UNCLASSIFIED **DTIC FILE COPY**

SECURITY CLASSIFICATION OF THIS PAGE

(2)

AD-A200 530

---- OCUMENTATION PAGE

Form Approved OMB No. 0704-0188

UNCLASSIFIED

1b. RESTRICTIVE MARKINGS
None

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE
-----

3. DISTRIBUTION / AVAILABILITY OF REPORT
Unrestricted

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
-----

5. MONITORING ORGANIZATION REPORT NUMBER(S)
----- AFOSR-TR- 88-1160

6a. NAME OF PERFORMING ORGANIZATION
Richard M. Warren
Univ. of Wisconsin-Milwaukee

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
AFOSR/PKD

6c. ADDRESS *(City, State, and ZIP Code)*
Department of Psychology
Milwaukee, WI 53201

7b. ADDRESS *(City, State, and ZIP Code)*
Building 410
Bolling Air Force Base, DC 20332-6448

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
AFOSR/NL

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR-85-0260A

8c. ADDRESS *(City, State, and ZIP Code)*
Building 410
Bolling Air Force Base, DC 20332-6448

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102F	2313	A6	

11. TITLE *(Include Security Classification)*
Mechanisms Mediating the Perception of Complex Acoustic Patterns (UNCLASSIFIED)

12. PERSONAL AUTHOR(S)
Warren, Richard M.

13a. TYPE OF REPORT
Final Report

13b. TIME COVERED
FROM 8/1/85 TO 7/30/88

14. DATE OF REPORT *(Year, Month, Day)*
88/09/26

15. PAGE COUNT
9 + Appendices

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
auditory perception, complex sounds, pitch

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The completed research program has dealt with the mechanisms and principles governing the perception of complex sounds. The main topics studied were: (1) Pitch averaging mechanisms for repetition pitch; (2) spectral and temporal mechanisms underlying some novel perceptual effects observed with complex tones mistuned from unison; (3) a comparison of tonal and infratonal auditory induction and their underlying mechanisms; (4) monaural ear advantages for infratonal periodicity detection and its implications for subcortical periodicity processing; and (5) evidence that perception of infratonal periodicity does not depend solely upon the recognition of the repetition of singularities, but also involves a holistic pattern recognition.

DTIC
S ELECTE
OCT 3 1 1988
E

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
John F. Tangney, Program Manager

22b. TELEPHONE *(Include Area Code)*
(202) 767-4278

22c. OFFICE SYMBOL
AFOSR/NL

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

88 10 28 02

0 3 OCT 1988

Report AFOSR-85-0260A

MECHANISMS MEDIATING THE PERCEPTION OF COMPLEX ACOUSTIC PATTERNS

Richard M. Warren
University of Wisconsin-Milwaukee
Department of Psychology
Milwaukee, Wisconsin 53201

31 August 1988

Final Report for Period 1 August 1985 - 30 July 1988

Prepared for
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Building 410
Bolling Air Force Base, DC 20332-6448

AFOSR-TR- 88-1160

Richard M. Warren
AFOSR Grant No. 85-0260A

Final Report AFOSR

SUMMARY

The completed research program has dealt with the mechanisms and principles governing the perception of complex sounds. The main topics studied were: (1) Pitch averaging mechanisms for repetition pitch; (2) spectral and temporal mechanisms underlying some novel perceptual effects observed with complex tones mistuned from unison; (3) a comparison of tonal and infratonal auditory induction and their underlying mechanisms; (4) monaural ear advantages for infratonal periodicity detection and its implications for subcortical periodicity processing; and (5) evidence that perception of infratonal periodicity does not depend solely upon the recognition of the repetition of singularities, but also involves a holistic pattern recognition.

STATEMENT OF WORK

The goal of this research program supported by AFOSR is to further our knowledge of mechanisms and principles governing the perception of complex sounds. The account which follows describes the essential aspects of work completed, with a full account being furnished by the papers included as appendices.

1. A study of repetition pitch or RP (see Appendix A, "Broadband repetition pitch: Spectral dominance or pitch averaging?" in press, *J. Acoust. Soc. Am.*) has examined RPs produced under a variety of delays and filtering conditions. Both normal or cophasic mixtures (RP+) and polarity inverted or antiphasic mixtures (RP-) were used. In keeping with earlier reports, RP+ having a delay of t seconds produced a pitch of 1/t Hz for all spectral regions examined. The conditions of interest to theory involved RP-. While broadband RP- diverged from 1/t Hz in keeping with the literature, the pitches heard under novel filtering conditions indicated that (contrary to some current theories) RP- is a weighted average of the different pitches contributed by different spectral regions. Polarity inversion of an echo introduces additional frequency-dependent delays, and it was suggested that the corresponding RP- at local regions of the basilar membrane reflect a temporal domain analysis (probably autocorrelational) based on the sum of these two types of delays.

2. The perceptual effects of complex tones mistuned from unison was examined, and some previously unreported effects were found (see Appendix B, "Perception of complex tone pairs mistuned from unison," submitted to *J. Acoust. Soc. Am.*). The complex tones employed had frequencies varying from 20 Hz through 400 Hz and contained all harmonics of the fundamental up to 8 kHz. All tone pairs had one member with a randomly determined phase spectrum, while the other member of the pair was either "correlated" (identical except for a slight difference in fundamental frequency) or "uncorrelated" (having a slight difference in fundamental frequency and independent, randomly-determined phases for corresponding harmonics). For the correlated tone pairs mistuned from unison, it was found that:

   (a) Glissandi could be heard with ease only when the fundamental frequencies of the tone pairs were roughly 400 Hz or less;

(b) There was no lower frequency limit for tone pairs producing glissandi, and pitch glides were also heard clearly for mistuned complex waveforms having infratonal repetition frequencies (that is, repetition frequencies below the 20 Hz tonal limit);

(c) Glissandi were not heard if the extent of mistuning exceeded a critical value. This limit for perception of pitch glides was a discontinuous function of the fundamental frequencies of the complex tones, with one function applying below 50 Hz and another above 50 Hz;

(d) Glissandi required the presence of higher but not the lower harmonics of the complex tones. When tone pairs producing glissandi were high-pass filtered above the seventh harmonic, perception of pitch glides was unimpaired. However, glissandi were not heard when the same tone pairs were low-pass filtered below the eighth harmonic (complex patterns of amplitude fluctuation were heard instead).

Rather different results were obtained when uncorrelated complex tone pairs were mistuned from unison:

(a) Pitch glides could not be perceived, but complex patterns of amplitude fluctuation were heard to repeat under the conditions (a) through (d) described above for correlated complex tones. Removal of the spectral fundamentals of the tones did not change either the clarity or ensemble repetition rate heard for the complex patterns produced by uncorrelated complex tones;

(b) The beating of the individual harmonic components of the complex tones was difficult to hear when their fundamental frequencies differed by 1 Hz or more. But when the uncorrelated tones were mistuned by less than about 0.5 Hz, the integrated or ensemble periodicity became less salient and the individual beat rates were clearly dominant. Appendix B describes these observations in more detail, along with others, and provides information concerning the nature of frequency domain and time domain mechanisms employed for the perception of iterated acoustic patterns.

3. Tonal and infratonal auditory induction were compared (see Appendix C, "Illusory continuity of tonal and infratonal periodic sounds," in press, *J. Acoust. Soc. Am.*). Auditory induction can produce illusory continuity of a tone alternated with a brief louder sound if the louder sound is capable of masking the tonal signal. This perceptual synthesis of obliterated tonal segments has been widely studied ("pulsation threshold" experiments), and the limiting conditions for this illusory continuity have been used to define characteristics of neural spectral (place) analysis of tones. The present study extends investigation of the apparent continuity of interrupted complex periodic sounds to infratonal frequencies, exploring the range of repetition frequencies from 2000 down to 10 Hz. The quantitative results obtained have suggested that the perceptual synthesis of missing segments involves both time domain and frequency domain mechanisms.

4. Auditory induction can also operate with complex, nonrepetitive signals such as speech, and we have examined verbal forms of illusory continuity in order to identify both specialized and general characteristics of perceptual restoration. A recent study examined effects of signal rate upon verbal induction (see Appendix D, "Illusory continuity of interrupted speech: Speech rate determines durational limits," in press, *J. Acoust. Soc. Am*). Using a recorded discourse passage periodically interrupted by noise, it was found that changes in speech playback rate produced proportional changes in the durational limits for apparent continuity, with the threshold gap duration at each playback rate approximating the average word duration in the recorded passage. These results indicate that the upper durational limit for perceptual synthesis of running speech corresponds to a fixed amount of information, and not a fixed temporal value. New work to be conduction under my renewal grant (AFOSR-88-0320) will employ periodic nonverbal signals ("word-length" segments of gaussian noise recycled at infratonal rates) to help determine whether an informational limit, rather than a fixed temporal limit, also applies to the perceptual restoration of time-varying patterns other than speech.

5. A study was completed dealing with ear advantages for monaural periodicity detection (Appendix E, "Ear advantages for monaural periodicity detection"). It was found that some listeners show a striking ear difference in the clarity of infratonal repetition heard when noise segments are repeated in what Guttman and Julesz have called the "whooshing" range (repetition frequency of 1 through 4 Hz) and the "motorboating" range (repetition frequency of 4 through 19 Hz). In Experiment 1, an overall left ear advantage was found for repeated noise delivered monaurally and opposed by contralateral silence. In Experiment 2, lateralization of the repeated monaural signal was abolished by simultaneous presentation of a louder on-line noise to the opposite ear (contralateral induction caused the monaural signal to be perceived as centered on the medial plane). Although this illusory centering of the periodic sound eliminated the possible influence of attentional biases favoring one of the sides, ear advantages were still obtained. These results suggest the possibility of asymmetry in active subcortical processing of periodicity information.

