This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the Department of the Air Force under Contract F19628-90-C-0002.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Hugh L. Southall

Hugh L. Southall, Lt. Col., USAF
Chief, ESC Lincoln Laboratory Project Office

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

MODELS OF PITCH PERCEPTION

T.C. AARSET
Group 41

B. GOLD
Group 21

TECHNICAL REPORT 964

28 AUGUST 1992

LEXINGTON MASSACHUSETTS

ABSTRACT

Two pitch perception modeling algorithms are described. The first algorithm models "periodicity" pitch perception, and the second algorithm models "place" pitch perception.

The two models are now applied to various psychoacoustic stimuli. Both periodicity and place models yield results that are in general agreement with psychoacoustic measurements for the missing fundamental and for inharmonic stimuli. The place algorithm proved to be a better approximation than periodicity for processing comb-filtered noise. Periodicity was more successful for periodic pulse train stimuli.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1. REVIEW OF SOME ASPECTS OF PSYCHOACOUSTICS

1.1 THE MISSING FUNDAMENTAL

When a listener matches a pure tone (i.e., sine wave) to a complex tone consisting of a set of harmonically related sum of sinusoids (e.g., harmonics 3, 4, and 5 of some fundamental frequency), the match will take place at f, which is the *missing* fundamental frequency of the complex tone.

1.2 PITCH OF INHARMONIC SIGNALS

When a listener is asked to match the sum of sinusoids at frequencies $3f$, $4f$, and $5f$ to a pure tone, that match will occur at f. What happens if each of the above three sinusoids is shifted in frequency by Δf? DeBoer [1] performed such an experiment, as did van den Brink [2] and Smoorenburg [3]. This research showed that subjects were still able to perceive pitch despite the inharmonicity of the signal.

1.3 PITCH OF REPEATED NOISE

Miller and Taylor [4] discovered that pitch could be perceived when the listener was presented with repeated bursts of noise. More current research found if white noise was comb-filtered (i.e., the output is the sum of the input and a delayed version of the input), a pitch of $1/T$ was perceived [5–8]. The stimulus "sounds like" noise; however, when T is systematically varied such that $1/T$ steps through the frequencies corresponding to the seven notes of the major scale, the pitch of these notes is heard [9].

1.4 PITCH PERCEPTION OF PULSE TRAINS

Flanagan and Guttman [10] discovered two distinct modes of pitch perception for periodic pulse train stimuli. To quote their paper, "In the first mode, for pulse rates less than 100 pps, the pulse trains are ascribed a pitch equal to the number of pulses per second, regardless of the polarity pattern of the pulses. In the second mode, for fundamental frequencies in excess of 200 Hz, the sounds are assigned a pitch equal to the fundamental frequency."

1.5 CIRCULARITY IN JUDGMENTS OF RELATIVE PITCH

Shepard [11] demonstrated that for specialized signals consisting of the sum of tones separated by octaves (e.g., 150, 300, 600, or 1200 Hz, etc.) listeners will often identify an *increase* in the tone frequencies as a *lowering* of pitch. On the average, if all tone frequencies are increased by less than one-half octave, the new stimulus is judged to be higher in pitch than the old one. If, however, all tone frequencies are increased by more than one-half octave (but less than one octave), the new stimulus is judged to be lower in pitch. Pollack [12] viewed this result as a further example of the decoupling of auditory pitch and stimulus frequency. Through a series of experiments he identifies important parameters of this phenomenon as "the number of components in the signal, the number of offsetting frequencies which are weighted against the direction of Shepard pitch and perhaps the spacing between the components." Deutsch [13] has discovered other interesting properties of "Shepard pitch." For example, a tone pattern can be heard as ascending when played in one key and descending when played in another.

2. DESCRIPTION OF PERIODICITY AND PLACE PROGRAMS TO MODEL PITCH PERCEPTION

2.1 PERIODICITY PITCH

Figure 1 shows block diagrams of both periodicity and place algorithms for pitch detection. The periodicity algorithm assumes a correspondence between the basilar membrane and the filter bank and a further correspondence between the hair cell-auditory nerve complex (on the one hand) and elementary pitch detectors EPD_1 through EPD_M. The filter designs are based on physiological measurements of Delgutte [14]. At present, 19 filters have been implemented; the frequency covers a 2-kHz range. The ability of these filters to resolve harmonics is a function of the pitch and formant structure of the incoming signal; thus, a given filter, representing a specific place on the basilar membrane, can sometimes perform a place function and, at other times, the *same* filter can perform a periodicity function.

