AN OPTIMUM PITCH PROCESSING MODEL FOR SIMULTANEOUS COMPLEX TONES

Adrianus. J.M. Houtsma and John. G. Beerends

Institute for Perception Research, P.O. Box 513,
5600 MB Eindhoven, The Netherlands.

ABSTRACT

An extension of Goldstein's Optimum Processing Theory is presented which can account for pitch perception behavior for simultaneous complex tones. The essence of the theory is that all aurally resolved stimulus frequencies are transformed into independent Gaussian random variables with a variance that depends only on the frequency of each partial. A central processor is assumed to use its prior knowledge about the number of simultaneously present tone complexes and the proper parsing of the observed random variables to find the respective fundamentals of the best fitting harmonic templates. In a series of pitch identification experiments for two simultaneous two-tone complexes with diotically and dichotically distributed partials, some model assumptions and their consequences were tested. It was found that (1) the processes of estimating two simultaneous (missing) fundamentals are to a large extent independent, (2) that the central processor tends to group the partial percepts on the basis of common fundamental and not on the basis of ear input, and (3) that pitch identification performance degrades only noticeably if none of the stimulus partials of both tone complexes are aurally resolved.

INTRODUCTION

The problem how we perceive the pitch of complex tones has kept psychoacousticians busy for more than a century. In particular the problem of the so called "missing fundamental", a pitch percept that corresponds with the fundamental frequency of a harmonic tone complex while that complex actually has only overtones, has been the object of many experimental and theoretical studies. Various pieces of important empirical evidence and theories to account for such evidence have been brought forward by Seebeck [1], Ohm [2], Helmholtz [3], Fletcher [4], Schouten [5] and Békésy [6].

More recent experiments by Plomp [7], Ritsma [8] and Houtsma and Goldstein [9] have progressively shown that the real cause of the "missing fundamental" phenomenon must not be sought in the peripheral, but rather in the central part of the auditory system. The new experimental evidence has led to the formulation of some new central pitch theories, of which the Virtual Pitch Theory of Terhardt [10] and the Optimum Processor Theory of Goldstein [11] are the principal variants. These theories were developed and quantified mostly on the basis of pitch perception data obtained with isolated complex tones or short sequences of such tones.

In music, especially in the Western hemisphere, we usually deal with harmonic or polyphonic sound patterns in which either a melody is accompanied with chords or several melodies are played simultaneously against one another. This poses the interesting problem how our auditory system is able to perceive two or more simultaneous pitches when it is acoustically exposed to a cluster of harmonics that belong to several different tone complexes. The same problem actually occurs when one tries to track the prosodic contours of two simultaneously spoken sentences or, more realistically, when one tries to follow the pitch contour of one spoken sentence against a background of other speech. Although both central pitch theories mentioned [10,11] are in principle able to cope with this problem, this has never been worked out specifically or tested against systematic empirical data.

In this study the Optimum Processor Theory of Goldstein will be extended and tested with experimentally obtained pitch identification data for two simultaneous complex tones. The model extension will be treated in Sect. I. Descriptions of the experimental procedure and the results are given in Sects. II and III. Computer simulations of model performance are discussed in Sect. IV, and conclusions of the study are presented in Sect. V.

I. EXTENSION OF THE OPTIMUM PROCESSOR THEORY

In Goldstein's Optimum Processor Theory [11] and in a later extension of that theory [12] it was assumed that :

1. the complex tone input in both ears is spectrally analyzed and only frequency information of sufficiently resolved partials is kept; phase and amplitude information is discarded;
2. independent Gaussian random variables r_i, of zero mean and with variance depending on frequency only, are added to each resolved frequency to form the noisy frequeny codes $x_i = f_i + r_i$;
3. a central processor rank-orders all noisy frequency codes from both ears and performs a maximum-likelihood estimate of the best-fitting harmonic numbers and fundamental of some underlying harmonic complex-tone template.

This model, which was originally formulated to describe perception of a single pitch from a single complex tone, can easily be extended to accomodate identification tasks of pitches from simultaneously sounding complex tones. In this study we will focus on the task of identifying two fundamental pitches in an acoustic stimulus that comprises two simultaneous two-tone complexes, each one having successive harmonics. Extension of the model to other cases, e.g., three or four simultaneous two-tone complexes or two simultaneous multi-tone complexes, is, in principle, not different but may be computationally more complex.