6. A study has been completed dealing with the mechanisms employed for the detection of infratonal repetition of complex waveforms. This study has not yet been written up (although a paper reporting this work was presented at an Acoustical Society of America meeting [B. S. Brubaker & R. M. Warren, *J. Acoust. Soc. Am.*, 1987, *82*, S93 (Abstract)]. The results address the question as to whether repetition detection for long-period complex waveforms is based upon the recognition of recurrence of unique features or singularities, or upon the more holistic recognition of the pattern formed by these events. In this study, "frozen" noise segments were divided into three sections of equal duration (A, B, C) which were reassembled and then repeated to form the periodic sounds $(ABC)_n$ and $(ACB)_n$. This manipulation changed the temporal arrangement between segments but preserved singularities and repetition rate. Untrained listeners heard a series of sequence bursts consisting of either one arrangement $[(ABC)_n, (ABC)_n, (ABC)_n \ldots]$ or two alternating arrangements $[(ACB)_n, (ABC)_n, (ACB)_n \ldots]$, and judged whether successive bursts were the same or different. Discrimination was possible when the duration of the entire iterated pattern (A+B+C) was 900 ms or less, indicating that a holistic recognition of patterns operates up to the limit of echoic storage.

7. A book chapter attempting to explain how speech evolved from prelinguistic auditory mechanisms shared with other animals was written [see Appendix F, "Perceptual bases for the evolution of speech," in The Genesis of Language, M. Landsberg (Ed.), Berlin: de Gruyter (in press)]. It was suggested that speech perception does not require the ability to identify individual sounds, but rather is based upon a holistic recognition of complex acoustic patterns of a sort studied under the current AFOSR grant. It appears that much confusion in the literature resulted from considering that speech perception requires the ability to recognize phonemes and their orders at some level of perceptual organization. There is experimental evidence that our ability to recognize acoustic patterns holistically is shared with other animals, and that speech perception evolved from this prelinguistic ability. Hence, the identification of component sounds and their orders may be a linguistic skill which is the consequence of, not the basis of, speech recognition.

PAPERS PUBLISHED OR IN PRESS

1. Bashford, J. A., Jr., & Warren, R. M. "Multiple phonemic restorations follow the rules for auditory induction." Perception & Psychophysics, 1987, 42, 114-121.

2. Warren, R. M., & Meyers, M. D. "Effects of listening to repeated syllables: Category boundary shifts versus verbal transformations." Journal of Phonetics, 1987, 15, 169-181.

3. Warren, R. M., & Bashford, J. A., Jr. "Broadband repetition pitch: Spectral dominance or pitch averaging?" J. Acoust. Soc. Am., (in press).

4. Warren, R. M., Wrightson, J. M., & Puretz, J. "Illusory continuity of tonal and infratonal sounds," J. Acoust. Soc. Am., (in press).

5. Warren, R. M. "Perceptual bases for the perception of speech," in The Genesis of Language, M. Landsberg (Ed.), Berlin: de Gruyter, (in press).

6. Bashford, J. A., Jr., Meyers, M. D., Brubaker, B. S., & Warren, R. M. "Illusory continuity of interrupted speech: Speech rate determines durational limit," J. Acoust. Soc.Am., (in press).

PROFESSIONAL PERSONNEL

In addition to R. M. Warren (Principal Investigator), James A. Bashford, Jr., Ph.D., participated in the project in the capacity of Associate Researcher. Graduate students who assisted in this research project were: Bradley S. Brubaker, Daniel A. Gardner, Robin L. Means, Mark D. Meyers, Keri R. Riener, Jill Robertson, and Daniel G. Zuck.

PROFESSIONAL INTERACTIONS

Presentation on pattern directed theory of auditory perception at the ONR workshop on "Auditory Perception" at NAS-NRC Joseph Henry Building, Washington, January 13-14, 1986. The workshop was organized by Dr. John J. O'Hare of ONR and hosted by Dr. Milton A. Whitcomb, Study Director for CHABA, NAS-NRC. The two other resource people at the Workshop were Dr. James A. Simmons and Dr. David M. Green.

Invited speaker at the Institute of Experimental Psychology, University of Sussex, Brighton, England, May, 1986. The topic was, "Perception of Temporal Order in Auditory Sequences."

Invited speaker at a special perception session of the Audio Engineering Society, Los Angeles, November, 1986. Paper topic "Acoustic Sequences: Explaining a Perceptual Paradox," (published abstract: Warren, R. M., "Acoustic sequences: Explaining a perceptual paradox." Journal of the Audio Engineering Society, 1986, 34, 1021). Other invited speakers at the session were: Edward C. Carterette, John R. Pierce, W. Dixon Ward, Floyd E. Toole, and Diana Deutsch.

Member of panel (chaired by William A. Yost) preparing report on "Classification of Complex Sounds" for the Committee on Hearing, Bioacoustics and Biomechanics (CHABA) of the National Research Council.

APPENDICES

A. Warren, R. M., & Bashford, J. A., Jr. "Broadband repetition pitch: Spectral dominance or pitch averaging?" *J. Acoust. Soc. Am.*, (in press).

B. Warren, R. M., Bashford, J. A., Jr., & Brubaker, B. S. "Perception of complex tones mistuned from unison," (submitted to *J. Acoust. Soc. Am.*).

C. Warren, R. M., Wrightson, J. M., & Puretz, J. "Illusory continuity of tonal and infratonal sounds," *J. Acoust. Soc. Am.*, (in press).

D. Bashford, J. A., Jr., Meyers, M. D., Brubaker, B. S., & Warren, R. M. "Illusory continuity of interrupted speech: Speech rate determines durational limit," *J. Acoust. Soc.Am.*, (in press).

E. Warren, R. M., & Bashford, J. A., Jr. "Ear advantages for monaural periodicity detection," (article draft).

F. Warren, R. M. "Perceptual bases for the perception of speech," in *The Genesis of Language*, M. Landsberg (Ed.), Berlin: de Gruyter, (in press).

APPENDIX A

AFOSR TR- 92-1160

Broadband repetition pitch:
Spectral dominance or pitch averaging?

by

Richard M. Warren
James A. Bashford, Jr.

University of Wisconsin-Milwaukee
Department of Psychology
Post Office Box 413
Milwaukee, Wisconsin 53201

Received:

ABSTRACT: Repetition pitch (RP) produced by mixing noise with its restatement was studied under a variety of delays and filtering conditions. Both normal or cophasic mixtures (RP+) and polarity inverted or antiphasic mixtures (RP-) were used. In keeping with earlier reports, RP+ having a delay of t seconds produced a pitch of 1/t Hz for all spectral regions examined. Broadband RP- diverged from 1/t Hz in keeping with the literature, but the pitches heard under novel filtering conditions indicated that (contrary to some current theories) RP- is a weighted average of the different pitches contributed by different spectral regions. Polarity inversion of an echo introduces additional frequency-dependent delays, and it is suggested that the corresponding RP- at local regions of the basilar membrane reflects a temporal domain analysis based on the sum of these two types of delays.

I. Introduction

When a broadband noise is added to itself following a delay of t seconds, a pitch corresponding to 1/t Hz can be heard for delays ranging from approximately 0.0005 s (corresponding to 2000 Hz) through 0.02 s (corresponding to 50 Hz) (Fourcin, 1965; Bilsen, 1966; Wilson, 1966). This "repetition pitch" or RP (Bilsen, 1966) has some interesting special characteristics.

When noise is mixed with its echo, the resulting rippled power spectrum has its first spectral peak at 1/t Hz and a harmonic succession of peaks at integral multiples of 1/t Hz. It seems reasonable to attribute the pitch of rippled noise to spectral cues provided by the loci of stimulation maxima on the basilar membrane. However, we shall see that effects produced by phase shifting suggest that temporal analysis of neural response can play an important role in determining the pitch of rippled noise.

The introduction of a relative phase shift between the delayed and nondelayed components of the rippled stimulus produces a displacement of all spectral peaks by the same absolute value, and results in a change in both pitch value and pitch strength (Fourcin, 1965; Wilson, 1966; Bilsen and Ritsma, 1967/68, 1969/70; Ritsma and Bilsen, 1970; Yost and Hill, 1978; Yost, Hill and Perez-Falcon, 1978). The phase-shift condition studied most extensively for repetition pitch involves a change of 180° (RP-). When the polarity of either component is inverted, the maxima of the spectral ripples are displaced downward in frequency by half the delay reciprocal, so that the peaks of the "antiphasic" RP- stimulus are found at the position of troughs in the corresponding "cophasic" stimulus (RP+). This change in spectral positioning produces a relatively small change in pitch (roughly 10% in most studies). In addition, the pitch is ambiguous, with values of roughly -10% and +10% both being heard (Fourcin, 1965; Bilsen, 1966; Wilson, 1966; Bilsen and Ritsma, 1967/68, 1969/70; Yost, Hill and Perez-Falcon, 1978).

Recent theories of pitch perception applied to both RP+ and RP- include a spectral pattern matching model (Bilsen, 1977; Bilsen and Goldstein, 1974) and a filtered autocorrelation model (Yost and Hill, 1978, 1979; Yost, Hill and Perez-Falcon, 1978; Yost, 1982). These models share a common assumption of "spectral dominance" which considers that ". . . if pitch information is available along a large part of the basilar membrane the ear uses only the information from a narrow band. This band is positioned at 3 to 5 times the frequency value of the pitch" (Ritsma and Bilsen, 1970). That is, each model considers that the pitch of broadband rippled noise is determined by information contained within a narrow frequency band centered in the vicinity of the fourth spectral peak. In this respect, modern theories of RP are similar to pitch theories dealing with the line spectra of complex tones (Goldstein, 1973; Wightman, 1973; Terhardt, 1974, 1979; Srulovicz and Goldstein, 1983).

The pattern matching theory explains the dual pitches of broadband RP- by considering that the spectral information available from the dominant region (neighborhood of the fourth spectral peak) is used for calculation of the fundamental of a cophasic harmonic sequence having peaks at frequencies close to the actual fourth and fifth peaks of the RP- stimulus. This extrapolation has two solutions, resulting in "pseudofundamentals" at approximately 0.9/t Hz and 1.1/t Hz, in agreement with most empirical findings. Yost's autocorrelation theory considers that temporal information from the fourth spectral mountain is used for an autocorrelational analysis, yielding a two-valued solution equivalent to that resulting from the spectral "pseudofundamental" calculation.

There are compelling reasons to believe that the region in the vicinity of the fourth and fifth spectral peaks plays an important role in pitch perception (see Plomp, 1976, pp. 114-118). However, it is not at all certain that this region is the exclusive determinant of RP. For example, Wilson (1966) has reported that the dual pitches evoked by broadband RP- vary in the extent of their deviation from 1/t Hz as a function of t (with a minimum deviation of approximately 6% at t = 25 ms, and a maximum deviation of approximately 20% at t = 1 ms). Since the pitches based on calculated pseudofundamentals or autocorrelation peaks for the dominant region of an RP- spectrum deviate from 1/t Hz by a constant percentage regardless of t, Wilson's observations would seem to indicate that different spectral regions dominate at different values of t, or that changes in components outside the dominant region can influence broadband RP-.

