characteristic frequencies.

The inadequacy of these cochlear-based conceptions of pitch processing for complex tones was recently proved by musical intelligibility tests (Houtsma and Goldstein, 1972) with the discovery that the pitch of complex tones can be heard equally well when two-tone sounds are presented monotically (both constituent tones in one ear) or dichotically (one simple tone in each ear). Moreover, systematic intelligibility experiments with sounds comprising two successive harmonics demonstrated that only harmonics of relatively low order ($<\sim 10$) were effective in communicating musical pitch. In some circumstances, the low harmonics that communicate pitch are provided by aural combination tones (Goldstein, 1967a). These results, added to the behavioral measures (Helmholtz, 1863; Plomp, 1964) and physiological measures (Békésy, 1960; Kiang, 1965; Goldstein, Baer, and Kiang, 1971) of the acuity of aural frequency resolution suggested the hypothesis that the pitch of these complex tones is necessarily mediated by a central processor that operates effectively *only* with signals derived from aurally resolved simple tones. This paper presents a unifying psychophysical theory of the central processor of pitch of complex tones, which was developed logically from this hypothesis, data from our reported musical intelligibility experiments, and an accumulation of past research. The logic of this development was drawn from statistical estimation theory and was motivated by related applications by Siebert (1968, 1970) and Colburn (1969).

Basically, the central processor may be viewed as a recognizer of spectral patterns supplied by the peripheral frequency analyzers. Recognition is accomplished through selection of the best matching stored pattern or template, where the templates correspond to periodic signals comprising successive harmonics. The power of the present formulation lies in the efficacy of an idealized representation of the peripheral spectral patterns for complex tones. The peripheral frequency analyzer extracts from complex-tone stimuli all constituent simple tones that differ in frequency from their neighbors by more than some resolution limit. The input spectral patterns are then defined by recording with some random error only the frequency of each resolved simple tone. Periodicity pitch is formally assigned by using a maximum likelihood statistical estimation procedure to fit the template.

Knowledge of the physiological implementation of the pitch processor is not a necessary ingredient of the

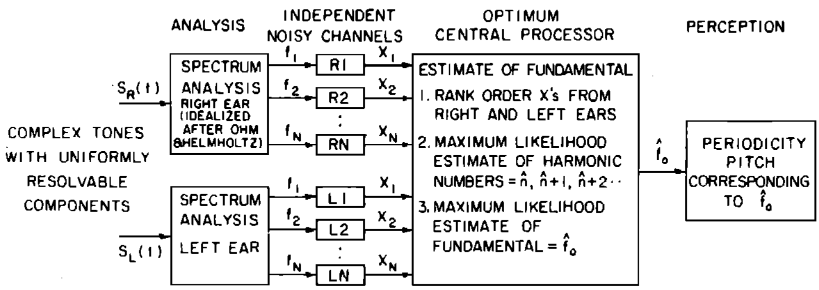

FIG. 1. Optimum processor theory of the central formation of the pitch of complex tones. The idealized spectrum analyzers supply the independent noisy channels with separate representations of each resolved spectral component in the stimulus. Information on frequency of each component, but not amplitude or phase, is conveyed stochastically to the optimum processor. The hypotheses of independent stochastic representations of the component frequencies and of optimum estimation of the fundamental are tested with psychophysical data. The stimuli are restricted to complex tones with spectral spacings such that aural resolution fails uniformly with increasing frequency of the components; viz, if any component is not resolved then no other component of higher frequency is separable from any neighboring component.

theory and is therefore treated as outside the scope of this report. It is important instead to appreciate the theoretical quantities and the logic defined by this pitch theory. The randomness with which the frequencies of resolved constituent tones are represented is found to be nearly independent of the frequency spacing between neighboring tones. This randomness is quantified in the theory by representing the frequency of each resolved simple tone with a sample from an independent Gaussian distribution centered on the true frequency. The variance of the Gaussian distribution is a function of the true frequency; it is the most fundamental quantity in the model.

Because of the stochastic nature of the frequency pattern seen by the central processor, the pitch estimated is a random function of any given complex tone stimulus. The most complete statement of predictions by the theory constitutes a probability distribution of periodicity pitch estimates for a given stimulus. In general, the probability distributions are found to be multimodal, where the modes are relatively narrow compared with their spacing; for some considerations these distributions may be treated as discrete.

For a periodic complex tone comprising successive harmonics the predicted probability distribution for periodicity pitch will include a principal mode centered on the true periodic frequency of the stimulus. Pitch estimates within the principal mode obtain when the harmonic numbers in the template are correctly aligned with the input spectral pattern. Misalignments generate secondary modes. The position, width, and probability of the various modes supply comprehensive, quantitative descriptions and predictions for periodicity pitch phenomena with complex tones.

In Sections I and II and Appendices A–C, the theory is deduced by assuming optimum processing of stimulus information, where this processing is constrained in ways inferred from empirical phenomena. Sections I and II-D supply an outline of the theory adequate for understanding the treatments of experimental phenomena in Secs. III–VII. These latter sections serve to demonstrate the existence of a unifying logic for behavioral phenomena on pitch of complex tones. Sections II-A–II-C and Appendices A–C contain the mathematical development of the theory based upon maximum likelihood estimation. A critical appraisal of the theory requires reading the mathematical development.

Sections VIII and IX discuss what the author believes to be the most relevant experimental and theoretical antecedants of the present work. Aural combination tones (Sec. VIII) misguided periodicity pitch theory for considerable time. Studies by many investigators over a long period of time provided the empirical bases for, and indeed in part anticipated, the constraints assumed in the present optimum processor formulation (Sec. IX). Finally, Sec. X offers suggestions for further work.

I. OUTLINE OF THEORY OF PITCH FORMATION

The heart of this theory is the hypothesis that a central processor makes an optimum estimate of the fundamental frequency on the basis of a noisy representation of the tonotopically organized stimulus (Fig. 1). The constraint on the central processor is that it presumes all stimuli are periodic (Schouten, 1940b; de Boer, 1956) with spectra comprising only successive harmonics. Estimates of both fundamental frequency ($\hat{f}_0$) and the harmonic numbers ($\hat{n}, \hat{n}+1, \hat{n}+2, \ldots$) must be obtained from the input to the processor. This constraint is natural for the stimuli comprising two unknown successive harmonics that were employed for our previously reported musical intelligibility experiments (Houtsma and Goldstein, 1972). This constraint, however, is maintained as a property of the central processor even for stimuli for which prior knowledge is incompatible, such as known harmonic numbers, or known inharmonicity.

At present the theory is restricted to complex-tone stimuli with spectral spacings that cause aural frequency resolution to fail uniformly, whenever it fails. In particular, if any component is not resolved no other component of higher frequency is separable from a higher or lower frequency component. This restriction is meant to exclude complex tones that could supply the central processor with useful information in the form of noncontiguous groups of tones (as discussed in Sec. X). By hypothesis, the central processor, when operating on the class of complex tones delineated, receives useful information from both ears only for spectral components that are individually resolved by the aural spectrum analyzer. Furthermore, it is hypothesized that (within limits) only information on frequency of each component but not on amplitude or phase is preserved in this measurement. Finally, it is hypothesized that frequency information for each aurally resolved simple tone is degraded in independent noisy channels. These last three hypotheses, (i) aural frequency resolution, (ii) retention only of frequency information for each resolved component, and (iii) independent noisy transmission of the frequency information, reduce to one hypothesis, namely, that the component frequencies in the stimulus are represented at the input to the central processor by independent stochastic signals. The present theory is a strictly mathematical development of the hypotheses of independent stochastic representations of component frequencies and of optimum estimation of the fundamental.

The stochastic frequency signals $\{X_k\}$ representing the aurally resolved simple tones with frequencies $\{f_k\}$ are characterized by static samples from independent Gaussian probability distributions each with mean f_k and standard deviation σ_k. This standard deviation is the only free parameter of the model. The finding of primary importance to be demonstrated in this paper is that known data on pitch of complex tones can be comprehensively treated with a standard deviation that is a function only of the frequency represented by each probability distribution, that is, $\sigma_k \sim \sigma[f_k]$, and this function appears to be similar among individuals.

Note in Fig. 1 that the simple view (circa 1863) of aural frequency analysis invoked here is all we need. Ohm (1843) provided us with the conception of aural signal decomposition in accord with Fourier's series. Helmholtz (1863) reminded us of the physical and psychophysical requirements for limited frequency resolution and gave the first evidence for aural insensitivity to the phases of resolved components. Conceptions of mechanisms for aural frequency resolution, however, need not be restricted to his peripheral bandpass filters, as neural processing could also be involved. The confounding role of aural combination tones will be treated in Secs. VII and VIII.

Note also that the concept that the optimum processor operates on signals representing the constituent frequencies of complex-tone stimuli does not necessarily imply the use of tonotopic or place information per se as the measure of frequency. For example, temporal periods of simple tones are not ruled out as the measure of frequency. To emphasize that this formal theory does not require place information per se, we assume that the frequency channels are not rank ordered or labeled according to characteristic frequencies. Thus the optimum processor must measure the rank order from the frequency signals. In principle, the rank order measured from unlabeled channels could differ from that indicated by labeled channels. In fact, however, this difference may be ignored because from the quantities that emerge from this analysis it is apparent that negligible probability of incorrect rank ordering results from ranking the stochastic samples of the frequency signals of interest. Periodicity pitch (Licklider, 1954) is suggested in Fig. 1 as an appropriate designation for the phenomenon of pitch of complex tones because the central processor presumes periodic stimuli, and not because temporal periods are presumed to mediate pitch as in earlier usage of the term.

II. MATHEMATICAL FORMULATION

A. Maximum Likelihood Estimation of Harmonic Numbers and Fundamental

The fundamental estimator $\hat{f}_0$ is given in this section for a two-component signal; this solution is extended in Sec. II-C to complex tones with more components. The stochastic frequency signals $\{X_k\}$, representing the aurally resolved simple tones with frequencies $\{f_k\}$ are characterized with static samples from Gaussian probability distributions with mean f_k and standard deviation σ_k. This standard deviation is the only free parameter of the model. A crucial theoretical question is whether empirical data on pitch of complex tones can be comprehensively described, interrelated, and predicted by treating the standard deviation as a function of a single variable, the frequency f_k. We show that it can, to a good approximation.

Given a two-component stimulus with resolvable tones of frequencies f_1 and f_2 $(f_2 > f_1)$, the transformations by the independent noisy channels (Fig. 1) are specified by Eq. 1. The assumption that the variance is a function of a single frequency variable is made for convenience in the mathematical development; it does not prejudice the test of this assumption in Sec. III:

$$\begin{aligned} f_1 &\rightarrow X_1, \quad pdf[x_1] = G[f_1,\sigma_1], \\ f_2 &\rightarrow X_2, \quad pdf[x_2] = G[f_2,\sigma_2], \end{aligned} \tag{1}$$

where

$$G[f_k,\sigma_k] \equiv (2\pi\sigma_k^2)^{-\frac{1}{2}} \exp[-(x_k - f_k)^2/2\sigma_k^2]$$

and

$$\sigma_k = \sigma[f_k].$$

The central processor makes an optimum estimate [maximum likelihood estimation (Van Trees, 1968)] of the unknown stimulus fundamental on the presumption that the stimulus frequencies are unknown successive harmonics. This estimate is obtained by choosing the unknown fundamental and harmonic numbers

$(\hat{f}_0;\hat{n},\hat{n}+1)$ to maximize the probability of the samples as given in Eq. 2,

$$L_{\max}=\left[\frac{1}{(2\pi)^{\frac{1}{2}}\hat{\sigma}_1}\exp-\frac{(x_1-\hat{n}\hat{f}_0)^2}{2\hat{\sigma}_1^2}\right]\times\left\{\frac{1}{(2\pi)^{\frac{1}{2}}\hat{\sigma}_2}\exp-\frac{[x_2-(\hat{n}+1)\hat{f}_0]^2}{2\hat{\sigma}_2^2}\right\},\quad(2)$$

where

$$\hat{\sigma}_k=\sigma[(\hat{n}-1+k)\hat{f}_0],\quad \sigma[\cdot]\text{ as in Eq. 1.}$$

Following standard procedure we replace Eq. 2 with a convenient monotonic transformation, its natural logarithm, Eq. 3:

$$\Lambda=-\ln 2\pi\hat{\sigma}_1\hat{\sigma}_2-(x_1-\hat{n}\hat{f}_0)^2/2\hat{\sigma}_1^2-[x_2-(\hat{n}+1)\hat{f}_0]^2/2\hat{\sigma}_2^2.\quad(3)$$

Equation 3 cannot be solved for the maximizing $\hat{n}$ and $\hat{f}_0$ without specifying the functional dependence of the standard deviation σ upon the estimated frequencies. The ideal system knows the form of this function. One approximate procedure is to assume that σ is constant in a local region, then calculate σ from empirical data on the basis of the approximate estimator, and finally iterate the calculations on the basis of this first approximation of σ. This procedure is abbreviated here by assuming a local behavior for σ which (after initial calculations) is known to approximate the σ function globally better than a constant; σ is taken as locally proportional to frequency:

$$\sigma[f_k]\sim K[f_k]\cdot f_k,\quad(4)$$

where $K[f_k]$ is locally constant. By substituting Expression 4 in Eq. 3 and maximizing with respect to $\hat{f}_0$, with $K[f]$ locally constant, we derive Eq. 5. (From calculated first approximations of σ, the first term on the right of Eq. 3 is known to be negligible.)

$$\hat{f}_0=\frac{(x_1/\hat{n})^2+[x_2/(\hat{n}+1)]^2}{x_1/\hat{n}+x_2/(\hat{n}+1)}.\quad(5)$$

The estimate of the fundamental, $\hat{f}_0$, is made on the basis of the frequency samples, (x_1,x_2), and the estimated harmonic numbers, $\hat{n}$ and $\hat{n}+1$. An optimum choice for $\hat{n}$ is given by choice of that integral $\hat{n}$ which maximizes Eq. 3. The sample space can be partitioned into regions for which different values of $\hat{n}$ are optimum. A simple solution for this partitioning obtains because for some values of (x_1,x_2) two different successive values of integral $\hat{n}$ give equal likelihoods. By substituting Eq. 5 in Eq. 3 and equating for $\hat{n}=m$ and $\hat{n}=m+1$, we get the desired partitioning Eq. 6. (The first term on the right of Eq. 3 is again negligible for relevant values of σ and m.)

$$x_2/x_1=[(m+2)/m]^{\frac{1}{2}}.\quad(6)$$

Figure 2 shows the estimation procedure of the optimum central processor for a two-tone signal. The