Suppose now that the acoustic stimulus consists of four frequencies: $f_1 = m f_{01}$, $f_2 = (m+1) f_{01}$, $f_3 = n f_{02}$, and $f_4 = (n+1) f_{02}$, and that these frequencies are all peripherally resolved by the auditory system. The frequencies f_1 through f_4 are then transformed into four independent Gaussian random variables x_1 through x_4, having means of f_1 through f_4 respectively, and standard deviations $\sigma(f_1)$ through $\sigma(f_4)$. If we denote $\sigma(f_i)$ simply as σ_i, the likelihood function to be optimized by the processor is given by the expression:

$$
\begin{aligned}
L(f_1, f_2, f_3, f_4) \;=\; & \frac{1}{4\pi^2\sigma_1\sigma_2\sigma_3\sigma_4}.exp[-\frac{(x_1-f_1)^2}{2\sigma_1^2}]. \quad (1)\\
& exp[-\frac{(x_2-f_2)^2}{2\sigma_2^2}].exp[-\frac{(x_3-f_3)^2}{2\sigma_3^2}].\\
& exp[-\frac{(x_4-f_4)^2}{2\sigma_4^2}].
\end{aligned}
$$

Maximizing Eq.(1) is equivalent to maximizing the log-likelihood function:

$$
\begin{aligned}
\Lambda(f_1, f_2, f_3, f_4) \;=\; & -\frac{(x_1-f_1)^2}{\sigma_1^2} - \frac{(x_2-f_2)^2}{\sigma_2^2} - \quad (2)\\
& -\frac{(x_3-f_3)^2}{\sigma_3^2} - \frac{(x_4-f_4)^2}{\sigma_4^2}.
\end{aligned}
$$

In interpreting this log-likelihood function, the knowledge the processor has about the make-up of the stimulus and the task to be performed becomes very important. We will fist discuss the case (A) in which the processor has full knowledge of the fact that there are two two-tone complexes, and hence two pitches to be found, as well as knowledge of the correct parsing, i.e., of the correct harmonic interpretation of each observed input x_i. We will then discuss another case (B) where the number of complex tones present is known, but the correct parsing is unknown to the processor.

Case A. When the number of fundamental pitches to be identified and also all parsing information is available to the processor, it makes the following substitutions in Eq.(2):

$$
\begin{aligned}
f_1 &= \hat{m}\hat{f}_{01} \quad (3)\\
f_2 &= (\hat{m}+1)\hat{f}_{01}\\
f_3 &= \hat{n}\hat{f}_{02}\\
f_4 &= (\hat{n}+1)\hat{f}_{02}
\end{aligned}
$$

and maximizes the expression with respect to the (lower) harmonic number estimates $\hat{m}$ and $\hat{n}$ and the fundamental pitch estimates $\hat{f}_{01}$ and $\hat{f}_{02}$. Because of the statistical independence of the input variables x_i, the first two terms and the last two terms of Eq. (2) can be maximized separately. The two independent fitting procedures, each one identical to the one described by Goldstein [11], yield the optimum harmonic-number estimates $\hat{m}$ and $\hat{n}$ as well as the fundamental pitch estimates $\hat{f}_{01}$ and $\hat{f}_{02}$. The probabilities $\Pr[\hat{m} = k]$ and $\Pr[\hat{n} = l]$, with k and l being integers, are discrete probabilities which can be computed from the stimulus frequencies f_i and the fixed and known frequency coding noise function $\sigma(f_i)$, and the fundamental pitch estimates are given by the expressions:

$$
\begin{aligned}
\hat{f}_{01} &= \frac{[x_1/\hat{m}]^2 + [x_2/(\hat{m}+1)]^2}{x_1/\hat{m} + x_2/(\hat{m}+1)} \quad (4)\\
\hat{f}_{02} &= \frac{[x_3/\hat{n}]^2 + [x_4/(\hat{n}+1)]^2}{x_3/\hat{n} + x_4/(\hat{n}+1)}.
\end{aligned}
$$

The probability density functions of the estimates $\hat{f}_{01}$ and $\hat{f}_{02}$ are nearly-discrete functions with the main modes at f_{01} and f_{02}, the correct fundamental estimates, with probabilities $\Pr[\hat{m} = m]$ and $\Pr[\hat{n} = n]$ respectively. Correct identification of the two pitches therefore boils down to two independent correct identifications of the respective lower harmonic numbers m and n.