One of the goals of the present study was to test the validity of spectral dominance theory by measuring RP+ and RP- under a variety of filtering conditions. As we shall see, the results obtained indicate that the "dominant" spectral region contributes to, but does not determine the pitch of broadband RP. The data obtained for RP together with other information in the literature suggest that a temporally based model can provide an explanation for both antiphasic and cophasic repetition pitch.

II. General Method

A. Preparation of Stimuli

For the preparation of rippled noise stimuli, white noise produced by a General Radio Model 1382 Noise Generator was band-pass filtered from 50 Hz to 8 kHz (General Radio Model 1952 Universal Filter: 30 dB/octave slopes) and then passed through a custom-modified Eventide Model BD955 Digital Delay Line (50 kHz sampling frequency and 10-bit coding) under the control of a Hewlett-Packard Model 3325A Frequency Synthesizer acting as an external clock. The delay line and external clock were adjusted to produce six values of the delay time, t, which corresponded to values of 1/t Hz ranging in whole tone steps from 110 Hz to 196 Hz (110, 123, 139, 156, 175, and 196 Hz, respectively). For each value of 1/t Hz, the delayed noise was added with unchanged polarity to the nondelayed noise to produce the stimuli for RP+, and with a polarity inversion (performed digitally within the delay line) to produce the stimuli for RP-. The delayed and nondelayed outputs from the delay line were each passed through separate matched Rockland Model 852 Dual Hi/Lo Filters (50 Hz to 8 kHz band-pass, with slopes of 48 dB/octave) and following this identical filtering were then mixed at equal amplitude by a Gately SPM-6 Stereo Mixer. These rippled noise stimuli were then passed successively through filters (Rockland Model 1042 and Wavetek/Rockland Model 751A) to produce the following five spectral ranges for both RP+ and RP- (filter slopes for all conditions were 211 dB/octave with cut-off frequencies set at the following positions): Broadband (50 Hz to 8000 Hz); band-pass from the third to the seventh spectral peak; low-pass up to the seventh spectral peak; band-reject between the third and seventh peak; and high-pass from the seventh spectral peak. Since peak frequencies change with delay setting and with polarity inversion, the cut-off frequencies of the filters were adjusted accordingly. In order to avoid edge-pitches produced by the steep filter slopes, for all conditions (except broadband) the rejected spectral components were replaced by uncorrelated white "filler" noise subjected to complementary filtering and having the same spectrum level (dB/Hz). All RP stimuli and filler noise bands, including the high- and low-pass components of the band-reject conditions, were recorded on separate tracks of an Ampex MM1200 16-track recorder at 15 ips, and were mixed down during the experiment using a Yamaha Model PM-430 audiomixer. The output of the mixer was subjected to a final low-pass filtering at 4 kHz (115 dB/octave slopes) to produce the rippled noise stimuli listed in Tables 1 and 3.

B. Subjects

Five listeners participated in this experiment. Two listeners (CG and JB) had had prior musical training and one (JB) had also had prior experience in repetition pitch matching. All listeners received between 1 and 8 hours of training in matching sinusoids to broadband RP+ using the method of adjustment. Each delay time was selected randomly from the range of 5 to 10 ms (20 μs steps), corresponding to pitch values from 200 to 100 Hz respectively. Listeners began their participation in the formal experiment when their pitch judgments corresponded to 1/t Hz within ±1.5% for each of six successive values of t in three successive blocks of practice trials.

C. Procedure

Preliminary training and formal testing were carried out in an audiometric room, with the rippled noise stimuli presented at 55 dBA SPL through diotically wired TDH-49 headphones. The rippled noise stimuli were matched with sinusoidal tones by the listener who adjusted the output of a Wavetek 166 function generator using both the main and vernier frequency control dials (calibration marks were concealed from view). The selected frequency match (measured with an accuracy of 0.01 Hz) was recorded by the experimenter who monitored a Hewlett-Packard Model 5316A Universal Counter/Timer, also concealed from the listener's view. During matching, the listener could switch between the rippled noise and the adjustable matching tone at will. The listener could also switch to an on-line white noise presented at the same intensity and having the same bandwidth (50 Hz to 4 kHz) as the RP stimulus. This flat-spectrum noise served as a neutral buffer, and listeners found it helpful when employed prior to the presentation of a new echo delay. No feedback or knowledge of results was provided during the study. Subjects could, at their option, defer matching at any particular value of t, and match at the next scheduled value before returning to the previous stimulus (this option was seldom used more than once per session). Listeners also had the option of canceling a session in progress if they did not wish to continue (this option was exercised 8 times out of a total of 293 sessions).

The five filtering conditions were presented in separate segments of the study and in the following order: (1) broadband, (2) high-pass, (3) low-pass, (4) band-reject, and (5) band-pass. Within each filtering condition, listeners completed all matches for RP+

before providing matches for RP-. There were five experimental sessions for RP+ matching, with listeners producing one match at each of the six delays in each session. Within sessions, the six echo-delays were presented in a pseudorandom order, with the restriction that the last delay in one session did not serve as the first delay in the next. The procedure for the RP- conditions was the same, except that some listeners needed more than 5 sessions. These additional sessions were necessary because of the dual pitches associated with RP-. Matches below 1/t Hz were much more frequent than those above (in keeping with reports by Fourcin, 1965; Bilsen, 1966; and Wilson, 1966), and listeners were required to repeat all judgments of an antiphasic condition until they had accumulated five matches below 1/t Hz at each value of t.

D. Results

The primary data employed for analysis were the averages of each listener's five adjustments of the matching sinusoidal tone under the various combinations of echo-delay, filtering condition and repetition phase shift (0° or 180°). The group results for the matching of RP+ under the five filtering conditions are presented in Table I, expressed as the average percent deviation of matches from 1/t Hz at each value of t. As shown in Table I, the accuracy of matching was high at all echo delays and under all filtering conditions, with an overall deviation from 1/t Hz averaging only 0.52%. The results for individual listeners, averaged across echo delays for each filtering condition, are presented in Table II.

-- Tables I and II About Here --

The group results for the matching of the antiphasic RP- stimuli are presented in Table III, and the results for individual listeners are presented in Table IV.

-- Tables III and IV About Here --

III. Discussion

Spectral Dominance versus Pitch Averaging. The data obtained in this study are not in accord with the spectral dominance theory, and an alternative broad-spectrum basis for pitch is proposed. Let us first relate our experimental data to predictions based on the spectral dominance theory.

The data gathered for normal or cophasic repetition (RP+) in this study serve mainly as a control for the measurements of the effect of filtering upon antiphasic repetition pitch (RP-). The broadband (unfiltered) pitch judgments for RP+ shown in Table I agree closely with the value of 1/t Hz (where t is the echo delay in seconds) reported by Bilsen and Ritsma (1969/70) and others. In addition, it was found that RP+ judgments approximated 1/t Hz for a variety of filtering conditions, including those for which the dominant spectral region for broadband rippled noise was absent. This finding is in keeping with the report that 1/3-octave bands of cophasic rippled noise outside the normally dominant region have values approximating 1/t Hz (Bilsen and Ritsma, 1969/70). The fact that other spectral regions produce the same repetition pitch does not conflict with spectral dominance theory since the theory considers that, although the region in the vicinity of the fourth spectral peak is the sole determinant of pitch when present, if absent, then other regions can give rise to repetition pitch. Still, the observation that the same pitch is heard for RP+ under various filtering conditions leaves open the possibility that spectral regions outside the range of "dominance" also contribute to the pitch of the broadband stimulus. Filtered antiphasic repetition pitch (RP-) can provide a critical test of this alternative to spectral dominance theory, and the current study was designed to provide such a test.

Our finding that RP- heard broadband differs from that heard for the dominant region when presented alone (band-pass condition) contradicts spectral dominance theory: If the dominant region were the sole determinant of pitch when present, then RP- for the broadband condition and for the dominant region band-pass condition should be the same. Data obtained under other filtering conditions suggest that the pitch heard broadband is based upon a pooling of the different pitches associated with particular spectral regions of the antiphasic spectrum. Thus, while the complementary band-pass and band-reject pitches each differ from the broadband pitch, their mean approximates that of the broadband condition, indicating that the effect produced by each alone is averaged when both are present simultaneously. Our data provide an additional example of pitch averaging: The pitch heard for the low-pass condition (which includes the dominant region and contains all peaks up to the seventh) deviates from that heard for the broadband condition by 4 units of standard error, but when averaged with the pitch associated with the complementary high-pass condition (all peaks down to the seventh), the value once again approximates that obtained for the broadband mixture of the complementary segments.

Polarity Inversion and Local Time Delays. If a cophasic RP+ with a time delay of t seconds is converted to antiphasic RP- by polarity inversion of the delayed sound, then an additional frequency-dependent time delay is introduced. For a 1/3-octave band (approximating a critical bandwidth) with a center frequency of f Hz, an additional delay of plus or minus half of the period of the center frequency is introduced, so that the overall delay is $t \pm 1/2f$ s, and pitch based upon the local repetition time becomes $1/(t \pm 1/2f)$ Hz. Using this simple expression, the calculated values for a decrease in pitch resulting from a polarity inversion at the regions of the fourth, fifth and sixth peaks are 12.5%, 10.0%, and 8.3%, respectively. The same expression for antiphasic repetition pitch was derived from the major positive peaks in the autocorrelation function (using simplifying assumptions) by Yost, Hill and Perez-Falcon (1978), and used by them to account for the empirical values for broadband RP- in terms of the pitch at the dominant spectral region.[1]

The empirical pitch values shown in Table III for the various filtering conditions of RP- correspond most closely to the delays calculated for the particular spectral peaks given in parentheses: Broadband (sixth); band-pass from third to seventh peaks (fifth); low-pass up to seventh peak (fifth); high-pass from the seventh peak (ninth); band-reject lacking third through seventh peaks (ninth). These values are consistent with the theory that repetition pitch (whether RP+ or RP-) is determined by the averaging of local temporally based pitch values.