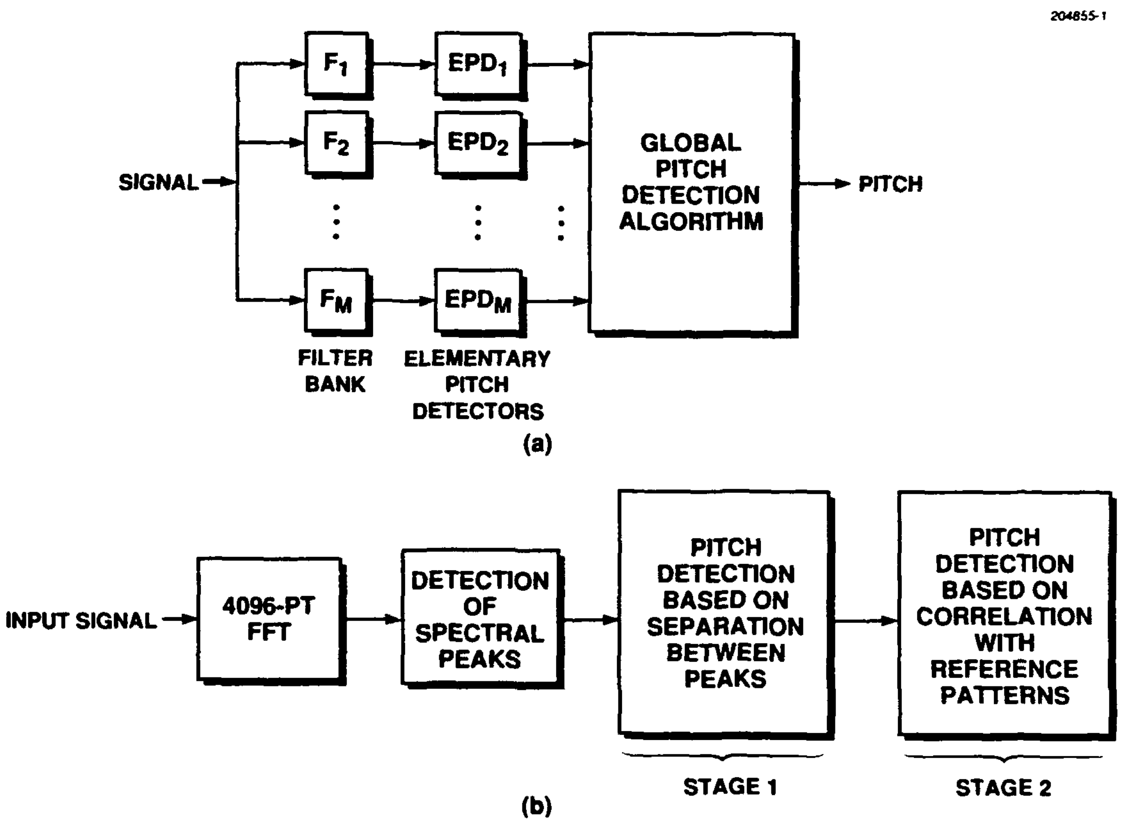

Figure 1. (a) Block diagram for the periodicity algorithm and (b) block diagram of the two-stage place algorithm.

Neural spiking tends to follow the peaks of the signal. Given an auditory nerve spike, that same nerve cannot respond to further stimulation for a time immediately following the spike; this is called the refractory period. Following this period, the potential differences inside the neuron gradually return to normal, thus steadily increasing the probability of subsequent firings.

The global algorithm shown in Figure 1(a) further processes the succession of time intervals between spike occurrences. The algorithm is based on the hypothesis that the higher auditory centers can interpret intervals separated by several spikes to produce five monotonically increasing intervals. Thus, at any instant, each elementary pitch detector (EPD) presents five numbers to the global detector. Each filter output excites two EPDs; the positive set of EPDs spikes on positive excursions of the signals, while the negative set spikes on negative excursions. Thus, there are 2 x 19 = 38 EPDs, each producing five intervals so that there are 5 x 38 = 190 intervals available. The final global periodicity decision is obtained by developing a histogram of these 190 intervals and choosing the mode or maximum of the resultant probability density function.