Case B. When the processor only knows the number of fundamental pitches to be identified, but does not have any information about the proper parsing of the input variables x_i, it tries, in principle, all possible interpretations of the x_is which are, in this case, 24 permutations. In practice, only the following three permutations are relevant in most cases because of simple ordinal properties of the input variables and their possible interpretations:

(a)	(b)	(c)
$f_1 = \hat{m}\hat{f}_{01}$	$f_1 = \hat{m}\hat{f}_{01}$	$f_1 = \hat{m}\hat{f}_{01}$
$f_2 = (\hat{m}+1)\hat{f}_{01}$	$f_2 = \hat{n}\hat{f}_{02}$	$f_2 = \hat{n}\hat{f}_{02}$
$f_3 = \hat{n}\hat{f}_{02}$	$f_3 = (\hat{m}+1)\hat{f}_{01}$	$f_3 = (\hat{n}+1)\hat{f}_{02}$
$f_4 = (\hat{n}+1)\hat{f}_{02}$	$f_4 = (\hat{n}+1)\hat{f}_{02}$	$f_4 = (\hat{m}+1)\hat{f}_{01}$

Group (a), of course, represents the correct parsing, but the interpretations of (b) and (c) may result in a larger likelihood function value and therefore in a better fit on a given trial because of the noise in the variables x_i. Interpretations (b) and (c) will almost always lead to incorrect pitch identifications, however. We will refer to such mistakes as parsing errors.

It is far from clear whether or not the extension of the Optimum Processor Theory as it has been described so far offers a realistic account of human pitch perception for situations of simultaneous complex tones. Some particular questions remain to be answered. Are the fundamental estimation processes for each complex tone in a chord really independent ? Does the processor actually have knowledge of the correct parsing and interpretation of the perceived partials, or can such knowledge be externally supplied ? When two partials of different complex tones have exactly or almost the same frequency, are they both unavailable to the processor because they are peripherally unresolved, or is some frequency information still transmitted to the processor ? These questions are investigated in the following set of experiments.

II. EXPERIMENTS

Musically experienced subjects performed a series of pitch identification experiments with two simultaneously sounding notes, each note made with a harmonic two-tone complex. One complex, representing the lower note, comprised the frequencies $f_1 = m f_{01}$, $f_2 = (m+1) f_{01}$, the other complex representing the higher note the frequencies $f_3 = n f_{02}$, $f_4 = (n+1) f_{02}$. The respective fundamentals f_{01} and f_{02} were both elements of the note set {do, re, mi, fa, so} or, equivalently, the frequency set {200, 225, 250, 267, 300} Hz, and could not be the same on any given trial. Both lower harmonic numbers m and n were independent random integers beteen 2 and 10. Note durations were 600 ms and intensities were 20 dB above threshold, with 30-dB SL broadband noise as a general masking background. Sound stimuli, which were computed and stored on a Philips P857 minicomputer, were played back through a 2-channel, 12-bit D/A converter and presented through headphones to the subject who was seated in a sound-insulated chamber. The task of the subject was to identify both simultaneously perceived (missing) fundamentals f_{01} and f_{02} on each trial by pressing two out of five buttons on a response box in any temporal order. There was unlimited response time, and each response triggered presentation of a new trial after a brief fixed delay.

Four stimulus conditions were investigated. In condition 1, which was diotic, all four stimulus frequencies f_1 through f_4 were presented to both ears. In condition 2, which was dichotic, one note (comprising the frequencies f_1 and f_2) was presented to one ear, while the other note (with the frequencies f_3 and f_4) went to the other ear. In conditions 3 and 4, which were also dichotic, the frequencies of both notes were split up between the ears. In condition 3, one ear received f_1 and f_3, while the other ear received f_2 and f_4. In condition 4, one ear received f_1 and f_4 while the other ear received f_2 and f_3.

Since there are ten different combinations of two notes in a total set of five, and since there were $9 \times 9 = 81$ different harmonic representations of each two-note combination, there was a total of 810 physically different stimuli and a total of 10 different response categories. Each of these stimuli was, on the average, presented six times to each of four subjects, for a total of 4500 identification trials per subject for each stimulus condition.

III. RESULTS

The raw data of all experiments consisted of a record for each trial of the presented fundamentals f_{01} and f_{02}, the lower harmonic numbers m and n, and the subject's two responses R_a and R_b. A response (R_a, R_b) could be an identification of the perceived fundamentals (f_{01}, f_{02}) or (f_{02}, f_{01}), since the order of pressing the response buttons was arbitrary.