A time-domain basis for RP is consistent with observations involving long repetition delays. Warren, Bashford and Wrightson (1980) reasoned that, if temporal processing were responsible for repetition pitch, it might be possible to detect repetition for delays extending beyond the limit for pitch even though spectral cues to repetition were unavailable. It was found that delays as long as 0.5 s could be detected and matched accurately. At this duration, neighboring spectral peaks were separated by only 2 Hz, which was much too close to permit resolution on the basilar membrane, and so temporal (autocorrelational?) analysis was responsible for detection of repetition. It was reported by Warren et al. that infrapitch repetition was insensitive to polarity inversion, so that antiphasic repetition was indistinguishable from cophasic repetition. This equivalence would be anticipated from a temporal theory since, for the long delays of infrapitch repetition, changes in delay times produced by polarity inversion drop below the just noticeable difference for all audible spectral regions.

In conclusion, it appears that while the region in the neighborhood of the fourth spectral peak contributes to the repetition pitch heard for broadband stimuli, it does not determine pitch as maintained by the spectral dominance theory. The results reported here for antiphasic repetition pitch, together with other evidence, indicate that the perceived pitch is based upon a pooling of information across critical bands. At each cochlear locus, the effective repetition period responsible for pitch is equal to the sum of the repetition delay and any additional local frequency-dependent delay produced by polarity inversion. The weighted average of these local time delays corresponds to the repetition pitch heard broadband.

Acknowledgments

The support of the Air Force Office of Scientific Research, Air Force Systems Command, USAF (grant agreement number AFOSR 85-0260) is gratefully acknowledged. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. This manuscript is submitted for publication with the understanding that the US Government is authorized to reproduce and distribute reprints for Governmental purposes.

We thank Bradley S. Brubaker for his valuable contributions.

References

Bilsen, F. A. (1966). "Repetition pitch: Monaural interaction of a sound with the repetition of the same, but phase shifted, sound," Acustica 17, 295-300.

Bilsen, F. A. (1977). "Pitch of noise signals: Evidence for a 'central spectrum'," J. Acoust. Soc. Am. 61, 150-161.

Bilsen, F. A., and Goldstein, J. L. (1974). "Pitch of dichotically delayed noise and its possible spectral basis," J. Acoust. Soc. Am. 55, 292-296.

Bilsen, F. A., and Ritsma, R. J. (1967/68). "Repetition pitch mediated by temporal fine structure at dominant spectral regions," Acustica 19, 114-115.

Bilsen, F.A., and Ritsma, R. J. (1969/70). "Repetition pitch and its implication for hearing theory," Acustica 22, 63-73.

Fourcin, A. J. (1965). "The pitch of noise with periodic spectral peaks," Reports of the 5th International Congress on Acoustics (Liège), 1A, B42.

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

Plomp, R. (1976). Aspects of tone sensation. London: Academic Press.

Ritsma, R. J. and Bilsen, F. A. (1970). "Spectral regions dominant in the perception of repetition pitch," Acustica 23, 334-339.

Srulovicz, P. and Goldstein, J. L. (1983). "A central spectrum model: A synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum," J. Acoust. Soc. Am. 73, 1266-1276.

Terhardt, E. (1974). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. 55, 1061-1069.

Terhardt, E. (1979). "Calculating virtual pitch," Hear. Res. 1, 155-182.

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

Wightman, F. L. (1973). "The pattern transformation model of pitch," J. Acoust. Soc. Am. 54, 407-416.

Wilson, J. P. (1966). "Psychoacoustics of obstacle detection using ambient or self-generated noise," in Animal Sonar Systems edited by R. G. Busnel (Louis-Jean, Gap, Hautes-Alpes, France) pp. 89-114.

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

Yost, W. A., and Hill, R. (1978). "Strength of pitches associated with ripple noise," J. Acoust. Soc. Am. 64, 485-492.

Yost, W. A., and Hill, R. (1979). "Models of the pitch and pitch strength of ripple noise," J. Acoust. Soc. Am. 66, 400-410.

Yost, W. A., Hill, R., and Perez-Falcon, T. (1978). "Pitch and pitch discrimination of broadband signals with rippled power spectra," J. Acoust. Soc. Am. 63, 1166-1173.

Footnote

[1] A temporal explanation for repetition pitch heard for band-passed filtered pulse pairs was proposed in a brief letter published by Bilsen and Ritsma (1967/68). They suggested that repetitive features of the fine structure of the waveform produced at discrete loci on the basilar membrane were responsible for the RPs heard. This temporal theory was subsequently discussed more fully by them (Bilsen and Ritsma, 1969/70), and they explicitly stated, (p. 67) "It is important to note that, in the case of continuous noise with its repetition, Repetition Pitch cannot possibly result from a process of detection of a temporal envelope because this is, essentially, missing . . . ." However, as pointed out by Yost et al. (1978), an autocorrelational analysis of neural patterns of stimulation could be used to determine delay times for iterated continuous noise (and hence RP) at individual loci on the basilar membrane.

Table I. Mean percent deviation from 1/t Hz for matches of pure tones to various spectral ranges of filtered RP+. Results shown are means and standard errors ($sd/\sqrt{5}$) for the average matches of 5 listeners.

	1/t Hz						
Filtering of RP+	110	123	139	156	175	196	Grand Mean
Broadband (50-4000 Hz)	0.05	0.34	0.53	0.41	0.47	0.68	0.41
	(0.08)	(0.15)	(0.29)	(0.13)	(0.25)	(0.23)	(0.09)
Band-pass (3rd-7th peak)	0.55	-0.03	0.11	0.25	0.39	0.95	0.37
	(0.23)	(0.46)	(2.24)	(0.97)	(0.61)	(1.74)	(0.14)
Low-pass (1st-7th peak)	0.32	0.61	1.47	0.60	0.59	2.11	0.95
	(0.31)	(0.23)	(0.87)	(0.31)	(0.13)	(1.34)	(0.33)
Band-reject (3rd-7th peak)	0.00	1.02	-0.70	-0.34	-0.77	-0.42	-0.20
	(1.01)	(1.34)	(0.17)	(0.15)	(0.38)	(0.50)	(0.25)
High-pass (7th peak-4000 Hz)	-0.92	-0.28	-0.41	-1.72	-0.04	-0.63	-0.67
	(0.57)	(0.40)	(0.53)	(0.31)	(0.41)	(0.43)	(0.12)

Table II. Mean percent deviation from 1/t Hz for pure tone matches to RP+ under five filtering conditions. Results shown are means and standard errors $(\mathrm{sd}/\sqrt{6})$ for the average matches of individual listeners at six values of t.

	Filtering Condition				
Subject	Broadband	Band-pass	Low-pass	Band-reject	High-pass
CG	0.57 (0.14)	0.18 (0.32)	0.82 (0.12)	-0.54 (0.30)	-0.90 (0.61)
DG	0.13 (0.15)	0.57 (0.68)	1.23 (0.15)	0.33 (1.17)	-0.28 (0.14)
JB	0.31 (0.09)	0.36 (0.12)	0.17 (0.19)	-0.11 (0.15)	0.01 (0.13)
BB	0.67 (0.24)	-0.41 (0.21)	1.48 (2.45)	-1.37 (0.21)	-3.64 (0.76)
MM	0.38 (0.23)	-1.30 (0.69)	1.05 (0.75)	0.68 (0.70)	1.54 (0.44)
Mean	0.41 (0.09)	0.37 (0.14)	0.95 (0.33)	-0.20 (0.25)	-0.67 (0.12)

Table III. Mean percent deviation from 1/t Hz for lower pitch matches of pure tones to various spectral ranges of filtered RP-. Results shown are means and standard errors (sd/$\sqrt{5}$) for the average matches of 5 listeners.

	1/t Hz						
Filtering of RP-	110	123	139	156	175	196	Grand Mean
Broadband (50-4000 Hz)	-7.08	-7.49	-8.18	-9.26	-9.12	-8.18	-8.22
	(0.10)	(0.31)	(0.67)	(0.98)	(0.77)	(0.65)	(0.35)
Band-pass (3rd-7th peak)	-8.62	-10.07	-10.28	-9.66	-13.98	-11.91	-10.75
	(0.85)	(0.49)	(1.12)	(1.18)	(2.24)	(5.28)	(0.78)
Low-pass (1st-7th peak)	-8.53	-9.62	-9.79	-10.00	-9.55	-10.30	-9.63
	(0.84)	(0.22)	(0.24)	(0.32)	(0.36)	(0.73)	(0.25)
Band-reject (3rd-7th peak)	-5.22	-5.96	-5.86	-6.39	-6.24	-6.36	-6.01
	(0.38)	(0.36)	(1.07)	(0.52)	(0.69)	(0.72)	(0.18)
High-pass (7th peak-4000 Hz)	-5.25	-5.15	-5.90	-5.51	-5.31	-5.72	-5.47
	(0.40)	(0.20)	(0.38)	(0.51)	(0.51)	(0.57)	(0.12)

Table IV. Mean percent deviation from 1/t Hz for the lower pitch matches of pure tones to RP- under five filtering conditions. Results shown are means and standard errors ($sd/\sqrt{6}$) for the average matches of individual listeners at six values of t.

	Filtering Condition									
Subject	Broadband		Band-pass		Low-pass		Band-reject		High-pass	
CG	-7.58	(0.45)	-9.89	(0.22)	-9.59	(0.31)	-5.06	(0.19)	-5.01	(0.12)
DG	-7.97	(0.33)	-9.36	(0.44)	-9.25	(0.21)	-5.72	(0.12)	-5.19	(0.31)
JB	-7.41	(0.14)	-9.68	(0.16)	-9.58	(0.10)	-5.36	(0.11)	-5.29	(0.19)
BB	-9.19	(0.93)	-15.93	(2.72)	-10.88	(0.44)	-6.75	(0.87)	-6.93	(0.30)
MM	-10.30	(0.54)	-10.07	(2.01)	8.76	(1.26)	-7.25	(0.91)	-5.48	(0.79)

APPENDIX B

AFOSR-TR- 88-1160

Perception of complex tone pairs mistuned from unison

by

Richard M. Warren

James A. Bashford, Jr.