2.2 PLACE PITCH

The underlying hypothesis of place detection is the ability of the auditory system to resolve enough harmonic peaks of the stimulus. This resolution can take place at the periphery or at higher levels. In fact, Houtsma and Goldstein [15] have shown that centrally located auditory processes can indeed perceive pitch even if each ear is subjected to a single harmonic of the fundamental frequency. Figure 1(b) shows an implementation of a high-resolution spectral analysis via a 4096-point fast Fourier transform.

A reliable algorithm that leads to the Goldstein algorithm [16] can be implemented as a two-stage process. Stage 1 is a version of the Seneff algorithm [17] that performs a statistical analysis of the frequency separation between spectral peaks. Stage 2 is related to the "harmonic sieve" algorithm of Goldstein's pitch perception model, as implemented by Duifhuis [18]. The spectral peaks are correlated with sets of harmonically spaced narrow windows. The chosen sets are based on the "winning pitch" of stage 1. If the winning pitch of stage 1 is f, the chosen sets include $f \pm 25$ Hz. Thus, the results of stage 1 greatly restrict the range of measurements of stage 2. The combination avoids many ambiguous pitch results.

3. PRELIMINARY MODEL RESULTS FOR PSYCHOACOUSTIC STIMULI

3.1 PITCH OF INHARMONIC SIGNALS (SHIFT OF VIRTUAL PITCH)

Figure 2 shows several cases for both place and periodicity model responses to inharmonic stimuli. Since people perceive pitch despite the absence of harmonic structure at the pitch frequency, the term "virtual pitch" is used to describe the results. Figures 2(a) and (c) show the model results for different harmonic structures. Figure 2(b) shows a comparison between the human response (dashed line) and the two models. It appears that *both* place and periodicity models respond similarly. Alsc, both models respond qualitatively in the same manner as the human. Because of our 10-kHz sampling rate, the model results are quantized.