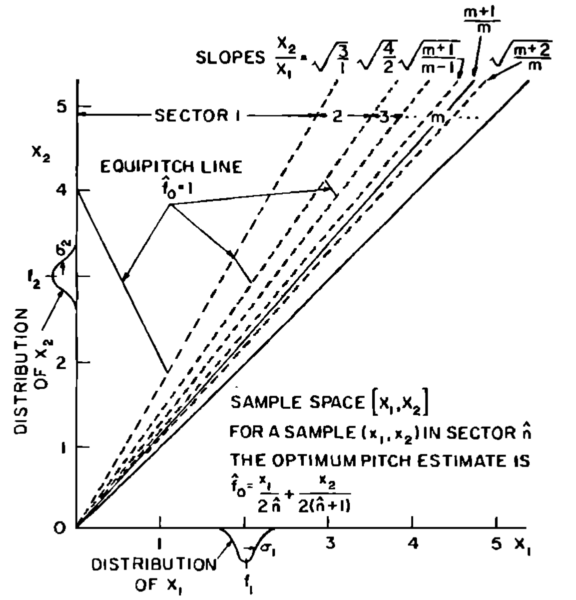

FIG. 2. Logical operation of the optimum processor for a two-component signal. Two independent Gaussian random variables X_1 and X_2 represent the frequencies f_1 and f_2, and define the two-dimensioned sample space. The sample space is partitioned into sectors such that all points within a given sector share a common maximum-likelihood estimate of harmonic numbers, $\hat{n}$ and $\hat{n}+1$. These harmonic numbers are used to calculate the maximum-likelihood estimate of fundamental frequency, $\hat{f}_0$. The discontinuity of this estimate for sample values at sector boundaries is illustrated by the breaks in the equipitch line for $\hat{f}_0=1$.

sample space $[X_1,X_2]$ is characterized by the independent Gaussian probability distributions of the frequency signals. Sectors for different harmonic numbers partition the sample space. Location of a sample point (x_1,x_2) in the mth sector fixes the estimate of the harmonic numbers as m and $m+1$. Finally, the fundamental is estimated by Eq. 5 which is approximated by the linear estimator 7 shown in Fig. 2. The linear estimator is derived by treating the variance function as locally constant when maximizing Eq. 3. The boundaries defined by Eq. 6 also apply to the linear estimator. The linear and quadratic estimators converge near the harmonic conditions $x_2/x_1\sim(\hat{n}+1)/\hat{n}$:

$$\hat{f}_0=x_1/2\hat{n}+x_2/2(\hat{n}+1).\quad(7)$$

An important property of the fundamental estimator is its discontinuity for sample points that traverse the boundaries of the sectors. This discontinuity is illustrated in Fig. 2 by the separation between lines of equal pitch (given by the linear estimator) in adjacent sectors. [*Note:* The theory allows for inversions $x_2<x_1$ for $f_2>f_1$, in which case the rank ordering specified in Fig. 1 exchanges the roles of x_1 and x_2 in the estimation procedure. However, for relevant signals and σ, the probability of inversions is negligible and is ignored in the remainder of this paper.]

B. Probability Distribution of the Estimated Fundamental

An accurate characterization of the optimum fundamental estimate $\hat{f}_0$ for signals with two resolvable spectral components is given by a Gaussian approximation of the estimates from each sector. The unimodal nature of the estimator within each sector can be deduced readily from Fig. 2. Beyond this fact, the degree of accuracy of the Gaussian approximation, though high, is not an issue here, because the pitch theory requires no more detail on the probability functions than is required for the Gaussian approximation. Given the signal frequencies and their standard deviations, one can calculate the probability that a sample falls in sector m, $\Pr[\hat{n}=m]$, the expected fundamental estimate in sector m, $\bar{f}_{0m}$, and the variance of this estimate, σ_{0m}^2. These quantities define a multimodal Gaussian characterization of $\hat{f}_0$ (Eq. 8),

$$pdf[\hat{f}_0]\cong\sum_{m=1}^{\infty}\Pr[\hat{n}=m]\cdot G[\bar{f}_{0m},\sigma_{0m}],$$

where

$$\Pr[\hat{n}=m]\equiv\mathrm{Prob}[\hat{n}=m\,|\,f_1,f_2,\sigma_1,\sigma_2],\qquad(8)$$
$$\bar{f}_{0m}\equiv E[\hat{f}_0\,|\,\hat{n}=m,f_1,f_2,\sigma_1,\sigma_2],$$
$$\sigma_{0m}^2\equiv\mathrm{var}[\hat{f}_0\,|\,\hat{n}=m,f_1,f_2,\sigma_1,\sigma_2].$$

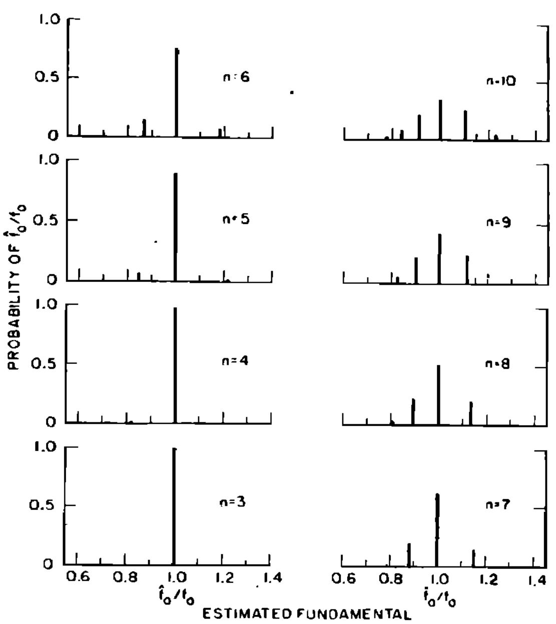

FIG. 3. Discretized example of the probability distributions of the optimum processor estimate of fundamental frequency, $\hat{f}_0$. The stimulus is periodic, comprising the successive harmonics, $f_1=nf_0$ and $f_2=(n+1)f_0$. The model parameters are specified by $\sigma_k/f_k=0.01/\sqrt{2}$. The distributions are multimodal, each mode being approximately Gaussian with standard deviation one-half percent of its mean. Errors by the central processor in estimating the harmonic numbers in the stimulus underlie the multiplicity of modes.

A principal sector, n, is defined as one in which (f_1,f_2) lies or on which it borders. The sector probabilities are then given exactly in Eq. 9. (See Appendix A for derivation.)

$$\Pr[\hat{n}=m=n]=\tfrac{1}{2}(\mathrm{erf}[R_{m1}/\sqrt{2}]+\mathrm{erf}[R_{m2}/\sqrt{2}]),$$
$$\Pr[\hat{n}=m<n]=\tfrac{1}{2}(\mathrm{erf}[R_{m1}/\sqrt{2}]-\mathrm{erf}[R_{m2}/\sqrt{2}]),$$
$$\Pr[\hat{n}=m>n]=\tfrac{1}{2}(\mathrm{erf}[R_{m2}/\sqrt{2}]-\mathrm{erf}[R_{m1}/\sqrt{2}]),\qquad(9)$$

where

$$R_{mk}=|(S_{mk}f_1-f_2)/\sigma_2|[(S_{mk}\sigma_1/\sigma_2)^2+1]^{-\frac{1}{2}}.$$

$S_{m1}>S_{m2}$ are the slopes of the lines bounding the mth sector. R_{mk} may be understood geometrically as the distance between the center of the distribution (f_1,f_2) and the line $x_2=S_{mk}x_1$ in normalized sample space (see Fig. A-1).

Restricting ourselves here to linear estimators of $\hat{f}_0$, we require only the first moments of the sample values within the mth sector. These moments are given exactly in Eq. 10. (See Appendix B for derivation.)

$$\bar{Y}_{1m}\equiv E\left[\frac{x_1-f_1}{\sigma_1}\middle|\hat{n}=m\right]=\left\{\frac{\exp-R_{m1}^2/2}{[(\sigma_2/\sigma_1S_{m1})^2+1]^{\frac{1}{2}}}-\frac{\exp-R_{m2}^2/2}{[(\sigma_2/\sigma_1S_{m2})^2+1]^{\frac{1}{2}}}\right\}\{(2\pi)^{\frac{1}{2}}\Pr[\hat{n}=m]\}^{-1},$$
$$\bar{Y}_{2m}\equiv E\left[\frac{x_2-f_2}{\sigma_2}\middle|\hat{n}=m\right]=\left\{\frac{\exp-R_{m2}^2/2}{[(\sigma_1S_{m2}/\sigma_2)^2+1]^{\frac{1}{2}}}-\frac{\exp-R_{m1}^2/2}{[(\sigma_1S_{m1}/\sigma_2)^2+1]^{\frac{1}{2}}}\right\}\{(2\pi)^{\frac{1}{2}}\Pr[\hat{n}=m]\}^{-1}.\qquad(10)$$

By taking the conditional probability of the linear estimator 7 and using the definitions in Eq. 10, we get the expected linear estimate of the fundamental within the mth sector (Eq. 11),

$$\bar{f}_{0m}=\frac{f_1}{2m}+\frac{f_2}{2(m+1)}+\frac{\sigma_1\bar{Y}_{1m}}{2m}+\frac{\sigma_2\bar{Y}_{2m}}{2(m+1)}.\qquad(11)$$

The exact closed form for the variance of the estimate is given in Appendix C. This result, though cumbersome, provides a check for simpler approximations of the variance given in Sec. II-C.

C. Some Useful Approximations

The linear fundamental estimate $\hat{f}_0$ for N-component signals 12 is obtained from the N-dimensional likelihood function (cf. Eq. 3) by treating the variance function locally constant during maximization. This solution converges to the exact solution when the sample values approach harmonic relations.

$$\hat{f}_0=\sum_{k=1}^{N}\frac{(\hat{n}-1+k)x_k}{\hat{\sigma}_k^2}\Big/\sum_{k=1}^{N}\left(\frac{\hat{n}-1+k}{\hat{\sigma}_k}\right)^2,\qquad(12)$$

where

$$\hat{\sigma}_k = \sigma[(\hat{n}-1+k)\hat{f}_0].$$

We have not found a useful formulation of $\Pr[\hat{n}=m]$ for the N-dimensional case ($N>2$).

Simple approximations of the first and second moments of Eq. 12 within the mth volume are obtained from the multivariate probability distribution along the harmonic "line," i.e., $x_k = x_1(m-1+k)/m$, $k=2, 3, \ldots, N$, as given by Eqs. 13 and 14,

$$\bar{f}_{0m} \approx \sum_{k=1}^{N} \frac{(m-1+k)f_k}{\sigma_k^2} \Big/ \sum_{k=1}^{N} \left(\frac{m-1+k}{\sigma_k}\right)^2, \tag{13}$$

$$\sigma_{0m} \approx \left[\sum_{k=1}^{N} \left(\frac{m-1+k}{\sigma_k}\right)^2\right]^{-\frac{1}{2}}, \tag{14}$$

where

$$\sigma_k = \sigma[f_k].$$

Combining Eqs. 13 and 14 for $f_k = (m-1+k)f_0$ gives a useful formula 15 that relates the precision of the estimate to the precision of the components,

$$(\bar{f}_{0m}/\sigma_{0m})^2 = \sum_{k=1}^{N} (f_k/\sigma_k)^2, \tag{15}$$

for

$$f_k = (m-1+k)f_0.$$

D. Examples of Probability Distribution for $\hat{f}_0$

The nature of the probability distributions 8 for the fundamental estimate $\hat{f}_0$ is illustrated in Fig. 3 for behaviorally relevant values of σ (as established in Sec. III). The distributions are given (Eqs. 8–11, and 14) for two-tone signals comprising successive harmonics n and $n+1$. Each of the two frequencies in the signal is communicated to the central pitch processor with a high precision of $\sigma_k/f_k = 0.01/\sqrt{2}$. Discretized distributions are shown in Fig. 3, because the precision of each mode is great ($\sigma_{0m}/\bar{f}_{0m} = 0.01/2$) compared with the spacing between modes.

The salient property of these distributions is the increasingly large probability of jump errors in estimating the fundamental f_0 that occur with increasing harmonic numbers in the stimulus. These occur because although the central system is allowed to operate optimally it nevertheless errs in estimating the harmonic numbers in the stimulus. Schouten, Ritsma and Cardozo (1962, Fig. 6) gave the first empirical demonstration of this ambiguity phenomenon for harmonic stimuli with harmonic numbers known to the listener, and thereby provided very important evidence for the relevance of the present theoretical approach. Note that these pitch ambiguities are closer to the fundamental than the commonly discussed octave ambiguity, whose occurrence appears to be either less significant or not directly related to pitch perception (Deutsch, 1972).

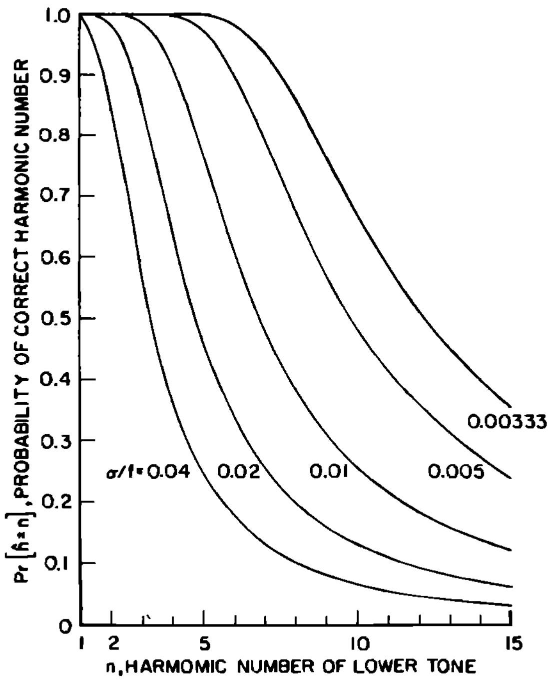

FIG. 4. Probability of correct estimation of harmonic numbers in a two-tone signal with frequencies nf_0 and $(n+1)f_0$. The parameter is the relative standard deviation of the random signals representing the component frequencies.