Bradley S. Brubaker

University of Wisconsin-Milwaukee

Department of Psychology

Milwaukee, Wisconsin 53201

Received:

ABSTRACT: When a broad-spectrum complex tone (CT) having a stable frequency of 600 Hz or less and containing several harmonics above the 8th is mixed with itself after a slight change in the waveform repetition frequency (1 Hz or less), listeners hear a rising glissando when corresponding portions of the correlated waveforms approach alignment and a falling glissando as they recede from alignment. Glissandi are unimpaired if harmonics below the 8th are absent, but if, instead, harmonics above the 8th are removed, only amplitude fluctuations are heard (not glissandi). When two broad-spectrum uncorrelated CTs mistuned slightly from unison are mixed, complex periodic patterns other than glissandi are heard. These observations, along with others involving CTs mistuned from unison, provide information concerning the nature of frequency domain and time domain mechanisms employed for the perception of iterated acoustic patterns.

INTRODUCTION

It appears that there has never been a systematic study of the perception of complex tones mistuned slightly from unison (frequency ratio of 1:1). Yet such a study can provide information of interest to auditory theory.

When a pair of mistuned complex tones having fundamental frequencies which differ by Δf_0 are mixed, their corresponding harmonics (that is, those with the same harmonic number, n), beat at a rate of $n\Delta f_0$. The beat rates thus form a harmonic series consisting of integral multiples of Δf_0. There has been a brief report (Warren, 1978) that under some conditions, these harmonically related beats are integrated perceptually to form a complex pattern of amplitude fluctuations heard to repeat at the same rate as the beats produced by the spectral fundamentals.

However, the interaction of complex tones mistuned from unison may be viewed from a different perspective. Consider two broad spectrum complex tones identical in every way (the same waveform and the same amplitude and phase spectra). If the waveform of one of these complex tones is stretched slightly (so that its repetition frequency drops by a small amount) and the two tones are then mixed, the temporal separation of corresponding portions of these "congruent" waveforms will be continually changing, with alignment occurring at a rate equal to the difference in frequency of the complex tones (Δf_0). It is known that when a complex sound is mixed with itself following a delay of t seconds, a "repetition pitch" of 1/t Hz may be heard (for review, see Plomp, 1976, pp. 138-139). Since the mistuned congruent tones have continually changing displacements from synchrony, pitch glides should be heard: As their waveforms move away from alignment a downward gliding pitch should be produced by the increasing value of t, followed by a rising pitch glide as the waveforms move past the point of maximum separation and back toward alignment. Such glissandi have been reported for pairs of pulsate periodic stimuli mistuned from unison (Thurlow and Small, 1955), and the changing pitch was attributed to the changing temporal separation of the discrete pulses (see also Small and McClellan, 1963). The present study demonstrates that these glides are not restricted to mistuned periodic pulses, but are found for the more general case of mistuned nonpulsate complex tones.[1] It should be emphasized that these glissandi would be anticipated only for the mistuning of complex tones with congruent or nearly congruent waveforms. When a pair of mistuned complex tones have independent waveforms with uncorrelated phase spectra for corresponding harmonics, there can be neither alignment nor delay from alignment of corresponding portions of the component waveforms, and hence no pitch glide would be expected.

I. PRELIMINARY OBSERVATIONS

A number of informal observations were made by a panel of four psychoacoustically experienced listeners who listened to the interaction of two broadband complex tones which had been mistuned slightly from unison and then mixed. These complex tones were presented diotically through headphones, and consisted of all harmonics up to 8 kHz. All tone pairs had one member with a randomly determined phase spectrum, while the other member of the pair was either "congruent" (identical except for a slight difference in frequency) or "uncorrelated" (having a slight difference in frequency and independent randomly determined phases for corresponding harmonics). Details of the manner in which such complex tone pairs were produced will be given subsequently in Section II (General Methods). Since these preliminary observations formed the basis for the formal study, they will be listed below.

A. Congruent Complex Tones Mistuned From Unison

1. Glissandi could be heard with ease only when the fundamental frequencies of the tone pairs were 600 Hz or less.
2. There was no lower frequency limit for tone pairs producing glissandi, and pitch glides were also heard clearly for mistuned complex waveforms having infratonal repetition frequencies (that is, repetition frequencies below the 20 Hz tonal limit).
3. Glissandi were not heard if the extent of mistuning exceeded a critical value. This limit for perception of pitch glides was a discontinuous function of the fundamental frequencies of the complex tones, with one function applying below 50 Hz, and another above 50 Hz.
4. Glissandi required the presence of higher but not the lower harmonics of the complex tones. When tone pairs producing glissandi were high-pass filtered above the 7th harmonic, perception of pitch glides appeared unimpaired. However, glissandi were not heard when the same tone pairs were low-pass filtered below the 8th harmonic (complex patterns of amplitude fluctuation were heard instead).

B. Uncorrelated Complex Tones Mistuned From Unison

1. Pitch glides could not be perceived, but complex patterns of amplitude fluctuation were heard to repeat under conditions A1 through A4 described above. Removal of the spectral fundamentals of the tones did not change either the clarity or ensemble repetition rate of these complex patterns.

2. The beating of individual harmonic components of the complex tones was difficult to hear when their fundamental frequencies differed by 1 Hz or more. But when the uncorrelated tones were mistuned by less than about 0.5 Hz, the integrated or ensemble periodicity became less salient and the individual beat rates were clearly dominant.

II. GENERAL METHODS

A. Subjects

Listeners were all trained in psychoacoustic experimentation. Between four and six experienced listeners served as observers for each of the phenomena described, and all of the phenomena were heard by each of the listeners unless otherwise stated.

B. Stimuli and Apparatus

Two types of complex tones were employed. "Frozen noise" tones were generated by the repetition of waveforms excised from a 100 to 8000 Hz band of pink noise. These tones had randomly determined amplitudes and phases for harmonic components which extended from the fundamental frequencies up to 8000 Hz (for a discussion of repeated noise segments as model periodic stimuli see Warren, 1982, pp. 78-80). "Synthesized" complex tones consisted of all harmonics lying between the fundamental frequency and 8 kHz (each harmonic had the same amplitude and an individually specified, randomly determined phase). These tones were generated from polynomial equations by a Data Precision Co. Polynomial Waveform Synthesizer Model 2020-100 [512,000/16-bit data point capacity, operated at a sampling frequency of 50 kHz].

Two matched factory-modified digital delay lines (Eventide Model BD955) were used in the production of all stimuli. When placed in recirculating or "looped" mode, the delay lines repeated stored input without change. The delay lines (operated in conjunction with appropriate antialiasing and reconstruction filters) had a flat frequency response (± 1 dB) from 50 Hz through 16000 Hz and a 60 dB dynamic range, based upon a sampling frequency of 50 kHz and 10-bit coding. The storage times (and repetition periods) used in the present study ranged from 2.5 ms through 1 s. To produce the congruent periodic stimuli mistuned from unison, the same waveform (either a noise segment or a synthesized waveform) was repeated on both delay lines at identical repetition frequencies, and then one of these identical waveforms was temporally stretched by lowering the clock frequency driving one of the delay lines by the desired amount. For the uncorrelated tones based on frozen noise, members of a tonal pair were derived from separate segments of a noise, and for the synthesized uncorrelated tones derived from polynomial equations, a different set of randomly determined phases were assigned to the equations for the corresponding harmonics of a complex tone pair. The initial 8 MHz clock frequency driving the two delay lines was produced by separate Hewlett-Packard Model 3325A Frequency Synthesizers locked to the same time base. These synthesizers were adjustable in steps of one milliHertz, and changes in the repetition period of the stored waveforms were executed by a Hewlett-Packard Model 85 Computer which controlled the frequency of one of the delay line clocks. The outputs of the delay lines were combined using an audio mixer.

Spectra of stimuli were monitored with a Brüel & Kjaer Model 2033 Spectrum Analyzer, and waveforms were monitored with a 2-channel digital storage oscilloscope (Nicolet Model 3091). Sound spectrograms were generated by a Kay Model 7800 Digital Sonagraph.

When desired, high-pass and low-pass filtering of the mixed iterated waveforms was accomplished with a Wavetek/Rockland Model 751A Filter having attenuation slopes of 115 dB/octave.

C. General Procedure

Subjects were seated in an audiometric room, and the stimuli were delivered diotically through a matched set of headphones (either TDH-49 or Sennheiser HD230) at a comfortable level selected by the listener (usually between 40 and 70 dBA) unless otherwise noted.

III. EXPERIMENT 1: LIMITS OF MISTUNING PRODUCING PITCH GLIDES

Preliminary experiments indicated that there were two linear functions representing the maximum deviation from unison of congruent complex tones for which pitch glides could be heard--one function applied to pairs with fundamental frequencies above 50 Hz, and the other to pairs with fundamental frequencies below 50 Hz. The present experiment measured these functions and, as we shall see, revealed that a common rule applied to both. It was also found that the linear function found for the lower tonal range (50 Hz to 20 Hz) applied as well to periodic sounds with infratonal fundamental frequencies (20 Hz to 1 Hz).

A. Stimuli and Procedure

The fundamental frequencies of the congruent pairs of frozen noise tones before mistuning from unison were 25, 50, 100, 200, and 400 Hz. In addition, congruent frozen noise waveforms repeated at infratonal frequencies of 1, 2, 4, 8, and 16 Hz were employed.[2] Each pair of mistuned periodic stimuli were derived from a single randomly selected pink noise segment of desired duration which was repeated on both delay lines. The mistuning from unison was in steps which varied with the initial iteration frequency. For repetition periods from 1 s through 20 ms, these steps were integral multiples of 0.5% of the initial values, and for periods from 10 through 2.5 ms, steps were multiples of 0.1% (as described earlier, mistuning was accomplished through a computer program controlling the clock frequency driving one of the delay lines).

Subjects were instructed to report hearing glissandi only when a gliding pitch was clearly perceived. They received 5 blocks of trials, each of the blocks consisting of judgments involving the 10 repetition frequencies (ranging from 1 Hz to 400 Hz) which were presented in random order. Every repetition frequency within a block was judged twice using the Method of Limits. For the first of these judgments, the experimenter mistuned one frequency beyond the point where glissandi could be heard (the initial mistuning was randomly selected from the range of 5 to 12 steps of the size described above). Mistuning was decreased stepwise until glissandi were heard. The experimenter then decreased the extent of mistuning further by an additional two to four units (subjects always indicated that they heard glissandi for this extent of mistuning). The extent of mistuning from unison was then increased systematically, using steps of the size described above, until the subject reported that a glissando could not be heard. The upper limit for hearing glissandi on that pair of trials was considered to be the average of the ascending and descending orders of presentation. A particular frozen noise segment was used for only one pair of trials for a single subject.