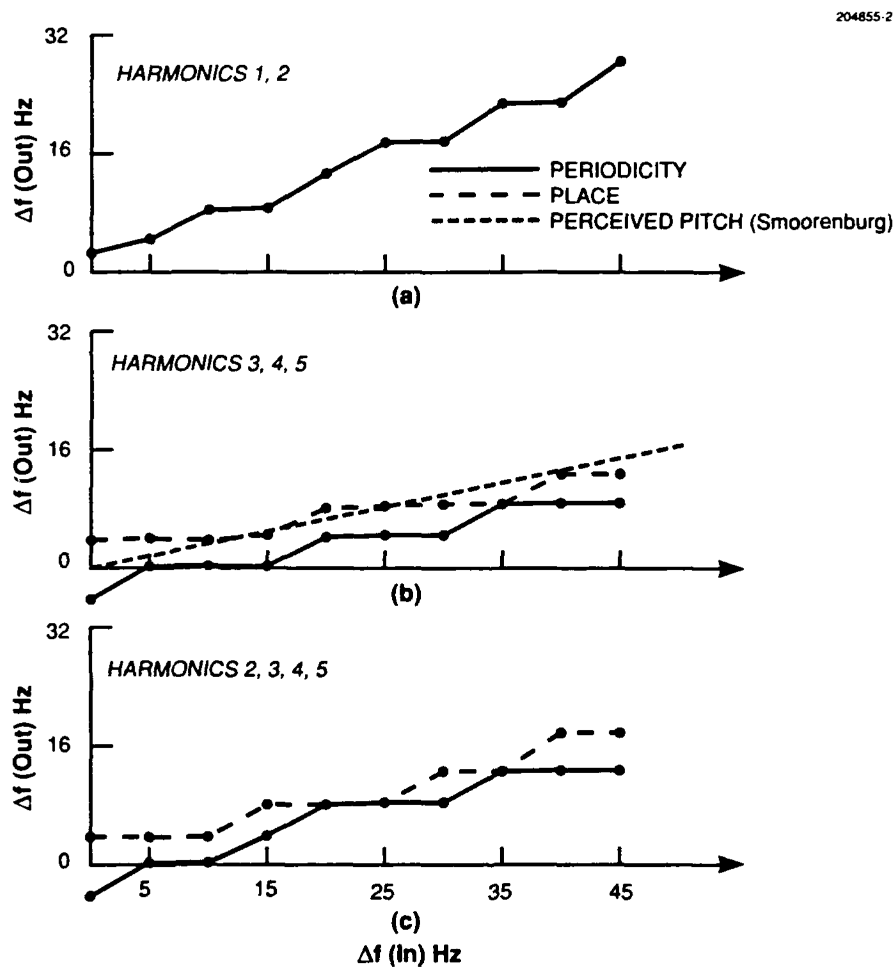

Figure 2. Shift of Virtual Pitch. (a) Periodicity pitch track for a stimulus consisting of harmonics 1 and 2, (b) three pitch curves for a stimulus consisting of harmonics 3, 4, and 5, and (c) a stimulus consisting of harmonics 2 through 5 inclusive.

3.2 PITCH OF COMB-FILTERED NOISE

Figure 3 shows how the models respond to comb-filtered white noise. Ten delays were imposed; they ranged from 12.0 to 2.1 ms. Each of the comb-filtered signals is processed by both models for 160 ms and then followed by a pause of 55 ms. Figure 3 shows the pitch period; the dips in period are due to the "off" parts of the signal. The place algorithm follows the results obtained from psychoacoustics for the initial six of the ten cases. Periodicity also tends to follow this pattern but much less reliably. Interestingly, however, the higher pitches are better represented by the periodicity models. A speculative hypothesis could attribute lower pitch results to a place model and higher pitch results to a periodicity model.

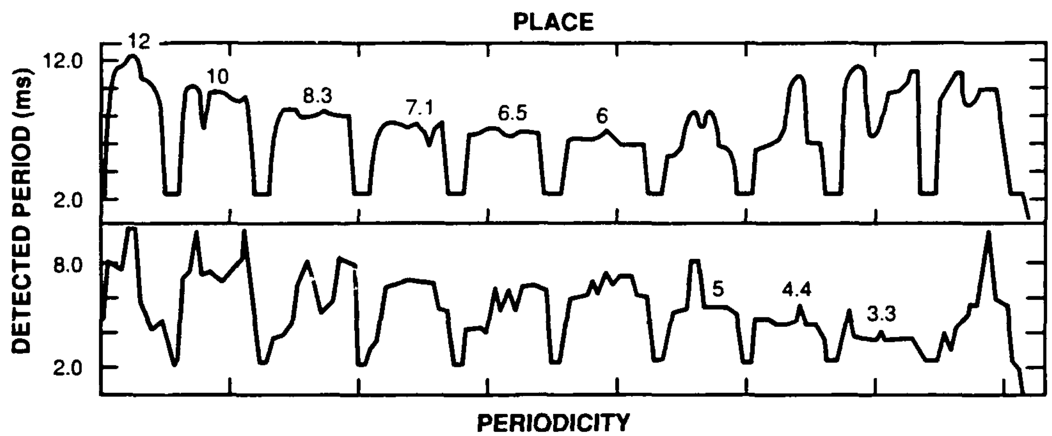

Figure 3. Pitch of comb-filtered noise, which shows the tracks of the detected periods from both place and periodicity pitch detection algorithms.

3.3 TRANSITION FROM RATE TO FUNDAMENTAL FREQUENCY

Figure 4 shows the results of the periodicity pitch model. A single period of the repetitive stimulus is shown in the box at the upper left. Every 0.18 s, the pulse rate r is increased in steps. The fundamental frequency f is always $r/4$. Periodicity pitch follows the rate until r = 300 pulses/s and then abruptly switches to follow the fundamental. This is quite analogous to the behavior of human listeners to the same stimulus. For this stimulus, the periodicity model appears to work properly, but the place model does not.

3.4 CIRCULARITY IN PITCH PERCEPTION

Figure 5 shows the pitch periods generated by the periodicity model for various fundamental frequencies of the Shepard stimuli. Figure 5 shows that this result is consistent (for the periodicity model) over an octave range of lowest tones. The place model, on the other hand, appears to yield completely ambiguous results.

Jean-Claude Risset experimented with a complex Shepard pitch signal consisting of ten components, each of which descends ten octaves but are perceived together as an endless glissando or pitch slide that remains within a single octave register Risset [19–21]. Figure 6 shows the responses of both periodicity and place models to Risset's "Endless Glissando." Both models remain within the single octave during the cycle with the exception of an ambiguous region for the periodicity model near the half octave point. The place algorithm is not ambiguous and seems to track the physical stimulus well. This is in contrast to its response to Shepard tones. A possible reason for this apparent discrepancy is that in our version of the Shepard tones all harmonics were equal, while our version of the Risset stimulus included his amplitude window. Future experiments may shed further light on this issue.