The distributions given in Fig. 3 are closely approximated (cf. Eq. 8) by Eq. 16,

$$pdf[\hat{f}_0] \cong \sum_{m=1}^{\infty} \Pr[\hat{n}=m] \cdot G\left[\frac{n+\frac{1}{2}}{m+\frac{1}{2}} f_0, \frac{\sigma_k}{\sqrt{2} f_k} \frac{n+\frac{1}{2}}{m+\frac{1}{2}} f_0\right]. \tag{16}$$

For many practical purposes the most important quantity in Eqs. 16 and 8 is the probability of correctly estimating the harmonic numbers in the stimulus ($\Pr[\hat{n}=n]$), because a correct $\hat{n}$ leads to a precise measurement of the stimulus period. Figures 4 and 5 give these probabilities 9 for component frequency precisions covering the range of interest; the effect of harmonic number is manifest.

III. QUANTIFICATION OF THE MODEL PARAMETER σ_k

Quantification of the model parameter σ_k is obtained with data from our previously reported experiments on musical intelligibility (Houtsma and Goldstein 1970, 1971, 1972). The ability of musically skilled subjects to identify eight different standard musical intervals was measured in an eight-alternative forced-choice experiment as a function of the stimulus fundamental and harmonic numbers. Each musical interval was communicated by two sounds presented in succession, each sound comprising two unknown successive harmonics, n and $n+1$, of the desired fundamental. The fundamental

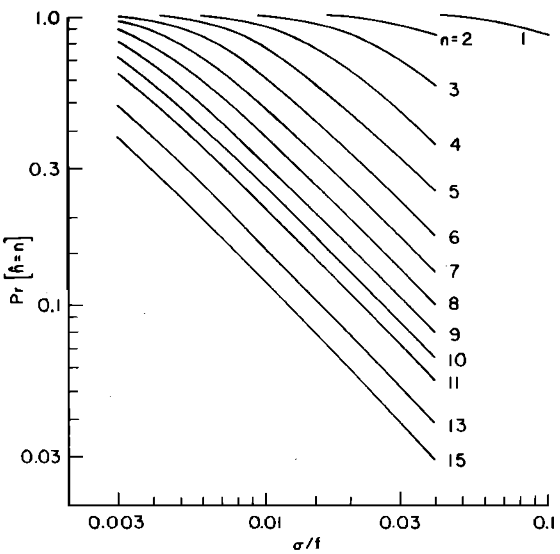

FIG. 5. Same as Fig. 4 except with lower harmonic number as parameter.

of the first note was fixed for each intelligibility test at f_0. The fundamental of the second note f'_0 was located above and below f_0 at approximately semitone (0.06) intervals over the range $4/5 \leqslant f'_0/f_0 \leqslant 5/4$. The harmonic number, n, was chosen randomly from a range of three for each sound presentation so that the subjects were forced to use information from both spectral components in each sound rather than to base their responses on a single component from each sound.

Contours of percentage of correct identification as a function of the stimulus parameters are reproduced (Houtsma and Goldstein 1971, 1972) in Fig. 6 for subject Adrian Houtsma (see also Figs. 8–10). These data are the average of the three presentation conditions that minimized the influence of aural combination tones and gave similar results. These conditions were low-intensity monotic, low-intensity dichotic, and moderate intensity dichotic. Overlaying the data in Fig. 6 are hyperbolae that define constant average frequency for the stimulus components. If the variance of the frequency signals received by the central processor depends primarily upon the frequency represented by each signal and not upon the spacing of the frequencies, then the musical intelligibility performance along each hyperbola should be closely predicted by a constant value of σ.

The subjects' decision strategy in utilizing the output of the central processor of pitch (see Fig. 1) must be specified for the musical intelligibility tests. An optimum strategy would ignore *individual* responses to the first note of each interval, because this note is the same for each presentation in a test run. The pitch of the first note could be remembered with great precision as a reference for interval measurement on the second note. Or, equivalent performance could be obtained by ignoring the first note and responding only to the second. Next, optimum decision strategy would partition the pitch scale to isolate the multiple modes of the central processor response in accord with the estimated harmonic numbers. The partitioning required can be appreciated by referring to Fig. 3 and considering how one could discriminate between several distributions located along a common scale when their principal modes are spaced by approximately 6%. The compactness of the various modes in the processor response would enable correct responses despite ambiguities of pitch. Data and central processor theory agree if, instead of an optimum strategy, a simple partitioning of the pitch scale is presumed, and the subject reports the stimulus note (i.e., one of the 8 standards) that is closest to the sample response of the central processor. No distinction between log or linear scale is required here. Because the spacing between modes within each probability distribution is comparable to, or larger than, the spacing between principal modes, a sample response from a nonprincipal (ambiguous) mode will, except for edge effects, result in an incorrect reported note. Optimal strategy for pitch ambiguities below or above the lowest and highest stimulus notes, would be to report the lowest and highest notes. The edge effects do not appear to be utilized optimally; instead the assumption of random response when edge ambiguities occur gives better agreement between data and theory. This nonoptimum decision strategy assumed for reporting notes gives the simple rule that performance in our musical intelligibility experiment is closely predicted by the probability of correctly estimating the harmonic numbers in the stimulus and thereby perceiving a pitch from the principal mode.

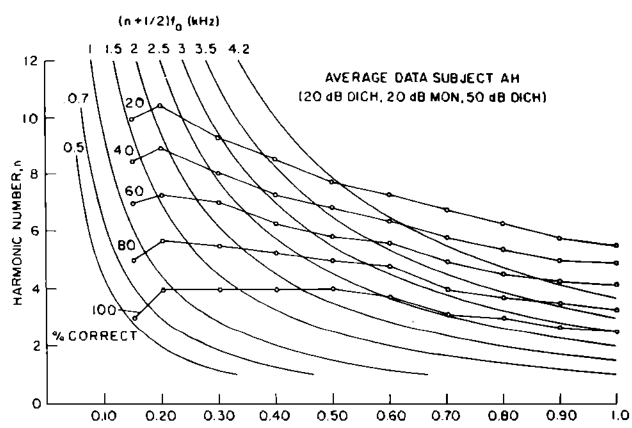

FIG. 6. Contours of equal intelligibility (percent correct identification) of 8 neighboring, standard, two-note, musical intervals as a function of the average parameters of the two-tone periodic stimuli used to communicate each note. These data are the average of the three presentation conditions that minimized the influence of aural combination tones and gave similar results, viz, low-intensity monotic and dichotic, and moderate intensity dichotic. (Reproduced from Houtsma and Goldstein, 1971, 1972.) Overlaying the data are hyperbolae that define constant average frequency for the stimulus components.

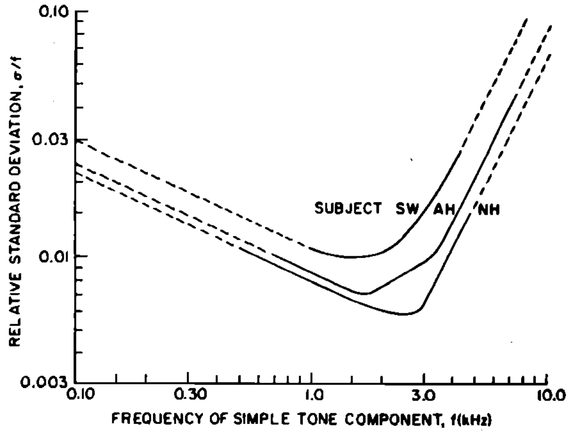

FIG. 7. Relative standard deviations characterizing the precision with which frequency information from aurally resolved tones is communicated to the central processor. Theoretically derived from musical intelligibility data for three subjects (data from Houtsma and Goldstein, 1971, 1972). Dashed lines indicate extrapolations.

The standard deviation σ was calculated at all points of intersection in Fig. 6 between the four data curves for less than 100% correct and the hyperbolae of constant average stimulus frequency. Probability of correct estimation of stimulus harmonics, $\Pr[\hat{n}=n]$, was taken as the predictor of percentage correct, and the value of σ required to fit the data was calculated from Fig. 5. Values found for σ along a given hyperbola were indeed very similar, except for higher, inconsistent values often found from the 20% correct contour. For example, σ calculated from the 20% contour was often nearly twice as large as the remaining three σ values which would span a range of only 15%. Consistent values of σ at each average frequency were geometrically averaged to obtain the single valued function given in Fig. 7. A similar procedure was followed for the data from two other subjects (SW and NH).

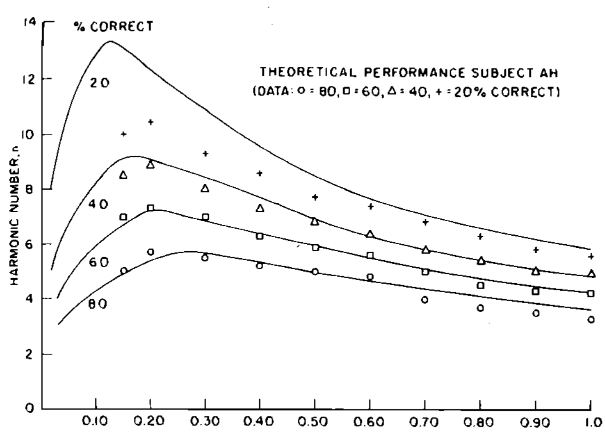

FIG. 8. Theoretical (based on Fig. 7) and measured (Houtsma and Goldstein, 1971, 1972) performance in musical intelligibility experiment for subject AH.

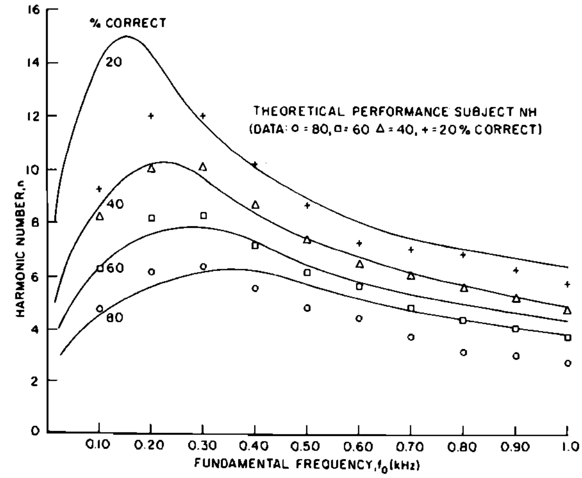

FIG. 9. Subject NH, as in Fig. 8.

Figures 8–10 show the relationship between the derived $\sigma[f]$ functions and the musical intelligibility performance for each subject. Data for each subject are compared directly with the theoretical performance corresponding to $\sigma[f]$. Extensions of the theoretical curves beyond available data reflect the extrapolations assumed in Fig. 7. Figure 8 shows that data and theory are everywhere consistent except for the 20% contour, where empirical performance is consistently poorer than predicted. Subject AH (Adrian Houtsma) was the most practiced (Houtsma, 1971) and probably produced the most reliable data of the three subjects. Data for the two other subjects (Figs. 9 and 10) also show good consistency, although somewhat poorer than that of AH. Both show some departures between theory and data for the 20% contour at low fundamentals, and in addition for the 80% contour at high fundamentals. Subject AH's 20% contour departure is similar for both dichotic and low-intensity monotic data. We have as yet no explanation for these small and probably unimportant inconsistencies. They may be caused by the variability of data between sessions, they may come from data processing, or represent an additional in-

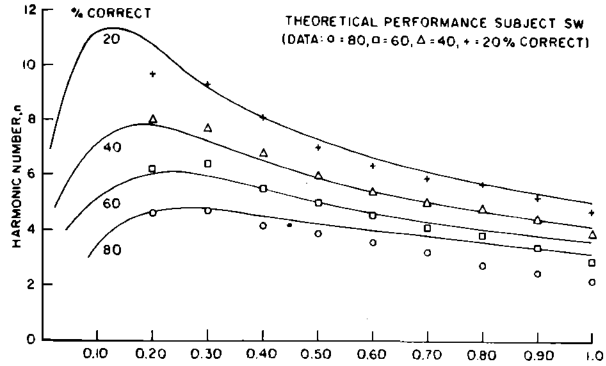

FIG. 10. Subject SW, as in Fig. 8.

crease in σ for very narrow spacing between stimulus components, or additional confusions caused by the first note when this reference note has a very low probability of being correctly heard and reinforced.

In the hypothesized theory the standard deviation σ_k describes the randomness with which frequency information of aurally resolved spectral components is transmitted to the central pitch processor. The demonstration given here (Figs. 7–10) that σ_k can reasonably be regarded as a function only of the frequency transmitted is an important result. A possible outcome of the analysis of the musical intelligibility data could have been a σ_k that depended strongly upon the spacing between stimulus frequencies. The actual outcome (that $\sigma_k[f_k,f_l]\sim\sigma[f_k]$) demonstrates that deterioration of musical intelligibility with increasing harmonic number is attributable primarily to deterioration in the central processor's optimum estimation of harmonic number. It is unnecessary to attribute this deteriorating performance to a *gradual* failure of aural frequency resolution with increasing harmonic number. At some high harmonic number presumably the failure of aural resolution blocks the operation of the central processor; evidence will be given in Sec. VII to support this presumption.

In the following sections it will be demonstrated that the optimum processor theory as quantified in this section consolidates data on pitch of complex tones published by other investigators.

IV. DOMINANT REGION PHENOMENON

Ritsma (1967a), reporting the first successful experiments on periodicity pitch of sounds comprising two successive harmonics, found that for stimuli with fundamental frequencies of 100–400 Hz, harmonics in the range 3–5 were most effective in communicating pitch changes. These results were interpreted as confirming evidence that the region of the third to fifth harmonics is dominant in providing periodicity pitch information for wideband stimuli as well.

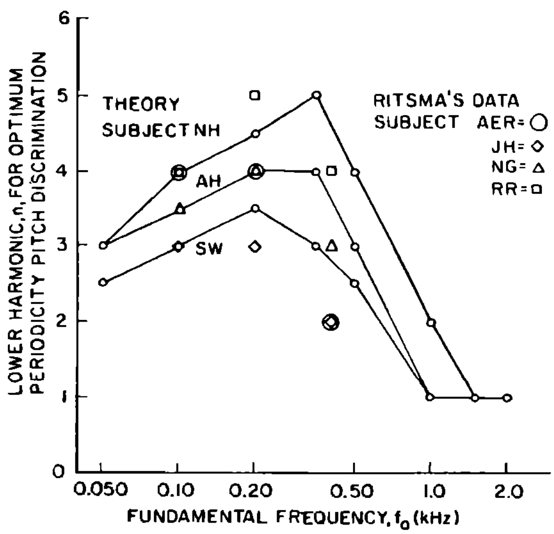

FIG. 11. Dominant region phenomenon for a two-tone signal with frequencies nf_0 and $(n+1)f_0$. Changes in fundamental f_0 are discriminated via periodicity pitch optimally with stimuli comprising the harmonic numbers indicated by the ordinate. (Broad optima spanning more than one integer are represented by their nonintegral average.) Theoretical optimum processor predictions for each of three subjects are based on the variance functions in Fig. 7. Ritsma (1967a) discovered the empirical phenomenon.