To obtain some insight in the perceptual independence of the identification processes for each of the two simultaneous notes, the raw data were processed by two different methods. In the first method all trials were counted for every (m, n) combination where both f_{01} and f_{02} were identified correctly. The results of this way of counting yielded half-matrices of 'percent correct' scores, $Pc(k, l)$, for each subject, in which k and l are integers representing the harmonic numbers (m, n) or (n, m). They are half-matrices because of the built-in symmetry around the main diagonal, which makes both halves of the matrix mirror images. In the second method only the correct identification of one of the two simultaneous notes was considered as a function of both (lower) harmonic numbers but regardless of the identification response to the other note. The resulting score, designated as $Pc(k|l)$, represents the percentage correct identifications of f_{01} for $k = m$ and $l = n$, as well as the correct identifications of f_{02} for $k = n$ and $l = m$. The total count for each subject yielded full 9x9 matrices.

Both processed data matrices $Pc(k, l)$ and $Pc(k|l)$ can be used to find an underlying processor performance function $\Pr[\hat{k} = k]$, the processor's probability of correctly estimating the harmonic order of any complex tone. This was done with a minimum chi-square fitting procedure which looked for those $\Pr[\hat{k} = k]$ functions that provided the most likely account of the empirically obtained data matrices $Pc(k, l)$ and $Pc(k|l)$. The details of this procedure, which also involved some assumptions about the decision process for the particular experimental paradigm that was used, are discussed in a recent publication by the authors [13].

The functions $\Pr_1[\hat{k} = k]$ derived from the matrix $Pc(k|l)$and $\Pr_2[\hat{k} = k]$ derived from $Pc(k, l)$ are shown in Fig. 1a-d as triangles and squares respectively for the experimental conditions 1 through 4. One can show that, if $\Pr_1[\hat{k} = k] > \Pr_2[\hat{k} = k]$ for low values of k and $\Pr_1[\hat{k} = k] < \Pr_2[\hat{k} = k]$ for large values of k, the two fundamental pitch identification processes are mutually dependent in the sense that the perception of the more salient pitch, i.e., the one represented by the lowest harmonic numbers, inhibits correct perception of the less salient pitch [13]. Figure 1a-d shows that in condition 2 only subject JH noticeably exhibits this effect of mutual dependence of the two identification processes, but in conditions 1, 3 and 4 all subjects except MZ seem to show a small amount of mutual dependence.

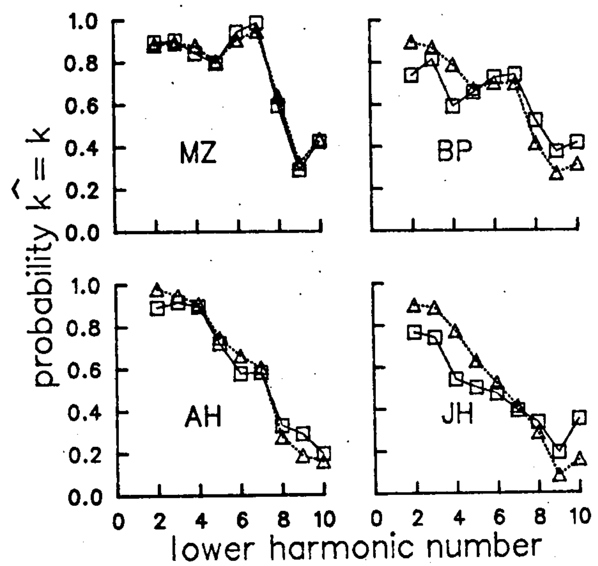

Fig. 1a. The processor's probability of identifying the correct harmonic order of a complex tone. The harmonic order is shown on the abscissa. Triangles designate $\Pr_1[\hat{k} = k]$, squares $\Pr_2[\hat{k} = k]$. Computed from data of condition 1.