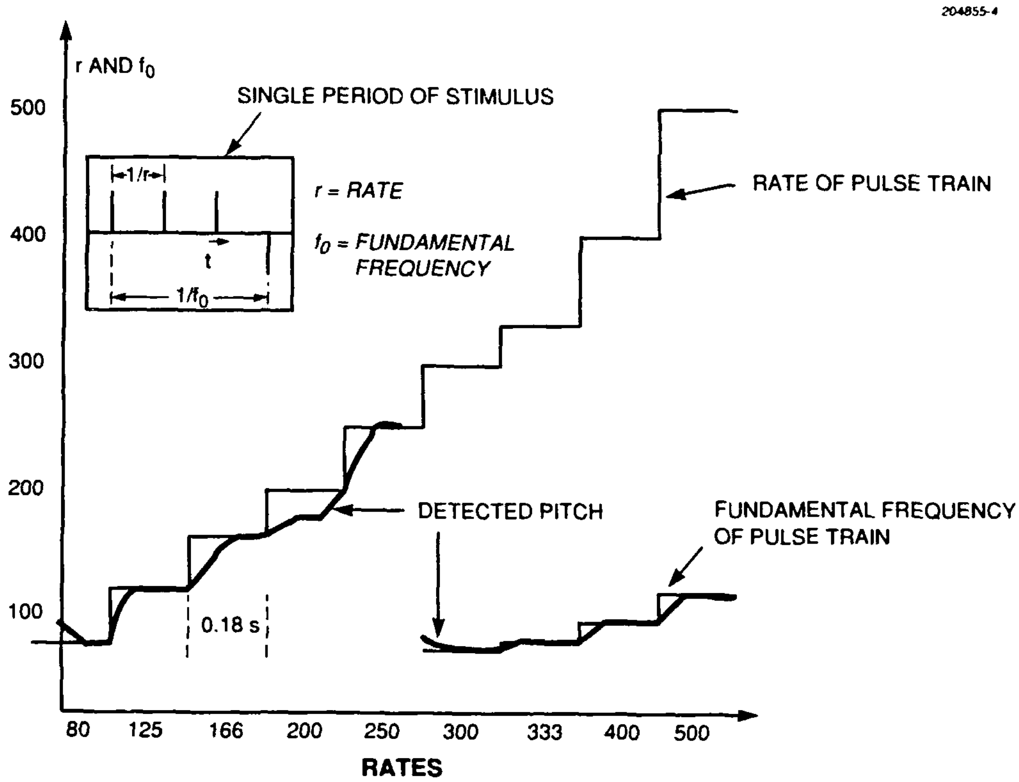

Figure 4. Rate versus frequency plot from the periodicity pitch algorithm for the Flanagan-Guttman pulse stimulus.