A dominant-region phenomenon similar to Ritsma's is predicted by the present optimum processor theory for discrimination of periodicity pitch. Given two signals each with similar harmonics n and $n+1$ but with fundamentals that differ by $\pm|\Delta f_0|$ from a standard f_0, their discriminability from the standard can be calculated on the assumptions that the subject is attending only to the principal mode and that he reports randomly from nonprincipal modes. The percent of correct responses, $P(c)$, in a one-interval, two-alternative, forced-choice experiment (Green and Swets, 1966) is readily calculated from Eq. 8 for unbiased responses.

$$P(c)=\tfrac{1}{2}+\Pr[\hat{n}=n]\cdot\int_0^{|\Delta f_0|/\sigma_{0n}} du(2\pi)^{-\frac{1}{2}}\exp-u^2/2.$$

For a unimodal probability distribution, we would have

$$P(c)=\tfrac{1}{2}+\int_0^{d'} du(2\pi)^{-\frac{1}{2}}\exp-u^2/2.$$

For small d' that is not much greater than unity the last integral is approximated by $d'(2\pi)^{-\frac{1}{2}}$. Thus the measure of discriminability d' for periodicity pitch can be approximated as in Eq. 17,

$$d'\sim|\Delta f_0|\cdot\Pr[\hat{n}=n]/\sigma_{0n}. \tag{17}$$

Equation 17 can be evaluated conveniently by applying the approximation of Eq. 4 to the average stimulus frequency, so that from Eq. 16 $\sigma_{0n}\cong\sigma[(n+\frac{1}{2})f_0]\cdot f_0/\sqrt{2}(n+\frac{1}{2})f_0$. The values of $\sigma[f]/f$ and $\Pr[\hat{n}=n]$ are given by Figs. 7 and 5, respectively. Computations of d' as a function of n for fixed f_0 and Δf_0 reveal that pitch discriminability is usually best, i.e., d' is largest, at some intermediate n. The connected curves in Fig. 11 give these optima for each of the three subjects represented in Fig. 7. The nonintegral entries in Fig. 11 reflect the broadness of the theoretical maxima which often span adjacent harmonic numbers. Ritsma's data are similarly plotted in Fig. 11.

For fundamentals below 500 Hz the theoretical optimum may be understood as the interaction of two opposing tendencies with changing harmonic number. Higher harmonic numbers benefit from increased precision (lower σ/f in Fig. 7), but suffer from more frequent errors in central estimation of harmonic numbers. At fundamentals above 500 Hz the deteriorating precision of higher frequencies (Fig. 7) dominates the theoretical prediction. It is clear from the differing theoretical curves in Fig. 11 that the magnitude as well as the shape of the derived variance functions (Fig. 7) are reflected in the predicted optima.

The agreement with Ritsma's (1967a) data shown in Fig. 11 within and across subjects appears highly significant. According to the present theory, the dominant region effect for low fundamentals would not apply to wideband complex tones, because the probability of correct central estimation of harmonic numbers is essentially unity when the lowest harmonics are present.

V. PITCH OF MULTICOMPONENT ($N>2$) COMPLEX TONES

Plomp (1967) has reported pitch judgments by naive subjects using novel 12-tone stimuli, created by modifying standard harmonic stimuli through lowering the frequencies of the first m harmonics by 10% while increasing the remaining harmonics by 10%. Equation 18 defines Plomp's test and standard stimuli, with the number of tones $N=12$.

$$P_{\text{test}}(t)=\sum_{k=1}^{m}\cos(1.8\pi kf_0t)+\sum_{k=m+1}^{N}\cos(2.2\pi kf_0t),$$
$$P_{\text{stnd}}(t)=\sum_{k=1}^{N}\cos(2\pi kf_0t). \quad (18)$$

By measuring as a function of m and f_0 whether the test stimulus was judged higher or lower in pitch than the standard, Plomp sought to discover the conditions under which the lower harmonics determine pitch. Plomp's data are abstracted in Fig. 12 with the stimulus conditions (m,f_0) for which the test stimulus had equal probability (50%) of being judged above or below the standard.

Optimum processor theory Eq. 12 predicts the periodicity pitch for N-tone complexes. Given that the lower harmonics including the "fundamental" are all present, then with probability approaching unity the harmonic numbers are estimated centrally as $\hat{n}-1+k=k$. Substitution in Eq. 12 and rearranging gives the estimator for periodicity pitch

$$\frac{\hat{f}_0}{f_0}=\sum_{k=1}^{N}\left(\frac{kf_0}{\sigma[k\hat{f}_0]}\right)^2\frac{x_k}{kf_0}\Big/\sum_{k=1}^{N}\left(\frac{kf_0}{\sigma[k\hat{f}_0]}\right)^2. \quad (19)$$

Equation 19 can be solved for the conditions (m,f_0) for which the expected pitch of Plomp's test and standard stimuli are equal, that is, $E[\hat{f}_0/f_0]=1$. By approximating $\sigma[k\hat{f}_0]\approx\sigma[kf_0]$ and substituting $E[x_k/k]=0.9f_0$ for $k\leqslant m$ and $1.1f_0$ for $k>m$, we get

$$E\left[\frac{\hat{f}_0}{f_0}\right]=1\approx\left[0.9\sum_{k=1}^{m}\left(\frac{kf_0}{\sigma[kf_0]}\right)^2+1.1\sum_{k=m+1}^{N}\left(\frac{kf_0}{\sigma[kf_0]}\right)^2\right]\times\left[\sum_{k=1}^{N}\left(\frac{kf_0}{\sigma[kf_0]}\right)^2\right]^{-1}. \quad (20)$$

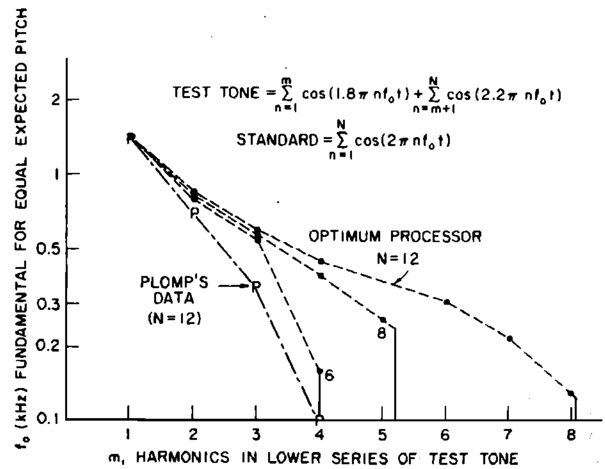

FIG. 12. Fundamental frequency at which the expected pitch of Plomp's (1967) test and standard stimuli are equal. Theoretical predictions are based upon the average of the variance functions in Fig. 7, and optimum processing of only the lowest N stimulus components.

Equation 20 was solved numerically by use of the geometric average of the relative variance functions σ/f for the three subjects shown in Fig. 7. These theoretical solutions are given in Fig. 12 for three different values of N. The large discrepancy between the data and optimum processor predictions for $N=12$ implies that all 12 components in Plomp's stimuli do not contribute their frequency information optimally to periodicity pitch. Instead, the closer agreement between data and theory for $N=6$ suggests that only approximately the first six harmonics contribute optimally. (The theoretical predictions for $N=5$ agree more closely with Plomp's data for $m=2$ and 3, but no solution exists for $m=4$.) On the basis of the earlier theoretical account of musical intelligibility data for two-tone stimuli (Sec. III, Figs. 8–10), one could expect optimal processing for harmonics as high as 8–10; the situation may be different, however, when the stimuli comprise more than two tones (cf., Plomp, 1964). Final conclusions should be reserved until the performance of musically sophisticated subjects is systematically measured.

VI. PRECISION OF PERIODICITY PITCH

What appear to be the most direct measures of the variance of periodicity pitch for the principal mode are the pitch-matching experiments reported by Ritsma (1963). Subjects made 80 repeated pitch matches (binaurally at 40 dB SL) between a periodic three-tone test stimulus and each fixed standard stimulus by adjusting the fundamental of the former. The test stimulus comprised the harmonics 6–8 and the standard comprised 9–11, except at the fundamental frequency of 600 Hz where the harmonics were 3–5 and 5–7, respectively. Matches were also made between simple tones. Ritsma's data are reproduced in Fig. 13; similar,

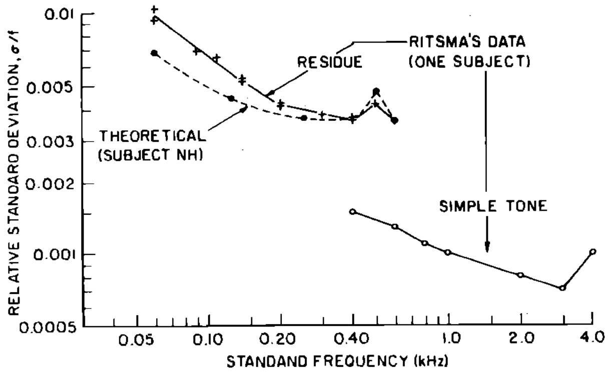

FIG. 13. Theoretically predicted precision of periodicity pitch (principal mode) for the three-tone periodic stimuli used by Ritsma (1963) to match (residue) pitches. Ritsma's data are the variances of rep ated adjustments to a standard. Data from matching simple tones are also reproduced.

though unpublished, data are claimed for another subject.

These data are predicted by optimum processor theory with the assumptions that the pitch of the standard was remembered with great precision relative to the measured precision, and that adjustments of test stimuli for ambiguous modes were rejected in data processing. Equation 15 with $m=6$ and $N=3$ ($m=3$ for $f_0=600$ Hz) gives the required prediction. The calculated precision is given in Fig. 13 for subject NH's variance (Fig. 7), as this choice suggested that Ritsma switched to lower harmonics at $f_0=600$ Hz to avoid the rapidly deteriorating precision of the internal representation of the component frequencies. However, the variance functions for the other two subjects (Fig. 7) are also compatible with Ritsma's data.

Ritsma's data also demonstrate that the precision for periodicity pitch is not equal to, nor theoretically accounted for (with Eq. 15) by the precision with which frequency of simple tones can be discriminated. The variance functions (Fig. 7) which describe the precision with which frequency information is conveyed to the central processor of periodicity pitch are much greater than those measured for simple tone frequency discrimination (Shower and Biddulph, 1931; Harris, 1952; Henning, 1966). Therefore, behavioral phenomena and aural signal processing associated with simple tones cannot fully account for the theoretically derived variance functions associated with periodicity pitch.

VII. PITCH SHIFT FOR INHARMONIC COMPLEX TONES

The outstanding empirical property demonstrated in classical studies of periodicity pitch (Hermann, 1912; Schouten, 1940b; de Boer, 1956; Schouten, Ritsma, and Cardozo, 1962) is that inharmonic complex tones evoke a pitch that does not correspond to a difference frequency. Periodicity pitch of a complex tone shifts nearly in proportion to a uniform frequency perturbation of its constituent frequencies away from an harmonic series. A central problem in classical theory has been to account for this proportionality factor, which can be interpreted in both spectral and temporal theories (Schouten, 1940b, p. 244; de Boer, 1956; Schouten, Ritsma, and Cardozo, 1962) as being inversely equal to the harmonic position of the effective carrier of the internal signal. Difficulties in accounting for this proportionality factor (second effect of de Boer, 1956 and Schouten *et al.*, 1962; see also Schroeder, 1966; Fischler, 1967; Fischler and Cern, 1968; Walliser, 1969) were attributed by the present author (Goldstein and Kiang, 1968, p. 990) to the influence of aurally generated combination tones, which had been regarded until recently (Goldstein, 1967a) as negligible at low sound levels (Zwicker, 1955; Plomp, 1965).

Smoorenburg (1970) demonstrated empirically with measurements of both periodicity pitch and aural combination tones for the same subjects that indeed aural combination tones do modify the pitch shift for inharmonic complex tones. These data are predicted by optimum processor theory, and in addition the limited resolution of aural frequency analysis is revealed. Given a successive harmonic series with the lowest frequency $n_l\cdot f_0$ and highest frequency $n_h\cdot f_0$, then from Eq. 12 the estimator that includes the principal mode of the unshifted signal is given by

$$\frac{\hat{f}_0}{f_0}=\sum_{k=n_l}^{n_h}\left(\frac{k}{\sigma[k\hat{f}_0]}\right)^2\frac{x_k}{kf_0}\Big/\sum_{k=n_l}^{n_h}\left(\frac{k}{\sigma[k\hat{f}_0]}\right)^2. \quad (21)$$

Let Δf_0 be the uniform frequency shift, so that $E[x_k]=kf_0+\Delta f_0$. For small Δf_0, if we approximate $\sigma[k\hat{f}_0]\approx\sigma[kf_0]$, then the expected pitch is given by

$$\frac{\bar{f}_0}{f_0}\doteq 1+\frac{\Delta f_0}{f_0}\sum_{k=n_l}^{n_h}\frac{1}{k}\left(\frac{kf_0}{\sigma[kf_0]}\right)^2\Big/\sum_{k=n_l}^{n_h}\left(\frac{kf_0}{\sigma[kf_0]}\right)^2. \quad (22)$$

The effective harmonic number n_e 23 is defined as the reciprocal of the proportionality factor in Eq. 22,

$$n_e\equiv\sum_{k=n_l}^{n_h}\left(\frac{kf_0}{\sigma[kf_0]}\right)^2\Big/\sum_{k=n_l}^{n_h}\frac{1}{k}\left(\frac{kf_0}{\sigma[kf_0]}\right)^2. \quad (23)$$

A useful bound follows from Eq. 23: $n_e=k$ when the precision of the kth harmonic predominates, therefore $n_l\leqslant n_e\leqslant n_h$. Approximating the variance function in Eq. 23 throughout the low-frequency region as $\sigma[f]\propto\sqrt{f}$ (cf., Fig. 7), gives

$$n_e=(n_l+n_h)/2. \quad (24)$$

Smoorenburg's (1970) data for the effective harmonic number measured with a two-tone stimulus (f_1,f_2) are reproduced in Fig. 14, along with the lowest frequency odd-order $[f_1-m(f_2-f_1)]$ aural combination tones that could be measured with f_1 fixed in amplitude and frequency (cf., Goldstein, 1970, p. 233). Theoretical calculations of effective harmonic numbers 23 were made by use of the geometric average of the variance functions for the three subjects in Fig. 7. For the present calculations, the more transparent Eq. 24 was found to offer nearly as precise an account of the data and it is useful and consistent to describe the theory in its terms.