For frequencies of 1 and 2 Hz, the minimum time required for a pitch glide period (a pair of rising and falling glissandi) was inconveniently long--therefore only the falling pitch glide following initial waveform alignment was judged, and each frozen noise segment was used for only one judgment.

B. Results and Discussion

As shown in Fig. 1, clear glissandi could not be heard for mistuned congruent tone pairs from 50 through 400 Hz if the difference in fundamental frequencies (Δf_o) was greater than about 1 Hz. For stimulus pairs with waveform repetition frequencies from 50 Hz down to 1 Hz, it can be seen that the limiting value of Δf_o for pitch glides was proportional to the repetition frequencies of the pairs of periodic sounds, so that two linear functions were obtained which intersected at 50 Hz.

-- Figure 1 About Here --

The acoustic interactions of the harmonics of complex tones mistuned from unison are illustrated in Fig. 2. The sound spectrograms show the amplitude changes produced by the beating harmonics for 99.5 and 100 Hz congruent tones (top spectrogram) and for 99.5 and 100 Hz uncorrelated tones (bottom spectrogram). It can be seen that the rate of beating of corresponding spectral components (that is, components of the mixture having the same harmonic numbers) is equal to the fundamental beat rate (0.5 Hz) multiplied by the harmonic number whether the tone pair is congruent or uncorrelated. Fig. 2 shows that for a congruent tone pair, beating harmonics all have their amplitude maxima occurring simultaneously once each two seconds at the alignment of the corresponding portions of the

-- Figure 2 About Here --

two waveforms. The upward sweeping pattern of the spectrogram is produced as the waveforms move toward alignment and the downward sweeping pattern as the waveforms move away from alignment. The sound spectrograms exhibit the same resolution in Hertz for the upper and lower harmonics, since the successive band-pass filters each have the same bandwidth (the sound spectrograms shown in Fig. 2 are based upon a fixed bandwidth of 150 Hz for all filters). However, the resolving power of our auditory system is rather different from that of a spectrograph, with auditory resolution in Hertz decreasing with increasing harmonic number (see Plomp, 1976, pp. 1-25). As will be shown in Experiment 2, glissandi require the presence of unresolved upper harmonics but not resolvable lower harmonics.

Some additional observations concerning the effects of mistuning congruent tones were made by our listeners. When the extent of mistuning was raised just above the limit for glissandi, the upward and downward pitch glides disappeared, and were replaced by brief "chirps" having no clear monotonic pitch changes. At mistunings about four times the glissando limit (corresponding to about 4 Hz mistuning for tonal frequencies from 50 through 400 Hz), chirps were no longer perceived, but instead a periodic amplitude pattern was heard to repeat at a rate equal to the difference between the fundamental frequencies.

The intersecting straight-line functions in Fig. 1 each conform to the following rule: Glissandi cannot be heard unless the individual glide durations last at least 0.5 s. Let us consider first how this rule applies to the tonal range from 50 to 400 Hz, and results in the horizontal line segment of Fig. 1. This horizontal line corresponds to a pitch-glide cycle extending from one waveform alignment to the next which is repeated at a frequency of 1 Hz (period of 1 s), with the temporally contiguous upward and downward glides each lasting 0.5 s. Adding a replica of a sound to itself after a delay of t seconds produces a "repetition pitch" corresponding to a tone of 1/t Hz (see Plomp, 1976, p. 139). Glissandi may be considered as gliding repetition pitches, with the greatest possible delay equal to half the average of the repetition periods of congruent tones. This maximum delay corresponds to a lower pitch limit of the glide one octave above that of the tones. Despite the changes in this lower pitch limit from 800 Hz (for the 400 Hz tonal pair) to 100 Hz (for the 50 Hz tonal pair), the limiting duration of each glide remained at 0.5 s, so that, rather surprisingly, it was the glide time rather than rate or extent of pitch change which determined the limit of mistuning for glissandi. The lowest pitch attained by a glide (100 Hz pitch corresponding to a waveform asynchrony of 10 ms) was reached for congruent stimulus pairs with one member having a repetition frequency of 50 Hz. For repetition frequencies below 50 Hz, a temporal gap was heard between the end of a downward glide and the start of an upward glide: Nevertheless, the 10 ms asynchrony or delay limit for hearing the gliding pitch still held, and the straight-line segment with a positive slope extending from 1 Hz through 50 Hz in Fig. 1 corresponded to mistunings which reached this 10 ms asynchrony (repetition pitch of 100 Hz) 0.5 s after fine structure alignment.

IV. EXPERIMENT 2: HARMONIC COMPONENTS REQUIRED FOR PITCH GLIDES

A glance at Fig. 2 suggests that glissandi might result from the alternating upward and downward frequency sweeps of the pattern produced by beat maxima. Were this the case, then the resolvable pairs of beating lower harmonics would be expected to produce clear glissandi for complex tones mistuned from unison. This possibility was tested by mixing 99.5 Hz and 100 Hz congruent tones derived from the same frozen noise segment which were low-pass filtered at the sixth harmonic (filter slopes of 115 dB/octave) before presentation to listeners. Glissandi could not be heard following removal of unresolved higher harmonics, and only periodic amplitude fluctuations produced by the beating lower harmonics could be heard. In order to confirm this observation and to determine more accurately the low-pass threshold for glissandi, congruent complex tones were synthesized from polynomial equations.

A. Stimuli and Procedure

The polynomial waveform synthesizer was used to generate a series of complex tones with fundamental frequencies of 100 Hz. The members of this series consisted of all harmonics up to the 4th, 6th, 8th, 10th, 12th, and 14th, respectively. Each of these complex tones had harmonics of equal amplitude and different randomly assigned phases, and was used to produce a congruent tone pair. A second series of complex tone pairs was synthesized which differed from the first series only in having different random assignments of phase. The complex tones of each series were used to produce congruent tone pairs mistuned from unison with fundamentals of 100 and 99.8 Hz, using the procedure described in the General Methods Section. The six listeners judged whether glissandi were heard. The Method of Limits was used, starting with the congruent tones having fourteen harmonic components (which always produced glissandi) and continuing with complex tones having successively decreasing harmonic numbers until glissandi were no longer heard. Judgments were then made starting with complex tones having only four harmonics (for which glissandi were never heard) and continuing with complex tones having successively increasing numbers of harmonics until glissandi were reported. After the first pair of judgments, the highest harmonic for the decreasing series was randomly selected as 14 or 12, and for the increasing series as 4 or 6. After three trials (pairs of judgments), an additional three trials were run with the second series of complex tones having different randomly assigned phases. The threshold value for each trial was the average for the descending and ascending modes of presentation.

B. Results and Discussion

Table I gives the thresholds obtained with six subjects for each of the complex tones. It can be seen that, on the average, harmonics above the eighth must be present for glissandi to be heard.

--Table I About Here--

Commenting on a preliminary report by Warren, Brubaker, and Gardner (1984) that described glissandi produced by mistuned congruent tones, Hartmann (1985) suggested that the pitches heard were caused by the effects of a "frequency domain grating" analogous to the effects of diffraction gratings used in optics. The glides were attributed to the orderly progression of spectral maxima, as if a filter were swept through the harmonics of the complex tones. Hartmann considered that not all harmonics entered into pitch glides (in the case of the first eight harmonics only four would enter into the orderly progression of maxima), and that when only a small number of harmonics formed a progression, glides were difficult to hear. However, there is reason to believe that the glissandi heard do not correspond to the frequencies of spectral maxima.

In order to determine if glissandi can reach pitches outside the range of spectral components, we synthesized mistuned congruent tones, each consisting of the 71 harmonics from the 10th through the 80th. One member of the pair had a (missing) fundamental of

100 Hz with harmonics of equal amplitude and randomly assigned phases. The other member of the pair was derived from the first, and had a (missing) fundamental of 99.95 Hz. When mixed, a glissando pair was produced having a period of 20 s with a falling pitch lasting 10 s and a rising pitch of the same duration. Of especial interest was the fact that at the transition from falling to rising pitch, all odd numbered harmonics were canceled and a complex tone was formed consisting of the 5th through the 40th harmonics of 200 Hz. A pitch corresponding to the new (missing) fundamental of 200 Hz was heard clearly at this point, and this pitch was continuous both in value and in quality with the falling and rising glides. Since Hartmann's theory would limit the range of pitches heard to those of the harmonics actually present, it would seem that this particular spectral explanation for glissandi does not apply.[3]

We have seen that mistuned congruent complex tones do not produce glissandi if harmonics above the 8th are absent. Spectral resolution of harmonic components can be accomplished by listeners up to about the 8th (Plomp, 1964; Plomp and Mimpen, 1968), so the presence of unresolved harmonics may be necessary for glissandi. Alternatively, glissandi may require more than eight harmonics for each member of the congruent tone pair. In order to test these hypotheses, a complex tone was synthesized consisting of eight harmonics of 100 Hz from the 10th through the 17th, each harmonic having the same amplitude and a different randomly determined phase. This complex tone was used to generate a congruent tone with a fundamental frequency of 99.8 Hz. All six listeners heard a pair of rising and falling pitch glides repeated each 5 seconds, suggesting that the lack of glissandi for congruent tones consisting of harmonics from the fundamental through the eighth is attributable to their spectral resolution rather than to the small number of components.

What are the bases for hearing glissandi? It appears that unresolved harmonics are required, and that the pitch of the glide at any moment seems to correspond to the delay separating corresponding portions of the congruent waveforms at that time. Both temporal and spectral explanations for glissandi are possible. First the spectral explanation: As the congruent waveforms move out of alignment, comb filtering produces peaks and troughs in the spectral envelope. With the initially small asynchronies, the interpeak spacing is large and the pitch heard (which corresponds to the frequency separation of adjacent peaks) is high. As asynchrony becomes larger, the interpeak separation becomes smaller, and the pitch drops until the comb filtering cancels all odd harmonics at the point of maximum asynchrony. The pitch then rises with the decreasing asynchrony, and the cycle repeats. A possible temporal explanation involves the superimposed patterns generated by the unresolved harmonics of each of the congruent complex tones. An autocorrelational analysis can lead to a pitch equivalent to the reciprocal of the delay separating these patterns (for a detailed discussion of repetition pitch based upon the autocorrelation of neural patterns, see Yost, Hill, and Perez-Falcon, 1978).