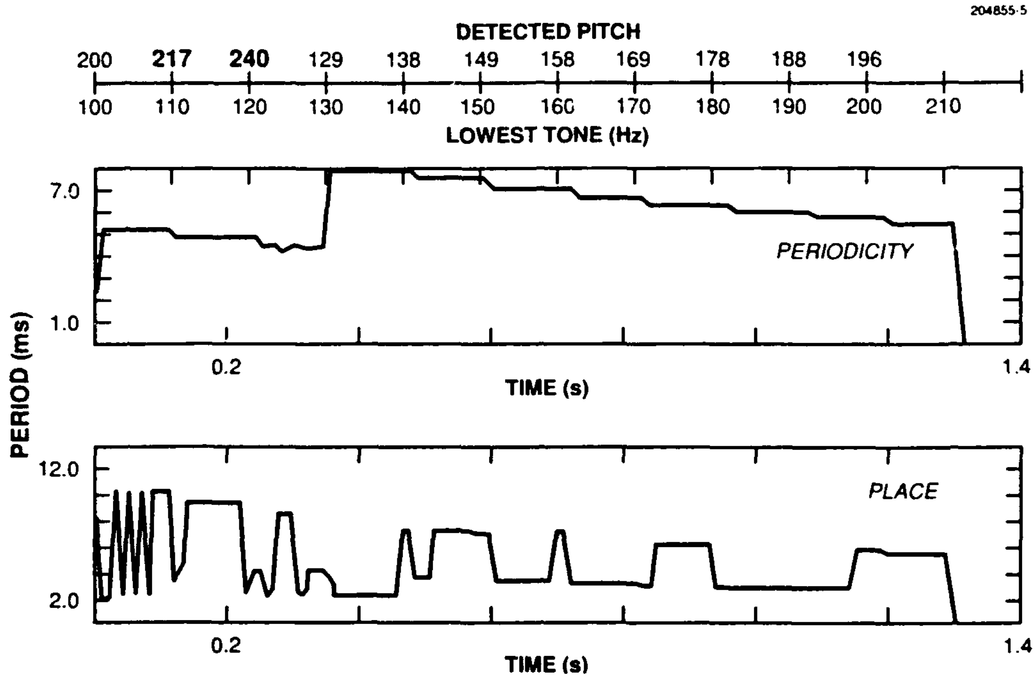

Figure 5. Comparison of periodicity and place algorithms for pitch circularity using a Shepard tone complex. The one-half octave shift (circularity) described in the psychoacoustics literature is contrasted to the arbitrary response from the place algorithm to the same stimuli.

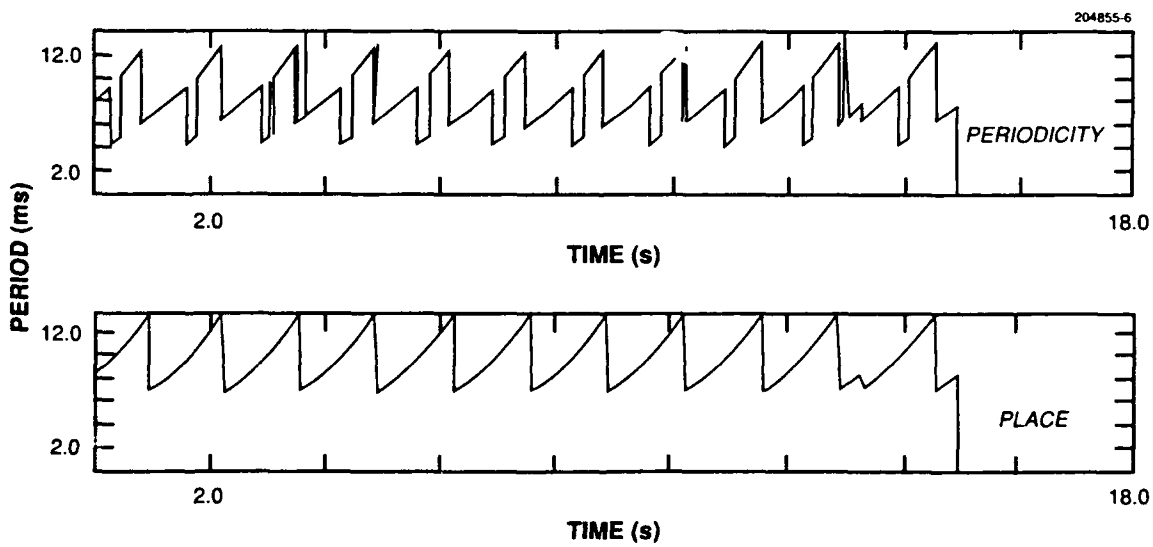

Figure 6. Periodicity and place model responses to Risset's "Endless Glissando."

4. SUMMARY AND CONCLUSIONS

Two models of pitch perception have been implemented, and the response of these models to various psychoacoustic stimuli have undergone preliminary study. Both models successfully track the pitch of a harmonic signal with missing fundamentals. The periodicity model corresponds to psychoacoustic results from human listeners for inharmonic stimuli, periodic pulse train stimuli, and Shepard stimuli. On the other hand, the place model corresponds to psychoacoustic results for inharmonic stimuli, comb-filtered noise, and nonsimultaneous harmonics. These results can help psychophysicists speculate on auditory nerve functions above the periphery, including a possible mechanism that might combine the optimal performance of both models.

REFERENCES

1. E. DeBoer, "On the 'residue' and auditory pitch perception," in W.D. Keidel and W.D. Neff (eds.), *Handbook of Sensory Physiology, Vol. V, Auditory System, Part 3: Clinical and Special Topics*, Berlin, Germany: Springer (1976), 479-583.