When the central processor receives information on only the stimulus frequencies $[f_1=nf_0+\Delta f_0$ and $f_2=(n+1)f_0+\Delta f_0]$, then the effective harmonic number is the average of the stimulus harmonic numbers, $n_e=n+\frac{1}{2}$. Figure 14 shows that the data are not predicted by this situation. If, instead, we allow for the presence of stimulus-like aural combination tones, which add lower harmonics to the stimulus, we can find the lowest harmonic number n_l in the effective stimulus which fits the n_e data. Following this calculation scheme, the predicted combination tones agree well with Smoorenburg's data only when neither stimulus harmonic exceeds 10 (i.e., the six lowest frequency entries in Fig. 14). To account for the remaining data it is necessary to add the rule that only constituent frequencies that are more than 10% remote from each other contribute to the central processor, i.e., limited frequency resolution.

Several points in this analysis should be noted. Both frequency resolution and the presence or absence of frequency signals were successfully idealized here as all-or-none. This supports the earlier theoretical conclusion from musical intelligibility data (Sec. III) that limited aural frequency resolution does not appear to cause a gradual deterioration of frequency information, for increasing harmonic number, at the input to the central pitch processor. Second, the idealization of the amplitudes of the constituent stimulus tones as all-or-none is confirmed by the equal roles played by the different aural combination tones, despite their manifestly different amplitudes, all below those in the two-tone primary stimulus. Finally, van den Brink's (1970) unique evidence that the pitch of inharmonic complex tones is nonlinearly related to the uniform frequency shift, though qualitatively consistent with the nonlinear nature of the estimator (5, 21), is not remotely confirmed quantitatively by theoretical calculations. Note, that the nonlinear estimator for the low-frequency region where $\sigma[f]\propto\sim\sqrt{f}$ is given by Eq. 25. For $N=2$, the partitioning of sample space is given by Eq. 6.

$$\hat{f}_0=\left[\frac{2}{N(2\hat{n}+N-1)}\sum_{k=1}^{N}\frac{x_k^2}{\hat{n}+k-1}\right]^{\frac{1}{2}}. \tag{25}$$

Instead, a large contribution to nonlinearity in the pitch shift phenomenon often appears to be caused by changes in the number of combination tones present at different frequency shifts. For positive shifts ($\Delta f_0>0$) the relative frequency spacing between stimulus tones decreases; since narrower spacing creates the possibility of more combination tones, the effective harmonic number for positive shifts is often smaller than for negative shifts.

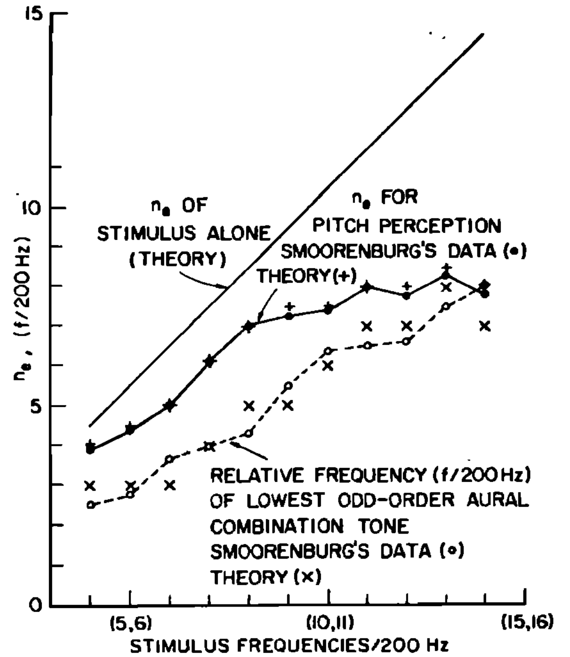

FIG. 14. Reduction by aural combination tones and limited resolution of the effective harmonic number, n_e, in pitch perception with inharmonic two-tone signals. Smoorenburg's (1970) data on effective harmonic number are fitted to theory on the assumption of optimum processing of all effective stimulus tones. The lowest frequency odd-order aural combination tone is predicted from the theoretical n_e and the assumption that only components with spectral spacings exceeding 10% contribute to periodicity pitch. Also shown is the effective harmonic number if the stimulus frequencies alone comprised the effective stimulus.

VIII. THE CONFOUNDING ROLE OF AURAL COMBINATION TONES

Two empirical characteristics of periodicity pitch have served as primary evidence in classical theory for mediation of pitch of complex tones directly by temporal intervals measured in the cochlear output (Schouten, 1940a; Licklider, 1956). First, periodic complex tones comprising harmonics too high to be aurally analyzable as simple tones can evoke a low periodicity pitch corresponding to the fundamental frequency (Seebeck, 1841, 1843; Schouten, 1940a). This was interpreted to imply that the failure of aural spectrum analysis is essential for permitting the stimulus harmonics to beat and define their common temporal period. Second, phase effects upon pitch have been reported for stimuli with more than two harmonics higher than approximately the fifth (Licklider, 1955; de Boer, 1956; Ritsma and Engel, 1964). This was interpreted as reflecting the influence of phase upon the complex waveform responsible for mediating pitch.

De Boer (1956) performed systematic studies of the pitch of inharmonic complex tones and found that periodicity pitch exists *as well* for sounds comprising only low harmonics (spacings greater than approximately 15%) for which behavioral frequency analysis is operative and phase effects are negligible. Earlier, Schouten (1940b) had proposed that the pitch of inharmonic complex tones could be described as corresponding to the fundamental frequency of an harmonic series which closely matches those stimulus frequencies that excite the pitch extractor; the actual form of the mediating signals, however, was presumed to be temporal. De Boer (1956) synthesized these ideas into a composite pitch mediation theory. Complex tones with wide spectral spacings are tonotopically analyzed and periodicity pitch is assigned on the basis of an harmonic series that best matches the *analyzed* spectrum; de Boer's least-squares nonstochastic approach anticipates the present more complete theory (Sec. IX). Complex tones with spectral spacings narrower than the bandwidth of the cochlear filters provide residual complex waveforms at the cochlear outputs and pitch is assigned on the basis of temporal intervals defined by the *fine structure* in these waveforms; this view is now untenable (Houtsma and Goldstein, 1972; Goldstein, 1972).

The temporal fine-structure mechanism of de Boer was accepted by Schouten, Ritsma and Cardozo (1962) as most consistent with periodicity pitch phenomena, including the aforementioned and the multiple-valued (ambiguous) nature of periodicity pitch for both harmonic and inharmonic complex tones. The absence of phase effects and the presence of behavioral resolution for complex tones with wide spectral spacing does not logically preclude co-existing weak spectral interactions that create residual complex waveforms in the cochlear output. For example, it is known (Mathes and Miller, 1947; Ritsma and Engel, 1964; Goldstein, 1965) that perceptibility of phase changes can be greater with stimuli comprising three tones than with two tones. A property consistent with neither of Schouten and de Boer's formulations was the frequent theoretical underestimation of pitch shifts for inharmonic complex tones, i.e., the second effect (de Boer, 1956; Schouten, Ritsma, and Cardozo, 1962), which often could not be accounted for regardless of the filtering or weighting assumed in aural signal processing (Goldstein and Kiang, 1968, p. 990; Smoorenburg, 1970). This discrepancy became increasingly greater as the stimulus was chosen to comprise higher harmonics (see Sec. VII).

Recognizing that wideband inharmonic stimuli produce different temporal periodicities at different characteristic frequencies of the aural spectrum analyzer, Ritsma (1967a) sought to determine which spectral region is dominant in exciting the pitch extractor. He employed an experimental paradigm similar to Plomp's (1967), (see Sec. V), and concluded that the dominant region is contained within the lowest harmonics. Moreover, from additional experiments in which the lower series contained only two successive harmonics, Ritsma (see Sec. IV) concluded that the spectral region covering the harmonics 3–5 dominates pitch perception for fundamentals in the range 100–400 Hz. Ritsma's results constituted a major puzzle for the temporal waveform formulation, because it had been assumed that the existence of behavioral frequency analysis and weak phase effects for low harmonic stimuli implied weak temporal interactions and poorly defined periodicity pitch.

New possibilities for interpreting data on periodicity pitch arose when it was discovered (Goldstein, 1967a) that significant aural combination tones existed at lower sound levels than previously expected. Stimuli containing only upper harmonics are augmented with a series of stimulus-like lower harmonics by an essential nonlinearity in peripheral signal processing. The second effect of the residue was no longer a puzzle (Goldstein and Kiang, 1968; Ritsma, 1970; Smoorenburg, 1970); all available data on the pitch of inharmonic complex tones (de Boer, 1956; Schouten, Ritsma and Cardozo, 1962; Sutton and Williams, 1970; Houtsma, 1971) could be fully accounted for with aural combination tones (where relevant). It has become increasingly evident that complex tones which would otherwise be musically ineffectual because of dense spectral spacing can be augmented by aural combination tones that are musically effective as well as behaviorally resolvable (Goldstein, 1967a, p. 684; Smoorenburg, 1970, see Sec. VII of this paper; Houtsma and Goldstein, 1972).

Aural combination tones are also responsible for at least some phase effects in periodicity pitch, as it has been shown (Goldstein, 1970) that the very presence of some combination tones is phase-dependent for stimuli comprising three tones. Data by Ritsma and Engel (1964) on pitch matchings for periodic three-tone, quasi-FM stimuli include examples where the fundamental frequency of the periodic stimulus does not correspond to periodicity pitch. This suggests an effective stimulus comprising some nonsuccessive harmonics, which is a condition that can be created by aural combination tones. Circumventing phase effects of peripheral origin, Houtsma (Houtsma and Goldstein, 1971) measured the discriminability between AM and quasi-FM (Mathes and Miller, 1947; Goldstein, 1967b) when these sounds are presented normally as well as when the center tone and side tones enter opposite ears. The latter dichotic presentation did not destroy periodicity pitch, but it did remove phase effects.

Were it not for the confounding role of aural combination tones it is very likely that periodicity pitch would have been attributed earlier exclusively to harmonics of low order. Periodicity pitch theory, in all likelihood, would then have developed much sooner with a clear recognition that the auditory system organizes stimuli tonotopically and then proceeds hierarchically to read across the tonotopic dimension. (Examples of recent theoretical formulations of periodicity pitch are those by Walliser, 1969; Terhardt, 1972; Wightman, 1972.)

IX. EMPIRICAL AND THEORETICAL ANTECEDENTS

A. Empirical Bases for Optimum Processor Constraints

An optimum processor theory is meaningful when it is structured by an appropriate set of constraints. These constraints may be inferred directly from empirical phenomena and they may be selected to yield the desired fit between data and theory. Optimum processor theory follows well-defined logical rules, while the discovery of appropriate constraints almost never does.

The constraints introduced in Sec. I are (i) the optimum processor receives information on only aurally resolved simple tones, (ii) phase is irrelevant, (iii) amplitude is irrelevant (within limits), (iv) the processor presumes periodic stimuli comprising successive harmonics, (v) the stimulus harmonics presumed must be measured by the processor for each stimulus presentation, and (vi) the processor is provided with independent stochastic information on the frequency of each resolved component.

The empirical bases for these inferred, but not logically proved, constraints are provided by the studies of many investigators over a long period of time. These studies have been referenced throughout this paper; it is useful to review them briefly.

The first two constraints concerning aural resolution and irrelevance of phase were anticipated by Helmholtz (1863). In his experimental investigations he found that musical properties of periodic complex tones are primarily determined by behaviorally resolvable, low-order harmonics, that is, the first 6–8. He was the first to demonstrate the absence of significant phase effects for low-order harmonics. But he was wrong in concluding that the fundamental component mediates musical pitch. De Boer (1956) demonstrated the existence of periodicity pitch for low-order harmonics. De Boer (1956) and Schouten, Ritsma, and Cardozo (1962) discovered the second-effect phenomenon in experiments on pitch shift for inharmonic complex tones (see Sec. VII), in which the effective harmonic number saturated at the relatively low harmonic number of approximately 10. De Boer proposed that this effect could be accounted for by assuming the auditory system emphasizes lower frequency stimulus components. Schouten *et al.* (1962), presented data which could not be so simply accommodated and tacitly suggested the existence of a basic enigma. The present author (Goldstein, 1967a, 1970) demonstrated the existence of odd-order aural combination tones below the stimulus frequencies at previously unsuspected low sound levels. Combination tones were found to affect musical quality greatly and generate significant phase effects when they were generated by complex tones with spacings near the limits of behavioral resolution. Goldstein and Kiang (1968) proposed on physical and psychophysical grounds that the second effect can be accounted for by peripherally generated odd-order combination tones. Smoorenburg (1970) demonstrated empirically the influence of aural combination tones in pitch shift experiments. Because of the presence of aural combination tones, one could reasonably hypothesize that only aurally resolvable simple tones mediate pitch. This hypothesis was fully consistent with the findings of Ritsma (1967a,b) and Plomp (1967) that relatively low-order harmonics are most effective in mediating periodicity pitch. Finally, the musical intelligibility experiments of Houtsma and Goldstein (1971, 1972) with monotic and dichotic stimuli demonstrated conclusively the primary role of low-order harmonics in periodicity pitch and the secondary roles of peripheral interaction phenomena including temporal beats, aural combination tones, and significant phase effects.

The third constraint concerning the minor roles of amplitude of spectral components can be inferred from the natural phenomenon that musical sounds with quite different distributions of spectral amplitude convey the same pitch. Perhaps it was this invariance that led Helmholtz to insist that pitch of periodic complex tones is mediated by the fundamental component. Controlled experiments on the effect of amplitude distribution were reported by Plomp (1967) who found similar pitch for uniform and $1/f$ spectral amplitude distributions. Ritsma (1967a) found that the outcome of his dominance experiments was quite insensitive to the relative amplitudes between his two competing sounds. Smoorenburg's (1970) data on the effects of aural combination tones on pitch demonstrate that presumably low-amplitude tones have great influence on pitch. With phase and amplitude (within limits) ignored, the central processor of pitch need only collect information on the frequencies of aurally resolved stimulus tones.

The fourth constraint, that the pitch processor presumes the stimulus is periodic and comprises successive harmonics, was an inference first made by Schouten (1940b) to describe the fact that pitch of inharmonic complex tones does not correspond to a difference frequency (f_2-f_1). De Boer (1956) gave the first empirical demonstration that periodicity pitch for periodic sounds comprising only upper odd harmonics differs from the true fundamental frequency; this would be expected if the auditory system presumed stimuli comprising successive harmonics.