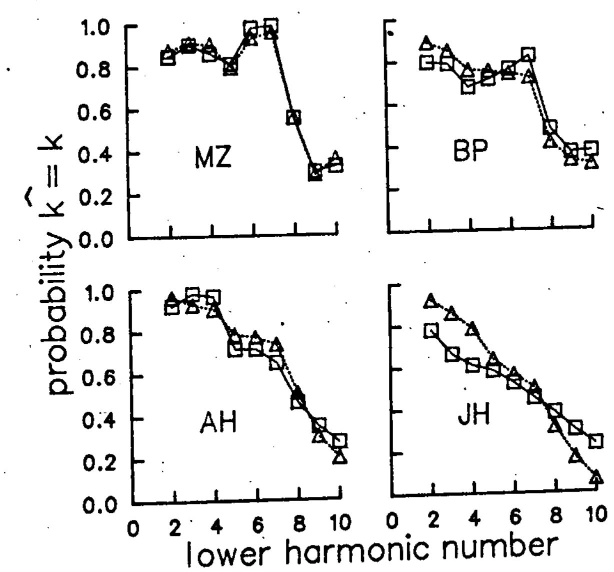

Fig. 1b. Same as Fig 1a, but computed from data of condition 2.

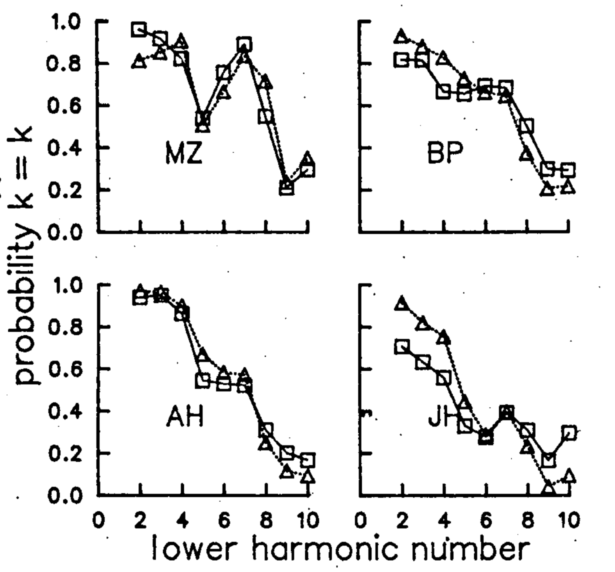

Fig. 1c. Same as Fig 1a, but computed from data of condition 3.

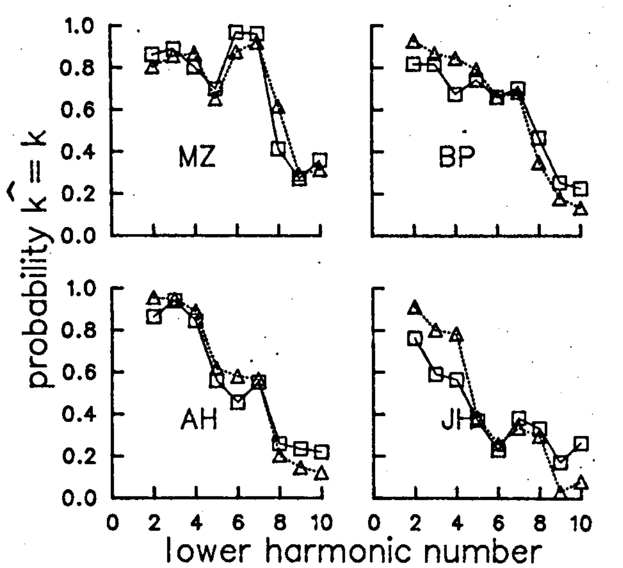

Fig. 1d. Same as Fig 1a, but computed from data of condition 4.

From either probability function $Pr_1[\hat{k} = k]$ or $Pr_2[\hat{k} = k]$ one can now compute the model's variance function $\sigma(f)$ which represents the frequency coding noise and is its only free parameter. A set of those sigma functions is shown in Fig. 2. The functions were computed from the averaged $Pr_1[\hat{k} = k]$ and $Pr_2[\hat{k} = k]$ functions obtained from the experimental data of dichotic condition 2. The $\sigma(f)/f$ functions have the typical U-shape which was also found in an earlier study [11], and have also the same general magnitude. The low-frequency slopes of these functions, however, are much steeper than the average slope found in that earlier study. We think this is caused by an over-estimate of $\sigma(f)$ at low frequencies in the present experiments. Partial frequencies below 1000 Hz, with fundamentals limited between 200 and 300 Hz, could occur only for very low harmonic numbers where identification is close to perfect and occasional mistakes are more made through carelessness or poor attention than through insufficient salience of pitches.