2. G. van den Brink, "Two experiments on pitch perception: diplacusis of harmonic AM signals and pitch of inharmonic AM signals," *J. Acoust. Soc. Am.* 48, 1355–1365 (1970).

3. G.F. Smoorenburg, "Pitch perception of two-frequency stimuli," *J. Acoust. Soc. Am.* 48, 924–942 (1970).

4. G.A. Miller and W.G. Taylor "The perception of repeated bursts of noise," *J. Acoust. Soc. Am.* 20, 171–180 (1948).

5. F.A. Bilsen and R.J. Ritsma, "Repetition pitch and its implication for hearing theory," *Acustica* 22, 63–73 (1969).

6. R.M. Warren, J.A. Bashford, Jr., and J.M. Wrightson, "Infrapitch echo," *J. Acoust. Soc. Am.* 68, 1301–1305 (1980).

7. W.A. Yost, "The dominance region and ripple noise pitch: A test of the peripheral weighting model," *J. Acoust. Soc. Am.* 72, 416—425 (1982).

8. R.M. Warren and J.A. Bashford, Jr., "Broadband repetition pitch: Spectral dominance or pitch averaging?" *J. Acoust. Soc. Am.* 84, 2058—2062 (1988).

9. A.J.M. Houtsma, T.D. Rossing, and W.H. Wagenaars, *Audio Demonstrations*, compact disk brochure, Institute for Perception Research (1987).

10. J.L. Flanagan and N. Guttman, "On the pitch of periodic pulses," *J. Acoust. Soc. Am.* 32, 1308–1319 (1960).

11. R.N. Shepard, "Circularity in judgments of relative pitch," *J. Acoust. Soc. Am.* 36, 2346–2353 (1964).

12. I. Pollack "Decoupling of auditory pitch and stimulus frequency: The Shepard demonstration revisited," *J. Acoust. Soc. Am.* 63, 202–206 (1978).

13. D. Deutsch, "A musical paradox," *Music Perception* 3, 275—280, (1986).

14. B. Delgutte (private communication, 1990).

15. 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).

16. 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).

17. S. Seneff, "Real-time harmonic pitch detector," *IEEE Trans. Acoust. Speech Signal Process.* ASSP-26, 358–365 (1978).

18. H. Duifhuis, L.F. Willems, and R.J. Sluyter, "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception," *J. Acoust. Soc. Am.* 71, 1568—1580 (1982).

19. J.C. Risset (Bell Labs, private communication, 1969).

20. C. Dodge and T.A. Jerse, "Synthesis fundamentals," *Computer Music: Synthesis, Composition, and Performance*, New York: Schirmer Books (1985), 95–97.

21. M.V. Mathews and J.R. Pierce (eds.), "Paradoxical sounds," *Current Directions in Computer Music Research*, Cambridge, Massachusetts: MIT Press (1989), 149–158.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	28 August 1992	Technical Report

4. TITLE AND SUBTITLE

Models of Pitch Perception

5. FUNDING NUMBERS

C — F19628-90-C-0002

6. AUTHOR(S)

Tim Aarset and Ben Gold

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Lincoln Laboratory, MIT
P.O. Box 73
Lexington, MA 02173-9108

8. PERFORMING ORGANIZATION REPORT NUMBER

TR-964

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

HQ Air Force Material Command
AFMC/STSC
Wright-Patterson AFB, OH 45433-5001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ESC-TR-92-119

11. SUPPLEMENTARY NOTES

None

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

Two pitch perception modeling algorithms are described. The first algorithm models "periodicity" pitch perception, and the second algorithm models "place" pitch perception.

The two models are now applied to various psychoacoustic stimuli. Both periodicity and place models yield results that are in general agreement with psychoacoustic measurements for the missing fundamental and for inharmonic stimuli. The place algorithm proved to be a better approximation than periodicity for processing comb-filtered noise. Periodicity was more successful for periodic pulse train stimuli.

14. SUBJECT TERMS

pitch	auditory physiology	computer modeling and simulation
pitch detection	acoustics, hearing	signal processing algorithms
pitch perception	psychoacoustics	acoustic/digital signal processing

15. NUMBER OF PAGES

14

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

END

FILMED

DATE:

11-92

DTIC