The fifth constraint, that the pitch processor must decide solely on the basis of each stimulus presentation which stimulus harmonics to presume, may be inferred from the natural phenomenon that pitch in music and speech appears to be appreciated without prior information of harmonic structure of specific stimuli. Experimental evidence was given by Schouten *et al.* (1962), who demonstrated the existence of ambiguities in periodicity pitch heard in *repeated* presentations of periodic stimuli comprising successive harmonics; ambiguities should not occur according to optimum processor theory if the processor could accumulate information on harmonic numbers from prior presentations. Proof that withholding information on harmonic structure need not interfere with reliable perception of periodicity pitch was given in experiments by Cross and Lane (1963) and Houtsma and Goldstein (1972).

The sixth and final constraint is that variability in pitch perception is accounted for by independent stochastic variability in central processor information on each stimulus frequency. This choice is based on the physiologically motivated supposition that frequency information of aurally resolved simple tones is conveyed to the central pitch processor by independent noisy channels. Furthermore, optimum processor theory requires noisy frequency information to account for ambiguities. The characterization of this randomness by static, stochastically independent Gaussian distributions is a simple and reasonable choice. The fact that the variance parameters of the Gaussian distributions are dependent only upon the frequency represented is a major outcome, and not a built-in assumption of the analysis of musical intelligibility data. In the final analysis, this assembly of constraints and their supporting basis is interesting primarily because the theory that was logically constructed from the constraints unifies a large body of data on pitch of complex tones.

B. Relation to Earlier Theory of Periodicity Pitch

Several theories of periodicity pitch have been proposed (Schouten, 1940a, b; de Boer, 1956; Licklider, 1956; Walliser, 1969; Terhardt, 1972a, b; Wightman, 1972). Of these, the "harmonic spectrum matching" suggested by Schouten (1940b) and developed by de Boer (1956) as least-squares matching is the most relevant for the theory given here. We demonstrate this relationship heuristically by further developing de Boer's least-squares formulation.

De Boer (1956) demonstrated that periodicity pitch exists for inharmonic complex tones comprising widely spaced as well as narrowly spaced simple tones. He proposed that for the stimuli with wide spectral spacing the auditory system actually operates in the frequency domain to match a harmonic spectral pattern to the stimulus spectrum. He suggested a least-squares formulation as a way of describing this matching. Equation 26 gives the square error as a function of the stimulus frequencies, f_k, and the matched frequencies, $\hat{n}\hat{f}_0$, $(\hat{n}+1)\hat{f}_0, \ldots,$

$$\mathcal{E}^2=\sum_{k=1}^{N} a_k[f_k-(\hat{n}+k-1)\hat{f}_0]^2. \tag{26}$$

Assuming the weights, a_k, are constants, de Boer gives the optimum $\hat{f}_0$ (our notation),

$$\hat{f}_0=\sum_{k=1}^{N} a_k(\hat{n}+k-1)f_k/\sum_{k=1}^{N} a_k(\hat{n}+k-1)^2. \tag{27}$$

We can bridge the gap between de Boer's expressions and those derived from maximum-likelihood estimation (mle) by postulating that the pattern matcher has available only samples from independent Gaussian representations of the stimulus frequencies, as specified earlier in Eq. 1. Then

$$\mathcal{E}^2=\sum_{k=1}^{N} a_k[x_k-(\hat{n}+k-1)\hat{f}_0]^2 \tag{28}$$

and

$$\hat{f}_0=\sum_{k=1}^{N} a_k(\hat{n}+k-1)x_k/\sum_{k=1}^{N} a_k(\hat{n}+k-1)^2. \tag{29}$$

Next, we require that the constants, a_k, be chosen to minimize the expected square error in estimating the fundamental. We already know from Sec. II that Eq. 29 has a multimodal probability distribution, so it is not trivial to solve this problem. Let us generally adopt the procedure that applies when the stimulus comprises successive harmonics beginning with a sufficiently low order so that the probability of estimating the correct harmonic numbers approaches unity. Then we can treat $\hat{n}$ as a constant in evaluating the expected square error of the fundamental estimate,

$$\mathcal{E}_0^2=E[(\hat{f}_0-E[\hat{f}_0])^2]. \tag{30}$$

From Eqs. 1, 29, and the independence of the frequency samples we get

$$\mathcal{E}_0^2=\sum_{k=1}^{N} [a_k(\hat{n}+k-1)\sigma_k]^2/[\sum_{k=1}^{N} a_k(\hat{n}+k-1)^2]^2. \tag{31}$$

Equation 31 can be minimized for each a_k; the solution within a common arbitrary constant is

$$a_k=\sigma_k^{-2}. \tag{32}$$

Next, we reason that the system has no information on the variances of the internal representations of the actual stimulus frequencies, but it can estimate these variances by using the variances it can associate with the individual frequencies in the matched spectrum. Thus

$$a_k^{-\frac{1}{2}}=\hat{\sigma}_k=\sigma[(\hat{n}+k-1)\hat{f}_0]. \tag{33}$$

Hence Eqs. 28 and 29 become

$$\mathcal{E}^2=\sum_{k=1}^{N} [x_k-(\hat{n}+k-1)\hat{f}_0]^2/\hat{\sigma}_k^2, \tag{34}$$

and

$$\hat{f}_0=\sum_{k=1}^{N}\frac{(\hat{n}+k-1)x_k}{\hat{\sigma}_k^2}\Bigg/\sum_{k=1}^{N}\left(\frac{\hat{n}+k-1}{\hat{\sigma}_k}\right)^2. \tag{35}$$

Clearly Eq. 34 is the statistic that is minimized by maximum-likelihood estimation (mle) and Eq. 35 is the linear estimator given by Eq. 12. Thus all treatments of data described in this paper could have been identically described in terms of this least-squares pattern matching. Important conceptual distinctions, of course, do exist between mle and least-squares matching. The mle approach frees one from unproved, though possibly true, presumptions on physiological mechanisms that favor place over time representations of resolved components. This is so because the square error 34 is derived with mle from basic probabilistic considerations and need not appeal to a spatial pattern interpretation, as in the least-squares matching approach. Second, the minimum squared error derivation given here is a nonrigorous one that benefited with hindsight from the known mle solution. Without the more basic logic of mle is it doubtful that we would have pursued the difficult mathematical problems encountered (see Sec. II). Finally, the optimality of the mle solution could not be appreciated from a least-squares matching approach.

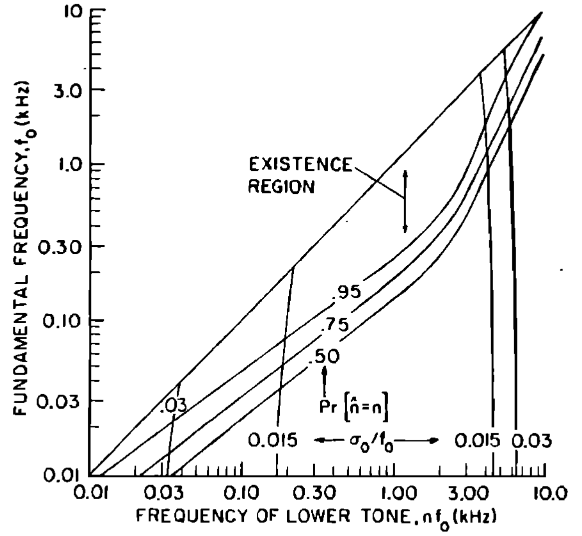

FIG. 15. Theoretically predicted existence region for periodicity pitch of an harmonic two-tone signal (aural combination tones not included), calculated from the geometric average of the variance functions in Fig. 7. Boundaries are set by choosing an acceptable precision (σ_0/f_0) and probability of nonambiguous responses (Pr[$\hat{n}=n$]).

X. FURTHER WORK

Three categories of further experimental and theoretical work can be distinguished: work within the theoretical paradigm presented in this paper, attempts to remove the limitations of present theory, and identification of the physiological mechanisms that mediate periodicity pitch.

Many issues within the present theoretical paradigm have been raised throughout this paper; some additional issues follow. Experimental research is needed on the options available to the central processor for choosing information from either one or both ears. Theoretical predictions of existence regions for musically interesting stimuli (Ritsma, 1962) require study. For example, Fig. 15 defines existence regions for stimuli comprising two successive harmonics on the bases of precision for periodicity pitch (σ_0/f_0) and the probability of nonambiguous responses (Pr[$\hat{n}=n$]). For other stimuli, such as lowpass filtered unipolar pulse trains and simple tones, precision alone (Eq. 15) would be the primary determinant of the existence region. Finally, periodicity pitch for periodic stimuli that do not contain energy at successive harmonics can have a weak or nonexistent mode at the fundamental frequency; experimental and theoretical investigations of all modes are needed (Flanagan and Guttman, 1962; Ritsma, 1967b; Secs. VII and VIII of this paper). For example, a seemingly paradoxical theoretical prediction obtains when the fundamental tone is added to a periodic complex tone comprising successive harmonics no lower than the third. When aural combination tones of odd and even orders (Goldstein, 1967a; Hall, 1972) do not fill the spectral gap between fundamental and upper harmonics, periodicity pitch differing from the fundamental should be heard—a condition clearly to be avoided in musical practice.

The most important limitation of the present theory is that it is formally restricted to complex tones (stimuli comprising line spectra, with each line of width less than the theoretically derived variance in Fig. 7), and it cannot treat transient and dynamic properties of stimuli. Thus, the extensive literature on pitch of click pairs and related stimuli which have continuous spectra (Bilsen, 1968; Ritsma, 1970; Small, 1970) cannot be treated in the present formulation. Repeating the click pairs at regular but infrequent intervals creates a line spectrum, but the phenomenon is still dependent upon dynamic properties of the stimulus. In Fourier terms, the central processor could determine pitch in the latter case by interrelating signals representing clusters of spectral lines that were grouped separately within the limits of aural resolution. Another view of this inability of present theory to treat the dynamic case is gained by considering any of the complex tone stimuli within the scope of present theory and multiplying it by a simple tone with frequency (Δ) small relative to the original spectral spacing. Each spectral line of the original stimulus will be split into two lines separated by 2Δ. Periodicity pitch should not disappear; instead, depending upon Δ the sound will fluctuate in loudness or probably suffer a deterioration in the precision of pitch. Finally, questions on the role of learning (Licklider, 1956; Thurlow, 1963; Whitfield, 1970) in organizing the properties of the central processor of periodicity pitch are moot in the present theory, although the existence of a consolidating logic for periodicity pitch phenomena weighs against a very plastic organization through learning. The relatively simple feedback procedure used successfully by Cross and Lane (1963) to train human subjects to respond consistently to periodicity pitch also suggests a minor role for learning. Their paradigm appears to offer a practical approach to nonhuman psychophysics of periodicity pitch, whereby learning can be more carefully controlled and studied.

No claims on physiological mechanism are made by the present theory; rather it is claimed that physiological details are outside the theory's scope. However, the present theory describes a unifying logic that underlies the behavioral phenomena of periodicity pitch and thereby defines a problem for physiology to solve.

(1) Periodicity pitch is not accountable in terms of peripheral events localized within a small range of characteristic frequencies; instead, the pitch processor synthesizes information from across central projections of the internal frequency maps from both ears.

(2) The decomposition by aural spectrum analysis of complex tone stimuli into signals representing the constituent tones (or separable groups of tones) is a necessary but not sufficient step in aural creation of periodicity pitch. When aural frequency analysis fails because of its limited resolution, periodicity pitch ceases to exist. More often, however, because of ambiguities periodicity pitch ceases as a consistent response well before frequency resolution imposes its limit.

(3) A noisy representation of only the frequencies of the aurally resolved components is communicated to the central pitch processor. This noisiness is quantified by a variance of independent Gaussian distributions which depends upon the frequency represented and not upon the presence or proximity of other resolved frequencies. This major constraint leaves to physiology the question of whether the tonotopically organized frequency signals are coded by their place or temporal course. It would be premature to regard the necessity for tonotopic organization as proof of the redundancy of temporal information in view of the uncertain problems and benefits provided by the limited dynamic range of neural firing rate and by lateral inhibitory interactions (Whitfield, 1967; Kiang, 1968; Hind *et al.*, 1971; Sachs, 1971; Suga, 1971; Tamar, 1972).

(4) The optimum processing formulation for the central pitch mechanism states the logic relating input and output while overriding all physiological details. The constraint that this mechanism presupposes stimuli comprising successive harmonics is a major clue for physiology. Such processing can be described in terms of operations by spatial filters upon tonotopically organized patterns which yield place responses (Cornsweet, 1970). But, apart from the question of the roles of time and place information in these patterns, what justification is there for assuming that periodicity pitch must be physically found in one place?

ACKNOWLEDGMENTS

I am indebted to my colleagues W. M. Siebert, H. S. Colburn, N. I. Durlach, and L. D. Braida for creating at M.I.T. the intellectual and laboratory environments for psychoacoustical research from which the reported work benefited. My association with the Eaton Peabody Laboratory of Auditory Physiology, directed by N. Y. S. Kiang, has provided valuable physiological perspective. For providing the psychophysical data that made this work possible, I thank my former thesis student and present colleague A. J. M. Houtsma. H. Duifhuis, a visitor from the Institute of Perception Research in Eindhoven, Holland, has made helpful comments. I am grateful to Jan F. Schouten who was most responsible for motivating scientific thinking on periodicity pitch. This work was supported by the National Institutes of Health (Grant 5 PO1 GM14940-06 and Grant 1 RO1 NS10737-01).