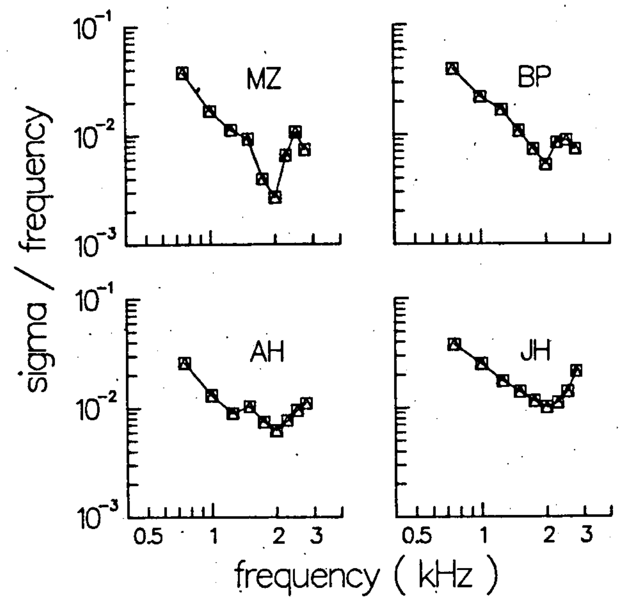

Fig. 2. Model variance or "noise" functions $\sigma(f)/f$ computed from the averaged functions Pr_1 and Pr_2 shown in Fig. 1b.

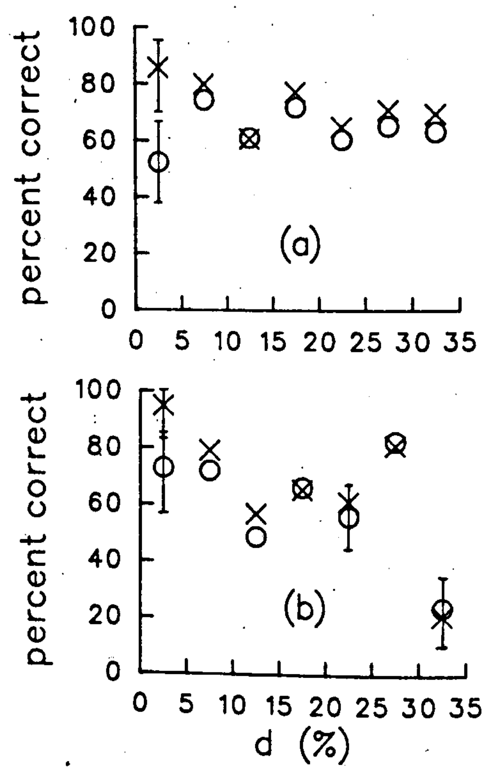

Fig. 3. Correct identification scores for both simultaneous fundamentals as a function of d defined by Eq. (5). (a) Diotic condition 1 (circles) and dichotic condition 2 (crosses). (b) Dichotic conditions 3 (circles) and 4 (crosses).

In order to examine the influence of spectral interference on pitch identification performance, all 810 different stimuli were mapped on a frequency-difference measure d, defined as:

$$d = \sqrt{d_1^2 + d_2^2}, \tag{5}$$

where d_1 represents the smallest frequency difference between any two harmonics and d_2 the next smallest difference in the total four-tone stimulus. Stimuli were grouped on this d-scale in bins of 5 % and, in order to limit the general degrading effect of high harmonic numbers on pitch identification performance, only stimuli were included having lower harmonic numbers m and n of 5 or less. Percentages correct pitch identification of both fundamentals as a function of d are shown in Fig. 3a for experimental conditions 1 and 2. Under diotic condition 1 values of d below 10 % imply that the partials mf_{01}, nf_{02} as well as the partials $(m+1)f_{01}, (n+1)f_{02}$ must have interfered with one another because of limited frequency resolution in the cochlea. Under dichotic condition 2 such interference was not possible because potentially interfering partials went to different ears. Figure 3a shows that only for the lowest d-values, between 0 and 5 %, there is a noticeable difference between the scores of conditions 1 and 2. The figure also shows, however, that performance for diotic condition 1, although degraded, is still well above the expected chance level of 10 % correct. Similar results were obtained with the data from conditions 3 and 4. They are shown in Fig. 3b.