V. MISCELLANEOUS OBSERVATIONS

A. Perception of Complex Patterns Produced by Uncorrelated Tone Pairs

Uncorrelated complex tones (that is, tones having independently assigned random phases for corresponding harmonic components) cannot produce pitch glides when mixed. There can be no movement toward or away from alignment of corresponding portions of the waveforms (and no corresponding systematic spectral interaction) as with mistuned congruent tone pairs (see Fig. 2). However, mistuned complex tones do produce patterns of amplitude fluctuation which are heard to repeat at a rate equal to the difference in fundamental frequency of the tones.

When 198 and 200 Hz uncorrelated complex tones were mixed, a complex amplitude pattern was heard clearly to repeat at 2 Hz. The beating of the fundamentals of the complex tones at 2 Hz was not necessary for hearing this repetition frequency, since a pattern was perceived to repeat twice a second with undiminished clarity when the fundamentals were absent. There appear to be two different mechanisms responsible for the 2 Hz iteration, one operating with unresolved upper harmonics, the other with resolved lower harmonics.

When the 198 and 200 Hz uncorrelated complex tones lacked the first nine harmonics (they were synthesized from polynomial equations and consisted of harmonics 10 through 40, each at equal amplitude), a clear 2 Hz periodicity was heard. When a tunable 1/3-octave filter (approximating a critical band) was swept slowly through the spectrum of the mixed complex tones, the 2 Hz repetition could be heard at all center frequencies. Examination of the waveforms of these bands showed periodic complex envelopes which were repeated each 500 ms at all center frequencies.

When the synthesized uncorrelated 198 and 200 Hz tones each consisted only of the first 7 harmonics, a complex 2 Hz pattern was again heard clearly. In addition, some (but not all) of the harmonically related simple beat rates of the resolved lower harmonic pairs could also be heard. When listeners could hear the beating of individual harmonic pairs, they were generally modulated or accented at the 2 Hz frequency. As with the mistuned pairs of broadband and high-pass complex tones, the ensemble complex beat rate of 2 Hz did not require the presence of the simple 2 Hz beat rate of the spectral fundamentals, for a pattern repeating twice a second was still heard clearly when only harmonics 2 through 7 were present. In order to minimize the possibility that the 2 Hz periodicity resulted from the interaction of harmonic distortion products, listening was carried out at low levels. When the signal was 15 dB SL, the 2 Hz periodicity was heard clearly by all listeners. An 80-300 Hz bandpass noise (filter slopes 96 dB/octave) was then introduced at 15 dB SL and the signal readjusted to 15 dB SL in the presence of the noise (which would be expected to mask any low level distortion products). Once again, all listeners heard a 2 Hz repetition matching the frequency of the missing fundamental beat rate (the beat rates present in the acoustic signal were 4, 6, 8, . . . 14 Hz). It should be emphasized that the ensemble complex beat produced by the lower harmonics was quite different from the simple waxing and waning of amplitude produced by a pair of beating sinusoidal tones, consisting rather of a complex periodic pattern of amplitude modulations. Unlike the 2 Hz iterance described earlier for 1/3-octave bands of unresolved harmonics, the 2 Hz periodicity heard with only the first 7 harmonics involved the integration of harmonically related periodic patterns across different neural frequency channels.

B. Glissandi Involving Vowels and Other Special Sounds

In keeping with the concept that iterated randomly derived waveforms can serve as exemplars or model periodic stimuli, the observations reported can be applied to other types of periodic stimuli as well (keeping in mind any special characteristics of particular sounds). Congruent pairs of broadband complex tones produced by standard laboratory generators (pulse trains, sawtooth waves, etc.) can be used to produce pitch glides. However, congruent tones consisting of only odd harmonics (e.g., square waves) produce glissando pairs (rising and falling glides) at rates twice that of all-harmonic tones having the same extent of mistuning. Glissandi can also be heard for mistuned vowels (although the glides are somewhat weaker than for broadband stimuli lacking pronounced formants): When a single glottal pulse of the vowel "ee" was repeated on two digital delay lines, and the clock frequency driving one delay line was changed slightly, glissandi were heard by all of our listeners. Faint pitch glides were heard when the vowel was mistuned slightly from unison with a pulse train (all harmonics in cosine phase). Apparently, glottal buzzes, even after passage through the vocal tract, are sufficiently similar to pulse trains to cause listeners to hear systematic pitch changes. However, any particular broadband complex tone (whether pulse train, vowel, or other) when mixed with an iterated frozen noise segment formed an uncorrelated pair, and no hint of glissandi could be heard.

Acknowledgment

The support of the Air Force Office of Scientific Research, Air Force Systems Command, USAF (grant agreement number AFOSR 85-0260) is gratefully acknowledged. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. This manuscript is submitted for publication with the understanding that the US Government is authorized to reproduce and distribute reprints for Governmental purposes.

References

Hartmann, W. M. (1985) "The frequency-domain grating," J. Acoust. Soc. Am. 78, 1421-1425.

Plomp, R. (1976). Aspects of tone sensation (Academic Press, London).

Plomp, R. (1964). "The ear as a frequency analyzer," J. Acoust. Soc. Am. 36, 1628-1636.

Plomp, R., and Mimpen, A. M. (1968). "The ear as a frequency analyzer. II," J. Acoust. Soc. Am. 43, 764-767.

Small, A. M., Jr., and McClellan, M. E. (1963). "Pitch associated with time delay between two pulse trains," J. Acoust. Soc. Am. 35, 1246-1255.

Thurlow, W. R., and Small, A. M., Jr. (1955). "Pitch perception for certain periodic auditory stimuli," J. Acoust. Soc. Am. 27, 132-137.

Warren, R. M. (1978). "Complex beats," J. Acoust. Soc. Am. 64, S38 (Abstract).

Warren, R. M. (1982). Auditory perception: A new synthesis (Pergamon, Elmsford, New York).

Warren, R. M., Brubaker, B. S., and Gardner, D. A. (1984). "Perception of complex tone pairs mistuned from unison: Waveform relations determine whether pitch glides or iterated complex auditory patterns are heard," J. Acoust. Soc. Am. 75, S20 (Abstract).

Warren, R. M., and Wrightson, J. M. (1981). "Stimuli producing conflicting temporal and spectral cues to frequency," J. Acoust. Soc. Am. 70, 1020-1024.

Yost, W. A., Hill, R., and Perez-Falcon, T. (1978) "Pitch and pitch discrimination of broadband signals with rippled power spectra," J. Acoust. Soc. Am. 63, 1166-1173.

FOOTNOTES

[1] In an earlier study, we reported that several other perceptual phenomena which have been reported for periodic pulse trains (and attributed to their pulsate nature) can also be observed with periodic nonpulsate sounds (Warren and Wrightson, 1981).

[2] It has been suggested that the tonal and infratonal repetition together form a single perceptual continuum of detectable acoustic repetition called "iterance" which extends from a lower limit of 1 Hz through the upper limit of audibility at about 16,000 Hz. There is evidence indicating that mechanisms subserving iterance detection have some degree of overlap in the tonal and infratonal ranges (see Warren, 1982, pp. 80-85).

[3] Pitch glides corresponding to changing spectral maxima of harmonics, or groups of harmonics were heard with careful listening, and were more evident at high sensation levels. These glides covered short frequency ranges, were much briefer than the major glissandi, and seemed to appear haphazardly at different carrier frequencies.

Table I. Number of successive harmonics for each complex tone (starting with the fundamental) required for perception of glissandi with congruent tones of 100 and 99.8 Hz (for further description, see text).

	Trial Number						Average
Subject	1	2	3	4	5	6	
JB	8	8	8	9	9	10	8.66
BB	11	11	11	9	9	10	10.16
PK	10	11	9	11	9	9	9.83
JR	8	8	7	6	6	6	6.83
DT	7	6	8	6	7	7	6.83
RW	8	7	9	8	7	7	7.67
Combined	8.67	8.50	8.67	8.17	7.83	8.17	8.33

Figure Captions

Fig. 1. Pitch glide limits: The greatest mistuning from unison for which glissandi can be heard. The frequency of one member of the pair is given by the abscissa, and the decrease in frequency of the second periodic sound (produced by stretching the waveform of the first) is given by the ordinate. For further details, see text.

Fig. 2. Sound spectrograms of mixtures of 99.5 Hz and 100 Hz complex tones. The top spectrogram is based upon congruent waveforms (both tones are derived from the same 10 ms segment of Gaussian noise, with the 99.5 Hz produced by a 0.5% "stretching" of the waveform). The bottom spectrogram is based upon uncorrelated waveforms (independent segments of Gaussian noise). Waveform alignment occurs only for the top spectrogram.

APPENDIX C

AFOSR-TR- 88-1160

Illusory continuity of tonal and infratonal periodic sounds

by

Richard M. Warren

John M. Wrightson[a]

Julian Puretz

University of Wisconsin-Milwaukee
Department of Psychology
P. O. Box 413
Milwaukee, Wisconsin 53201

Received:

---

[a] Current address: Joiner-Rose Group, 4125 Centurion Way, Dallas, TX 75244

ABSTRACT: Temporal induction can restore masked or obliterated portions of signals, so that tones may seem continuous when alternated with sounds having appropriate spectral composition and intensity. The upper intensity limits for the induction of tones (pulsation thresholds) are related to masking functions, and have been used to define the characteristics of frequency domain (place) analysis of tones. The present study has found that induction also occurs for infratonal periodic sounds which require a time domain analysis for perception of acoustic repetition. Limits for temporal induction were determined for iterated frozen noise segments from 10 Hz through 2000 Hz alternated with a louder on-line noise. Masked thresholds were also obtained for the pulsed signals presented along with continuous noise, and it was found that the relation between induction limits and masking changed with frequency. The results obtained for induction and masking are discussed in terms of general principles governing restoration of obliterated sounds.

INTRODUCTION

When portions of signals are replaced by louder sounds, listeners may believe they hear the missing fragments. The limiting conditions for these perceptual restorations have been studied in recent years as a source of information concerning auditory mechanisms.