APPENDIX A: PROBABILITY OF SAMPLE IN *M*th SECTOR

Equation 9 is derived in this appendix. Define the lines bounding the mth sector by $x_2=S_{m1}x_1$ and $x_2=S_{m2}x_1$, where $S_{m1}>S_{m2}$. The probability that a sample point (x_1,x_2) falls in the mth sector is formulated in Eq. A1. For mathematical convenience, the integral A1 includes both positive and negative samples, samples outside the first quadrant, $x_1>0$, $x_2>0$, make a totally negligible contribution to the integral for parameters of interest, and in addition such samples would have no physical significance.

$$\Pr[\hat{n}=m]=\int_{-\infty}^{\infty}dx_1\int_{S_{m2}x_1}^{S_{m1}x_1}\frac{dx_2}{2\pi\sigma_1\sigma_2}\times\exp\left[-\frac{(x_1-f_1)^2}{2\sigma_1^2}-\frac{(x_2-f_2)^2}{2\sigma_2^2}\right]. \quad \text{(A1)}$$

Normalize the variables in Eq. A1 with $y_1=(x_1-f_1)/\sigma_1$ and $y_2=(x_2-f_2)/\sigma_2$, to get

$$\Pr[\hat{n}=m]=\int_{-\infty}^{\infty}dy_1\int_{l_1}^{l_2}dy_2\frac{1}{2\pi}\exp[-y_1^2/2-y_2^2/2],$$

where

$$l_1=y_2=[S_{m2}(\sigma_1y_1+f_1)-f_2]/\sigma_2, \quad \text{(A2)}$$

and

$$l_2=y_2=[S_{m1}(\sigma_1y_1+f_1)-f_2]/\sigma_2.$$

Equation A2 can be evaluated in polar coordinates in terms of cumulative joint probabilities for sample points on and below the line $l=y_2=[S_{mk}(\sigma_1y_1+f_1)-f_2]/\sigma_2$, as given by Eq. A3. Equation A2 is the difference between two such cumulative probabilities. Figure A-1 defines the polar coordinate symbols.

$$I=\int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{l} dy_2 \frac{1}{2\pi} \exp[-y_1^2/2-y_2^2/2]$$

$$=\int_0^{\infty} dR \frac{R\phi(R)}{2\pi} \exp -R^2/2. \quad \text{(A3)}$$

First we evaluate Eq. A3 for a line above the origin (see Fig. A-1).

$$\phi(R)=2\pi, \qquad 0 \leqslant R \leqslant R_{mk}$$
$$=2(\pi-\arccos[R_{mk}/R]), \quad R \geqslant R_{mk}, \qquad \text{(A4)}$$

for

$$S_{mk}f_1-f_2 \geqslant 0.$$

Substitute Eq. A4 in A3 and designate the integral as I_A.

$$I_A=\int_0^{\infty} dR \cdot R \cdot \exp -R^2/2$$
$$-\int_{R_{mk}}^{\infty} dR \frac{R}{\pi} \arccos[R_{mk}/R] \cdot \exp -R^2/2.$$

The first integral on the right is unity; the second integral is treated by parts.

$$I_A=1+\frac{1}{\pi} \arccos[R_{mk}/R] \cdot \exp -R^2/2 \,|^{\infty}_{R_{mk}}$$
$$-\frac{1}{\pi}\int_{R_{mk}}^{\infty} \frac{dR}{R} \frac{R_{mk}}{(R^2-R_{mk}^2)^{\frac{1}{2}}} \exp -R^2/2.$$

Substitution of $u^2=R^2-R_{mk}^2$ in the integral above gives a standard form tabulated in Abramowitz and Stegun (1968), yielding

$$I_A=(1+\operatorname{erf}[R_{mk}/\sqrt{2}])/2,$$

where

$$\operatorname{erf}[x]=2\pi^{-\frac{1}{2}}\int_0^x du \exp -u^2. \qquad \text{(A5)}$$

With Eq. A5, the integral A3 for a line below the origin follows by symmetry,

$$I_B=1-I_A=(1-\operatorname{erf}[R_{mk}/\sqrt{2}])/2. \qquad \text{(A6)}$$

The probabilities $\Pr[\hat{n}=m]$ given in Eq. 9 follow immediately from Eqs. A5 and A6 after identifying the sector m, as lying above, containing, or lying below the centroid of the joint probability function (i.e., the origin in Fig. A-1).

The integrals A5 and A6 are functions only of the distance from the origin to the boundary line l, as shown in Fig. A-1. This distance is readily found by noting that the sides of the similar right triangles (aob) and (aco) are related in accord with Eq. A7,

$$R_{mk}/\overline{oa}=\overline{ob}/(\overline{oa}^2+\overline{ob}^2)^{\frac{1}{2}}. \qquad \text{(A7)}$$

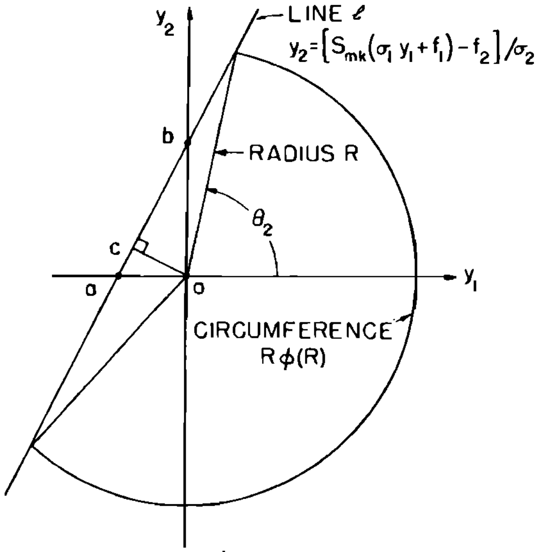

FIG. A-1. Evaluation of integrals below line l.

$\overline{oa}$ and $\overline{ob}$ are equal to the intercepts between the line l and the normalized axes; R_{mk} as given in Eq. 9 follows upon substitution.

APPENDIX B: EXPECTED VALUES OF SAMPLES IN Mth SECTOR

Equation 10 is derived in this appendix. Define the normalized variables $y_1=(x_1-f_1)/\sigma_1$ and $y_2=(x_2-f_2)/\sigma_2$. The expectations of these normalized variables, given that the sample point lies in the mth sector, are formulated in Eqs. B1 and B2,

$$\bar{Y}_{1m}=E\left[\frac{x_1-f_1}{\sigma_1}\middle|\hat{n}=m\right]=\frac{1}{\Pr[\hat{n}=m]}\int_{-\infty}^{\infty} dy_2 \int_{l_3}^{l_4} dy_1$$
$$\times\frac{y_1}{2\pi}\exp\left[-\frac{y_1^2}{2}-\frac{y_2^2}{2}\right], \qquad \text{(B1)}$$

where

$$l_3=y_1=[(\sigma_2 y_2+f_2)/S_{m1}-f_1]/\sigma_1,$$
$$l_4=y_1=[(\sigma_2 y_2+f_2)/S_{m2}-f_1]/\sigma_1,$$

$$\bar{Y}_{2m}=E\left[\frac{x_2-f_2}{\sigma_2}\middle|\hat{n}=m\right]=\frac{1}{\Pr[\hat{n}=m]}\int_{-\infty}^{\infty} dy_1 \int_{l_1}^{l_2} dy_2$$
$$\times\frac{y_2}{2\pi}\exp\left[-\frac{y_1^2}{2}-\frac{y_2^2}{2}\right], \qquad \text{(B2)}$$

where l_1 and l_2 are given in Eq. A2. These integrals can be evaluated directly in Cartesian coordinates, by the method of completing the square.

$$\bar{Y}_{1m}\cdot\Pr[\hat{n}=m]=\frac{1}{2\pi}\int_{-\infty}^{\infty} dy_2 \exp\left[-\frac{y_2^2}{2}-\frac{y_1^2}{2}\right]\Bigg|_{y_1=l_4}^{y_1=l_3}.$$

For each limit above, the argument of the exponential function can be put into the form below.

$$(-\tfrac{1}{2}[(\sigma_2/\sigma_1 S_{mk})^2+1]$$
$$\times\{y_2-(S_{mk}f_1-f_2)/[(S_{mk}\sigma_1/\sigma_2)^2+1]\sigma_2\}^2$$
$$-\tfrac{1}{2}[f_1-f_2/S_{mk}]^2/[(\sigma_2/\sigma_1 S_{mk})^2+1]\sigma_1^2).$$

By use of the definition of R_{mk} given in Eq. 9, the expression for $\bar{Y}_{1m}$ given in Eq. 10 follows directly. Comparison of Eqs. B1 and B2 reveals that the expression for $\bar{Y}_{2m}$ is obtained from $\bar{Y}_{1m}$ with the replacements $S_{m1}\leftrightarrow 1/S_{m2}$, $f_1\leftrightarrow f_2$, and $\sigma_1\leftrightarrow\sigma_2$.

APPENDIX C: VARIANCE OF FUNDAMENTAL ESTIMATE

This appendix describes an exact formulation for the variance of the linear, two-dimensional estimate of the fundamental for samples in the mth sector. With the definitions given in Eq. 8, the desired variance is given by

$$\sigma_{0m}^2=E[\hat{f}_0^2|\hat{n}=m]-(\bar{f}_{0m})^2. \qquad \text{(C1)}$$

The first moment of the estimate in the mth sector, $\bar{f}_{0m}$, is given by Eqs. 10 and 11. Equation C1 can be expressed conveniently in terms of the normalized variables $y_1=(x_1-f_1)/\sigma_1$ and $y_2=(x_2-f_2)/\sigma_2$, as given by

$$\sigma_{0m}^2=E[(A\sigma_1 y_1+B\sigma_2 y_2)^2|\hat{n}=m]-(A\sigma_1\bar{Y}_{1m}+B\sigma_2\bar{Y}_{2m})^2, \qquad \text{(C2)}$$

where

$$\hat{f}_0=Ax_1+Bx_2 \quad\text{and}\quad \bar{Y}_{km}=E[y_k|\hat{n}=m].$$

Evaluation of Eq. C2 requires the formulation of the expectations $E[y_1^2|\hat{n}=m]$, $E[y_2^2|\hat{n}=m]$ and $E[y_1y_2|\hat{n}=m]$; the mathematical procedures used are similar to those in Appendices A and B, respectively. Consider first the former expectation given by Eq. C3,

$$E[y_{1,2}^2|\hat{n}=m]=\frac{1}{\Pr[\hat{n}=m]}\int_{-\infty}^{\infty}dy_1\int_{l_1}^{l_2}dy_2\frac{y_{1,2}^2}{2\pi}\times\exp\left[-\frac{y_1^2}{2}-\frac{y_2^2}{2}\right], \qquad \text{(C3)}$$

where $y_{1,2}$ represents y_1 or y_2, l_1 and l_2 as in Eq. A2.

Equation C3 can be evaluated in polar coordinates, as in Appendix A. The integrals needed for y_1 and y_2 are given by I_1 and I_2 Eq. C4. Figure A-1 (Appendix A) defines the polar coordinate symbols.

$$I_1=\int_0^\infty dR\int_{\theta_1}^{\theta_2}R^2\cos^2\theta\frac{Rd\theta}{2\pi}\exp-R^2/2$$
$$I_2=\int_0^\infty dR\int_{\theta_1}^{\theta_2}R^2\sin^2\theta\frac{Rd\theta}{2\pi}\exp-R^2/2. \qquad \text{(C4)}$$

First we evaluate I_1 Eq. C4 for a line above the origin (see Fig. A-1).

$$\begin{aligned}\theta_2-\theta_1&=2\pi, && 0\leqslant R\leqslant R_{mk},\\ &=2(\pi-\arccos[R_{mk}/R]), && R\geqslant R_{mk},\\ \theta_2&=\pi-\arccos[R_{mk}/R]-\arccos[R_{mk}/\overline{oa}], && R\geqslant R_{mk}\end{aligned} \qquad \text{(C5)}$$

for

$$S_{mk}f_1-f_2\geqslant 0.$$

From Eq. A7,

$$R_{mk}/\overline{oa}=[(\sigma_2/\sigma_1 S_{mk})^2+1]^{-\frac{1}{2}}. \qquad \text{(C6)}$$

Integrating I_1 with respect to θ yields Eq. C7, cf. C5,

$$I_{1A}=\frac{1}{2}\int_0^{R_{mk}}dR\cdot R^3\exp-R^2/2+\frac{1}{4\pi}\int_{R_{mk}}^\infty dR[\theta_2-\theta_1+\tfrac{1}{2}\sin2\theta_2-\tfrac{1}{2}\sin2\theta_1]R^3\times\exp-R^2/2. \qquad \text{(C7)}$$

Substituting Eqs. C5 and C6 and rearranging yields C7′,

$$I_{1A}=\frac{1}{2}\int_0^\infty dR\cdot R^3\exp-R^2/2-\frac{1}{2\pi}\int_{R_{mk}}^\infty dR\left\{\arccos\frac{R_{mk}}{R}+\frac{R_{mk}}{R}\left[1-\left(\frac{R_{mk}}{R}\right)^2\right]^{\frac{1}{2}}\times\frac{(\sigma_1S_{mk}/\sigma_2)^2-1}{(\sigma_1S_{mk}/\sigma_2)^2+1}\right\}R^3\exp-R^2/2. \qquad \text{(C7}'\text{)}$$

The first integral on the right is trivial after we substitute $u=R^2/2$, which yields unity. The integral involving arc cos is solved via integration by parts, similarly as in Appendix A. The integral involving the radical is trivial after substitution of $r^2=R^2-R_{mk}^2$. Altogether, Eq. C7 is given by C8,

$$I_{1A}=\tfrac{1}{2}(1+\text{erf}[R_{mk}/\sqrt{2}])-\frac{R_{mk}}{(2\pi)^{\frac{1}{2}}}\frac{\exp-R_{mk}^2/2}{(\sigma_2/\sigma_1S_{mk})^2+1}. \qquad \text{(C8)}$$

Using Eq. C8, we evaluate I_1 Eq. C4 for a line below the origin by exploiting symmetry,

$$I_{1B}=1-I_{1A}\quad\text{(see Eq. C8).} \qquad \text{(C9)}$$

By identical methods, I_2 Eq. C4 is solved for lines above and below the origin, yielding I_{2A} and I_{2B}, respectively, Eq. C10,

$$I_{2A}=\tfrac{1}{2}(1+\text{erf}[R_{mk}/\sqrt{2}])-\frac{R_{mk}}{(2\pi)^{\frac{1}{2}}}\frac{\exp-R_{mk}^2/2}{(S_{mk}\sigma_1/\sigma_2)^2+1},$$
$$I_{2B}=1-I_{2A}. \qquad \text{(C10)}$$

Defining as a principal sector, n, one in which (f_1,f_2) lies or borders, we use Eqs. C8–C10 to develop the desired formulae C11,

$$E[y_1^2|\hat{n}=m=n]$$
$$=1-\left[\frac{R_{m1}\exp-R_{m1}^2/2}{(\sigma_2/\sigma_1 S_{m1})^2+1}+\frac{R_{m2}\exp-R_{m2}^2/2}{(\sigma_2/\sigma_1 S_{m2})^2+1}\right]$$
$$\times\{(2\pi)^{\frac{1}{2}}\Pr[\hat{n}=m]\}^{-1},$$

$$E[y_1^2|\hat{n}=m<n]$$
$$=1+\left[\frac{R_{m2}\exp-R_{m2}^2/2}{(\sigma_2/\sigma_1 S_{m2})^2+1}-\frac{R_{m1}\exp-R_{m1}^2/2}{(\sigma_2/\sigma_1 S_{m1})^2+1}\right] \quad \text{(C11)}$$
$$\times\{(2\pi)^{\frac{1}{2}}\Pr[\hat{n}=m]\}^{-1},$$

$$E[y_1^2|\hat{n}=m>m]$$
$$=1+\left[\frac{R_{m1}\exp-R_{m1}^2/2}{(\sigma_2/\sigma_1 S_{m1})^2+1}-\frac{R_{m2}\exp-R_{m2}^2/2}{(\sigma_2/\sigma_1 S_{m2})^2+1}\right]$$
$$\times\{(2\pi)^{\frac{1}{2}}\Pr[\hat{n}=m]\}^{-1}.$$

The probabilities $\Pr[\hat{n}=m]$ and the distances R_{mk} are given in Eq. 9. The conditional expectations for y_2^2 are given by Eq. C11 after making the replacement $(\sigma_2/\sigma_1 S_{mk})^2 \leftrightarrow (S_{mk}\sigma_1/\sigma_2)^2$.