IV. MODEL SIMULATIONS

The data presented in the previous section show a general performance deterioration with increasing (lower) harmonic numbers m and n, and also a dependence of performance on the the presentation conditions 1 through 4. The data still provide insufficient information, however, about the relative contribution of parsing errors compared with errors caused by interference of partials or mutual dependence of pitch identification processes. To study the influence of parsing errors in more detail, a computer simulation experiment was performed with the model discussed in Sect. I. To simulate each subject's performance, the $\sigma(f)/f$ functions derived from the data of condition 2 were substituted in the model to specify the exact amount of noise to be added to each frequency component of the simulation input. For all 810 stimuli 25 computations were made (with new noise samples added to partials each time) of the maximum log-likelihood function of Eq. (2) without knowledge of the correct parsing, as outlined in Case B of Sect. I. Simulations were made on a Vax 11/780 computer. Those stimuli for which the correct parsing was always obtained were put in a stimulus subset PNS (parsing-non-sensitive). The remaining stimuli, for which (occasionally) the likelihood function came out maximum with the wrong parsing, were put in the subset PS (parsing-sensitive). With the subsets PS, PNS and also with the entire set PS+PNS, the simulation experiment was now repeated for all four stimulus conditions (1 through 4) and with substitution of the appropriate $\sigma(f)/f$ function obtained from the data of each particular subject under that condition with stimulus subset PNS. Simulation was done with the use of Eq. (3), implying knowledge of the correct stimulus parsing, and with substitution of all 24 possible permutations outlined under Case B of Sect. I, implying the absence of this knowledge. The results, expressed as a percentage correct identifications of both fundamentals f_{01} and f_{02} pooled over all values of m and n, are shown in Fig. 4 for stimulus conditions 1 and 2. For each of the three stimulus (sub)sets, the solid line represents the actual performance of the subject, the dashed line the performance level simulated with parsing knowledge, and the dotted line the level simulated without this knowledge. For the PNS-subset one expects all three performance levels to be identical. The fact that this is not exactly the case is not a truncation effect in which the number of experimental and simulated trials was smaller than required by the Law of Large Numbers, but represents a small uncertainty about the details of the simultated decision strategy. One also observes that for the subset PS and for the entire stimulus set PS+PNS the performance level of all subjects (solid lines) is much closer to the performance level simulated with parsing knowledge (dashed lines) than to the level simulated without that knowledge (dotted lines). This is true for dichotic condition 2 as well as for diotic condition 1. Results similar to the ones shown in Fig. 4 were obtained for stimulus conditions 3 and 4. This finding is important because in the diotic condition 1 no explicit parsing information was supplied to the subjects, and in conditions 3 and 4 an explicit attempt was actually made to supply them with false parsing information. If this wrong information had been used by the subjects, their performance would have been at chance level, which was easily shown by simulation. Actual performance was well above chance level for those conditions, however, as is evident from Figs. 1c,d. The empirical and simulated results tell us that subjects somehow do have a fairly accurate knowledge of the proper interpretation of the various stimulus partials in the percept of simultaneous complex tones, but that this knowledge is not obtained on the basis of ear input. It is probably obtained on the basis of experience with the harmonics of the stimulus set and a general tendency to group perceived partials holistically on the basis of common fundamental. Something similar was also found by Deutsch [14] and Butler [15] who used entirely different musical paradigms.

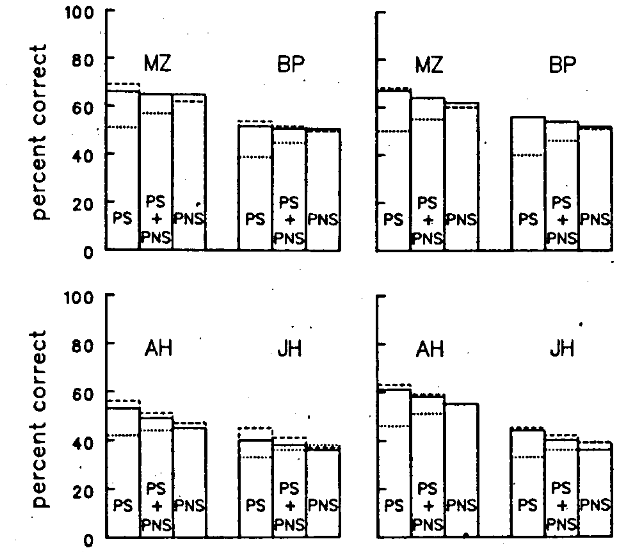

Fig. 4. Measured (solid lines) and simultated performance levels with (dashed lines) and without (dotted lines) knowledge of stimulus parsing. Columns PS are for the stimulus subset that is prone to parsing errors, columns PNS for the subset that does not induce such errors, the central columns for the entire set. Left: diotic condition 1; right: dichotic condition 2.