The derivation of the expectation $E[y_1y_2|\hat{n}=m]$ follows that of Appendix B so closely that only the result need be given here Eq. C12. The sign function introduced in Eq. C12 can be used to express Eqs. 9 and C11 more compactly for computer programming.

$$E[y_1y_2|\hat{n}=m]=\left\{\frac{\mathrm{sgn}[S_{m1}f_1-f_2]\cdot R_{m1}\exp-R_{m1}^2/2}{S_{m1}\sigma_1/\sigma_2+\sigma_2/\sigma_2 S_{m1}}\right.$$
$$\left.-\frac{\mathrm{sgn}[S_{m2}f_1-f_2]\cdot R_{m2}\exp-R_{m2}^2/2}{S_{m2}\sigma_1/\sigma_2+\sigma_2/\sigma_1 S_{m2}}\right\}$$
$$\times\{(2\pi)^{\frac{1}{2}}\Pr[\hat{n}=m]\}^{-1}. \quad \text{(C12)}$$

Substituting the expectations C11 and C12 in C2 gives us the desired expression for the variance σ_{0m}^2.

*Now with School of Enginneering, Tel Aviv University, Ramat Aviv, Tel Aviv, Israel.

Abramowitz, M., and Stegun, I. A. (1968). *Handbook of Mathematical Functions*, U.S. Natl. Bur. Std. (U.S. GPO, Washington, D.C.), p. 302.

von Békésy, G. (1960). *Experiments in Hearing* (McGraw–Hill, New York).

Bilsen, F. A. (1968). "On the Interactions of a Sound with its Repetitions," Ph.D. Thesis, Delft University of Technology.

de Boer, E. (1956). "On the Residue in Hearing," Ph.D. Thesis, University of Amsterdam.

van den Brink, G. (1970). "Two Experiments on Pitch Perception: Diplacusis of Harmonic AM Signals and Pitch of Inharmonic AM Signals," J. Acoust. Soc. Am. **48,** 1355–1365.

Colburn, H. S. (1969). "Some Physiological Limitations on Binaural Performance," Ph.D. Thesis, Dept. of Elec. Eng., M.I.T., Cambridge, Mass.

Cornsweet, T. N. (1970). *Visual Perception* (Academic, New York).

Cross, D., and Lane, H. (1963). "Attention to Single Stimulus Properties in the Identification of Complex Tones," in *Experimental Analysis of the Control of Speech Production and Perception*, U. of Michigan, ORA Rept. No. 05613-1-P (U. of Michigan, Ann Arbor, Michigan).

Deutsch, D. (1972). "Octave Generalization and Tune Recognition," Percept. Psychophys. **11,** 411–412.

Fischler, H. (1967). "Model of the Secondary Residue Effect in the Perception of Complex Tones," J. Acoust. Soc. Am. **42,** 759–764.

Fischler, H., and Cern, L. (1968). "Simulation of the Secondary Residue Effect by Digital Computer," J. Acoust. Soc. Am. **44,** 1379–1385.

Flanagan, J. L., and Guttman, N. (1960). "On the Pitch of Periodic Pulses," J. Acoust. Soc. Am. **32,** 1308–1319.

Fletcher, H. (1924). "The Physical Criterion for Determining the Pitch of a Musical Tone," Phys. Rev. **23,** 427–437.

Goldstein, J. L. (1965). "An Investigation of Monaural Phase Perception," Ph.D. Thesis, Univ. of Rochester, New York (University Microfilms 66–6852).

Goldstein, J. L. (1967a). "Auditory Nonlinearity," J. Acoust. Soc. Am. **41,** 676–689.

Goldstein, J. L. (1967b). "Auditory Spectral Filtering and Monaural Phase Perception," J. Acoust. Soc. Am. **41,** 458–479.

Goldstein, J. L. (1970). "Aural Combination Tones," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (A. W. Sijthoff Leiden, The Netherlands), pp. 230–247.

Goldstein, J. L. (1972). "Evidence from Aural Combination Tones and Musical Tones Against Classical Temporal Periodicity Theory," Symposium on Hearing Theory (IPO Eindhoven, Holland), pp. 186–208.

Goldstein, J. L., and Kiang, N. Y. S. (1968). "Neural Correlates of the Aural Combination Tone $2f_1-f_2$," Proc. IEEE **56,** 981–992.

Goldstein, J. L., Baer, T., and Kiang, N. Y. S. (1971). "A Theoretical Treatment of Latency, Group Delay, and Tuning Characteristics for Auditory-Nerve Responses to Clicks and Tones," in *Physiology of the Auditory System*, edited by M. B. Sachs (National Educational Consultants, Baltimore, Md.), pp. 133–141.

Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hall, J. L. (1972). "Auditory Distortion Products f_2-f_1 and $2f_1-f_2$," J. Acoust. Soc. Am. **51,** 1863–1871.

Harris, J. D. (1952). "Pitch Discrimination," J. Acoust. Soc. Am. **24,** 750–755.

von Helmholtz, H. L. F. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (F. Vieweg und Sohn, Braunschweig, Germany). Translated as: *On the Physiological Basis for the Theory of Music*, by A. J. Ellis from the 4th German edition, 1877, Leymans, London, 1885 (reprinted by Dover, New York, 1954).

Henning, G. B. (1966). "Frequency Discrimination of Random Amplitude Tones," J. Acoust. Soc. Am. **39,** 336–339.

Hermann, L. (1912). "Neue Versuche zur Frage der Unterbrechungstöne," Arch. ges. Physiol. **146,** 249–294.

Hind, J. E., Rose, J. E., Brugge, J. F., and Anderson, D. J. (1971). "Some Effects of Intensity on the Discharge of Auditory-Nerve Fibers," in *Physiology of the Auditory System*, edited by M. B. Sachs (National Educational Consultants, Baltimore, Md.), pp. 101–111.

Houtsma, A. J. M. (1971). "Perception of Musical Intervals: The Role of Aural Frequency Resolution in Tracking the Missing Fundamental," Ph.D. Thesis, Dept. of Elec. Eng., M.I.T., Cambridge, Mass.

Houtsma, A. J. M., and Goldstein, J. L. (1970). "Aural

Tracking of Musical Notes: The Problem of the Missing Fundamental," Quarterly Progress Report No. 98, Research Laboratory of Electronics, M.I.T., July 15, 1970, pp. 195–203.

Houtsma, A. J. M., and Goldstein, J. L. (1971). *Perception of Musical Intervals: Evidence for the Central Origin of Musical Pitch,* Technical Report 484, Research Laboratory of Electronics, M.I.T., October 1, 1971.

Houtsma, A. J. M., and Goldstein, J. L. (1972). "The Central Origin of the Pitch of Complex Tones: Evidence from Musical Interval Recognition," J. Acoust. Soc. Am. 51, 520–529.

Kiang, N. Y. S. (1965). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (M.I.T. Press, Cambridge, Mass.).

Kiang, N. Y. S. (1968). "A Survey of Recent Developments in the Study of Auditory Physiology," Ann. Otol. Rhinol. Laryngol. 77, 656–675.

Licklider, J. C. R. (1954). "Periodicity Pitch and Place Pitch," J. Acoust. Soc. Am. 26, 945(A).

Licklider, J. C. R. (1955). "Influence of Phase Coherence upon the Pitch of Complex Periodic Sounds," J. Acoust. Soc. Am. 27, 996(A).

Licklider, J. C. R. (1956). "Auditory Frequency Analysis," in *Information Theory*, edited by C. Cherry (Butterworths Scientific Publications, London), pp. 253–268.

Mathes, R. C., and Miller, R. L. (1947). "Phase Effects in Monaural Phase Perception," J. Acoust. Soc. Am. 19, 780–797.

Ohm, G. S. (1843). "Ueber die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen," Ann. Phys. Chem. 59, 513–565.

Plomp, R. (1964). "The Ear as a Frequency Analyzer," J. Acoust. Soc. Am. 36, 1628–1636.

Plomp, R. (1965). "Detectability Threshold for Combination Tones," J. Acoust. Soc. Am. 37, 1110–1123.

Plomp, R. (1967). "Pitch of Complex Tones," J. Acoust. Soc. Am. 41, 1526–1533.

Plomp, R., and Smoorenburg, G. F. (1970). *Frequency Analysis and Periodicity Detection in Hearing* (A. W. Sijthoff, Leiden, The Netherlands).

Ritsma, R. J. (1962). "Existence Region of the Tonal Residue. I," J. Acoust. Soc. Am. 34, 1224–1229.

Ritsma, R. J. (1963). "On Pitch Discrimination of Residue Tones," Int. Audiol. 2, 34–37.

Ritsma, R. J. (1967a). "Frequencies Dominant in the Perception of Pitch of Complex Sounds," J. Acoust. Soc. Am. 42, 191–198.

Ritsma, R. J. (1967b). "Frequencies Dominant in Pitch Perception of Periodic Pulses of Alternating Polarity," Annual Progress Report (IPO, Eindhoven, Holland), pp. 14–24.

Ritsma, R. J. (1970). "Periodicity Detection," in *Frequency Analysis and Periodicity Detection*, edited by R. Plomp and G. F. Smoorenburg (A. W. Sijthoff, Leiden, The Netherlands), pp. 250–266.

Ritsma, R. J., and Engel, F. L. (1964). "Pitch of Frequency Modulated Signals," J. Acoust. Soc. Am. 36, 1637–1644.

Sachs, M. B. (1971). *Physiology of the Auditory System* (National Educational Consultants, Baltimore, Md.).

Sachs, M. B. (1971). "Quantitative Studies of the Responses of Auditory-Nerve Fibers to Two-Tone Stimuli," in *Physiology of the Auditory System*, edited by M. B. Sachs (National Educational Consultants, Baltimore, Md.), pp. 113–123.

Schouten, J. F. (1938). "The Perception of Subjective Tones," Proc. K. Ned. Akad. Wet. 41, 1086–1093.

Schouten, J. F. (1940a). "The Residue and the Mechanism of Hearing," Proc. K. Ned. Akad. Wet. 43, 991–999.

Schouten, J. F. (1940b). "The Perception of Pitch," Philips Tech. Rev. 5, 286–294.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (1962). "Pitch of the Residue," J. Acoust. Soc. Am. 34, 1418–1424.

Schroeder, M. R. (1966). "Residue Pitch: A Remaining Paradox and a Possible Explanation," J. Acoust. Soc. Am. 40, 79–81.

Seebeck, A. (1841). "Beobachtungen über einige Bedingungen der Entstehung von Tönen," Ann. Phys. Chem. 53, 417–436.

Seebeck, A. (1843). "Ueber die Definition des Tones," Ann. Phys. Chem. 60, 449–481.

Shower, E. G., and Biddulph, R. (1931). "Differential Pitch Sensitivity of the Ear," J. Acoust. Soc. Am. 3, 275–287.

Siebert, W. M. (1968). "Stimulus Transformations in the Peripheral Auditory System," in *Recognizing Patterns*, edited by P. A. Kolers and M. Eden (M.I.T. Press, Cambridge, Mass.), pp. 104–133.

Siebert, W. M. (1970). "Frequency Discrimination in the Auditory System: Place or Periodicity Mechanisms?" Proc. IEEE 58, 723–730.

Small, A. M., Jr. (1970). "Periodicity Pitch," in *Foundations of Modern Auditory Theory*, edited by J. R. Tobias (Academic, New York), Vol. I, pp. 1–54.

Smoorenburg, G. F. (1970). "Pitch Perception of Two-Frequency Stimuli," J. Acoust. Soc. Am. 48, 924–942.

Suga, N. (1971). "Feature Detection in the Cochlear Nucleus, Inferior Colliculus, and Auditory Cortex," in *Physiology of the Auditory System*, edited by M. B. Sachs (National Educational Consultants, Baltimore, Md.), pp. 197–206.

Sutton, R. A., and Williams, R. P. (1970). "Residue Pitches from Two-Tone Complexes," J. Sound Vib. 13, 195–199.

Tamar, H. (1972). *Principles of Sensory Physiology* (Ch. C. Thomas, Springfield, Illinois).

Terhardt, E. (1972a). "Zur Tonhöhenwahrnehmung von Klängen I. Psychoakustische Grundlagen," Acustica 26, 173–186.

Terhardt, E. (1972b). "Zur Tonhöhenwahrnehmung von Klängen II. Ein Funktionsschema," Acustica 26, 187–199.

Thurlow, W. R. (1963). "Perception of Low Auditory Pitch," Psychol. Rev. 70, 461–470.

van Trees, H. L. (1968). *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York).

Walliser, K. (1969). "Ueber ein Funktionsschema für die Bildung der Periodentonhöhe aus dem Schallreiz," Kybernetik 6, 65–72.

Whitfield, I. C. (1967). *The Auditory Pathway* (Edward Arnold and Co., London).

Whitfield, I. C. (1970). "Neural Integration and Pitch Perception," in *Excitatory Synaptic Mechanisms*, edited by P. Andersen and J. K. S. Jansen (Universitetsforlaget, Oslo), pp. 277–285.

Wightman, F. L. (1972). "Pitch as Auditory Pattern-Recognition," Symposium on Hearing Theory (IPO, Eindhoven, Holland), pp. 161–171.

Zwicker, E. (1955). "Der ungewöhnliche Amplitudengang der nichtlinearen Verzerrungen des Ohres," Acustica 5, 67–74.