V. CONCLUSIONS

From the experimental and simulated results of this study the following conclusions are drawn:

1. The task of identifying two pitches when exposed to two simultaneous complex tones is separable into two pitch identification processes which are to a large extent independent. The small amount of mutual dependence that is sometimes observed tends to support the notion that the more salient pitch, represented by lower-order harmonics, is processed first and interferes with the processing of the less salient pitch. This mutual dependence of the two processes, small as it may seem, is largely responsible for the degradation of performance when going from dichotic condition 2 to diotic condition 1 and finally to dichotic conditions 3 and 4.
2. Information of the correct parsing and interpretation of perceived stimulus partials is, to a large extent, available to the central pitch processor. It is independent of the manner in which partials are distributed between the ears. This is consistent with other results on simultaneous-melody perception in the literature [14,15], and with informal observation of ordinary musical practice in which both ears are always exposed to all partials of simultaneously-playing musical instruments.
3. Interference of spectrally close partials has a surprisingly small effect on pitch identification for two complex tones, at least as long as either tone is represented by harmonics of sufficiently low order. Although it is known that high-order harmonics of a single complex tone do not contribute much to fundamental pitch sensation [9] and are as such not available to the central processor [11], it now appears that aurally non-resolved harmonics belonging to different tone complexes are not entirely discarded. They may instead be transformed into a single (noisy) percept that can be used more than once by the processor when filling in the variables of Eqs. (1) or (2). This idea will be investigated further in a future study.
4. The human central pitch processor appears not to be hardwired or specifically programmed for one particular way of processing stimulus tones. On the contrary, its processing algorithm appears to be quite *cooperative* and *interactive* with the task it has to execute.

VI. ACKNOWLEDGMENTS

This work, initiated under a grant from the U.S.A. National Institutes of Health (NS 11680) was executed with a grant from the Netherlands Foundation for the Advancement of Pure Research (ZWO 560-260-009).

VII. REFERENCES

[1] A. Seebeck, "Beobachtungen über einige Bedingungen der Entstehung von Tönen", Ann. Phys. Chem. 53, 417-436, 1841.

[2] G. Ohm, "Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen", Ann. Phys. Chem. 59, 513-565, 1843.

[3] H.L.F. von Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, F. Vieweg & Sohn, Braunschweig, 1863.

[4] H. Fletcher, "The physical criterion for determining the pitch of musical tone", Phys. Rev. 23, 427-437,1924.

[5] J.F. Schouten, "The residue and the mechanism of hearing", Proc. Ned. Akad. Wetenschap 43, 991-999, 1940.

[6] G. von Békésy, "The missing fundameental and periodicity detection in hearing", J. Acoust. Soc. Am. 51, 631-637, 1972.

[7] R. Plomp, "Pitch of complex tones", J. Acoust. Soc. Am. 41, 1526-1533, 1967.

[8] R.J. Ritsma, "Frequencies dominant in the perception of the pitch of complex sounds", J. Acoust. Soc. Am. 42, 191-198, 1967.

[9] A.J.M. Houtsma and J.L. Goldstein, "The central origin of the pitch of complex tones: evidence from musical interval recognition", J. Acoust. Soc. Am. 51, 520-529, 1972.

[10] E. Terhardt, "Zur Tonhöhewahrnehmung von Klängen II: Ein Funktionsschema", Acustica 26, 187-199, 1972.

[11] J.L. Goldstein, "An optimum processor theory for the central formation of the pitch of complex tones", J. Acoust. Soc. Am. 54, 1496-1516, 1973.

[12] A. Gerson and J.L. Goldstein, "Evidence for a general template in central optimal processing for pitch of complex tones", J. Acoust. Soc. Am. 63, 498-510, 1978.

[13] J.G. Beerends and A.J.M. Houtsma, "Pitch identification of simultaneous dichotic two-tone complexes", J. Acoust. Soc. Am. 80, 1048-1055, 1986.

[14] D. Deutsch, "Two-channel listening to musical scales", J. Acoust. Soc. Am. 57, 1156-1160, 1975.

[15] D. Butler, "A further study of melodic channeling", Percept. Psychophys. 25, 264-268, 1979.