Miller and Licklider (1950) appear to have published the first report of illusory continuity of signals interrupted by noise. They found that when two sounds differing in intensity and quality are alternated in a regular fashion, the fainter sound may seem to remain on continuously. Thurlow's (1957) rediscovery of this effect led to a number of subsequent studies (Thurlow & Elfner, 1959; Thurlow & Marten, 1962; Elfner & Caskey, 1965; Elfner & Homick, 1966, 1967; Elfner, 1969, 1971).

Houtgast (1972) and Warren, Obusek, and Ackroff (1972) independently proposed rules considering the apparent continuity of the fainter of two alternating sounds as the inverse of masking. Houtgast's rule describing what he called the *pulsation threshold* for tones was: "When a tone and a stimulus *S* are alternated (alternation cycle about 4 Hz), the tone is perceived as being continuous when the transition from *S* to tone causes no (perceptible) increase of nervous activity in any frequency region. The pulsation threshold, thus, is the highest level of the tone at which this condition still holds." This rule has been used to infer the characteristics of spectral filtering at the basilar membrane, with experimental findings being interpreted in terms of both topographical excitation patterns produced by tones and lateral suppression at loci contiguous to the stimulated regions (see, for example, Houtgast, 1974; Aldrich & Barry, 1980; Shannon & Houtgast, 1986).

The rule for *temporal induction* described by Warren et al. (1972) also requires an overlap of peripheral neural excitation, but is somewhat different in scope, stating that: "If there is contextual evidence that a sound may be present at a given time, and if the peripheral units stimulated by a louder sound include those which would be stimulated by the anticipated fainter sound, then the fainter sound may be heard as present." This principle applies not only to the perceptual restoration of fragments of steady-state tones, but also to the restoration of several types of time-varying signals which include speech, tonal glides, and melodic tonal sequences (see Warren, 1984). The present investigation extends the study of temporal induction to complex sounds having repetition frequencies below the limit of pitch. These long-period sounds are perceived as possessing a repetitive temporal texture or time-varying pattern.

In their investigation of such sounds, Guttman and Julesz (1963) used iterated segments of Gaussian noise (repeated "frozen noises" or RFNs). They described the perceptual quality as a repetitive "whooshing" from 1 Hz (the approximate lower limit of periodicity detection of RFNs) to 4 Hz, and as "motorboating" from 4 Hz through 19 Hz. At 20 Hz and above, RFNs are considered to be complex tones possessing pitch, and were not investigated by them. Warren and Bashford (1981) examined both tonal and infratonal RFNs, and reported that a noisy pitch with a hiss-like quality was heard from 20 Hz up to about 100 Hz, with RFNs of higher frequencies appearing to be completely tonal with no hint of a noisy quality. They noted similarities in the rules governing perception of pitch and infrapitch, and suggested that some of the mechanisms for the detection of acoustic repetition may operate on both sides of the tonal/infratonal boundary. It was suggested that RFNs could serve as useful model stimuli for studying the continuum of detectable acoustic iterance, with observations in the infratonal and the tonal ranges each enhancing understanding of the other (for further discussion, see Warren, 1982, pp. 78-90).

The present study was designed to compare the upper limits of temporal induction for complex tones and for infratonal periodic sounds. Since experiments have demonstrated a close relation between masking and illusory continuity (Houtgast, 1972; Warren, Obusek & Ackroff, 1972), in the present study, both induction limits and various types of masking limits were measured on both sides of the pitch boundary.

I. METHOD

A. Subjects

Four listeners participated in the study. Each was familiar with psychoacoustic experimentation and had served as a subject in other studies of auditory perception.

B. Stimuli

The periodic stimuli consisted of repeated frozen noises (RFNs). The output voltage from a Gaussian noise generator was sampled every 20 μsec and coded in 12-bit form by a digital delay line built to our specifications by the Physical Data Company. The delay was adjusted to correspond to the desired period, and then by closing a "recycle" switch, input to the delay line was rejected, and the signal looped or repeated indefinitely in digital form. Appropriate filtering removed the spectral artifacts associated with digital processing. The RFNs had periods of 100, 50, 20, 10, 5, 2, 1, and 0.5 ms which corresponded to frequencies of 10, 20, 50, 100, 200, 500, 1000, and 2000 Hz, respectively. Two modes of stimulus presentation were used: For one mode the periodic signal was alternated with on-line noise (the signal and the noise were each on for 300 ms); for the other mode the noise was on continuously, and the signal was pulsed (the superimposed signal was on for 300 ms and off for 300 ms). The on-line noise was always delivered at 80 dB SPL, and the intensity of the periodic signal was adjusted to a particular criterion level by the listener. Timing was controlled by a preset counter driven by a Rockland 5100 Frequency Synthesizer, and electronic switches used for alternating the stimuli were set for a linear rise-fall time of 25 ms. For repetition frequencies of 50 Hz and above, both the noise and the RFNs were high-pass filtered at the lowest frequency of the periodic signal (the spectral fundamental) and low-pass filtered at 8,000 Hz, using filters with slopes of 48 dB/octave. At repetition frequencies below the 50 Hz response limit of both the delay line and the headphones, high-pass filtering was maintained at 50 Hz and the low-pass filtering was again 8,000 Hz for both the signal and the on-line noise. The intensity of the RFNs was increased or decreased as desired by the listener by turning the unseen dial of an attenuator having 1 dB steps. Stimuli were presented diotically through matched TDH-49 headphones having a flat response (± 1 dB) from 50-8000 Hz while listeners were seated in an audiometric room having an ambient SPL of 25 dBA.

C. Procedure

Listeners were presented with each of the eight signal frequencies once during an experimental session. At each frequency, they were instructed to make the five different types of judgments described in detail below. For the first three judgments, the RFNs were alternated with 80 dB SPL on-line noise, with each on for 300 ms before switching. For the last two judgments the 80 dB noise was continuous, and the mixed (added) periodic sound was alternately on for 300 ms and off for 300 ms.

Five types of judgments were made in the following order:

1. Continuity/Discontinuity Transition (Upper Limit of Temporal Induction). The intensity of the RFN alternated with noise was adjusted to the lowest level at which it

seemed discontinuous or pulsate (just below this limit, listeners reported hearing continuous iterance which was either pitch or "motorboating").

2. Threshold for Detection of Signal Presence When Alternated With Noise. The RFN was adjusted to the lowest level at which its presence could be detected (i.e., it was noticeably different from silence).

3. Threshold for Detection of Signal Repetition When Alternated With Noise. The RFN was adjusted to the lowest level at which iteration (either pitch or infrapitch repetition) could be heard.

4. Threshold for Detection of Signal Presence When Superimposed Upon Noise. The RFN was adjusted to the lowest level at which its presence could be detected as an intermittent addition to the continuous noise.

5. Threshold for Detection of Signal Repetition When Superimposed Upon Noise. The RFN was adjusted to the lowest level at which iteration (either pitch or infrapitch) could be heard for the intermittent addition to the continuous noise.

There were six experimental sessions. Each session was split into two parts separated by a five-minute rest period. During Part A, listeners were presented with four of the eight RFN frequencies (10, 50, 200, and 1000 Hz) presented in a randomly determined order. The five types of judgments described above were made successively in the order listed for each of the frequencies. Part B was the same as Part A, except that the remaining four repetition stimulus frequencies were employed (20, 100, 500, and 2000 Hz). In the first, third, and fifth sessions, Part A was presented first, followed by Part B. In the second, fourth, and sixth sessions, this order was reversed. By the end of the study, each listener had made six judgments for each of the five types of thresholds with each of the eight iterated noise segment frequencies. It should be noted that each frozen waveform was used for only one session and one listener.

II. RESULTS

The experimental data obtained are summarized in Table 1. It can be seen that induction was greatest (the continuity/discontinuity boundaries were at the highest amplitudes) at infratonal and low tonal repetition frequencies. A one-way analysis of variance with repeated measures yielded a significant effect of frequency ($F(7, 21) = 20.53$, $p < .001$), and subsequent Neuman-Keuls tests indicated that continuity/discontinuity thresholds were higher ($p < .05$ or better) at repetition frequencies from 10 Hz to 100 Hz than at repetition frequencies from 200 Hz to 2000 Hz.

-- Table 1 About Here --

Table 1 also shows that when the RFN was alternated with noise, the threshold for detecting signal repetition was several dB above the threshold for detecting signal presence for the infratonal and low tonal frequencies. Planned orthogonal comparisons (Kirk, 1968, pp. 73-76) indicated that the two thresholds differed reliably ($p < .01$) at repetition frequencies of 10, 20 and 50 Hz. As the intensity level was raised for these low frequencies, the repeated frozen noise (RFN) was heard first only as a faint continuous hiss without detectable iteration: An appreciable increase in amplitude (3-5 dB) above the absolute detection threshold was required before effects attributable to repetition could be heard. The boundary between pitch which seems completely homogeneous and tonal, and pitch with *a noisy or hiss-like quality occurs at about 100 Hz (Warren & Bashford, 1981).* As shown in Table 1, the pitch corresponding to these purely tonal RFNs was detected at signal intensities approximating the absolute detection threshold.

Although simultaneous masking was absent when the periodic sounds were alternated with on-line noise, the possibility of forward and backward masking produced by noise bursts preceding and following the signal needs to be considered. The 300 ms duration of interruptions used for both induction and threshold measurements would be expected to produce only a slight, if any, increase in thresholds in the present study (for a discussion of the limits of forward and backward masking and their interactions, see Elliot, 1971; Wilson & Carhart, 1971). Nevertheless, in order to compensate for any residual masking of this type, the threshold for detection of repetition when the signal was alternated with noise was subtracted from the amplitude corresponding to the continuity/discontinuity transition to obtain the sensation level (SL) at the upper limit of auditory induction for each stimulus frequency. These values were used to construct Figure 1 showing the existence regions for temporal induction (illusory continuity of acoustic repetition) and for pulsation (perception of discontinuity). Listeners' SLs for induction were subjected to an analysis of variance which yielded a significant effect of frequency ($F(7, 21) = 7.08$, $p < .001$). Subsequent Neuman-Keuls tests indicated that the existence region for iterance was diminished ($p < .05$) for repetition frequencies of 1 kHz and 2 kHz.

-- Figure 1 About Here --

The correspondence between the continuity/discontinuity transition and the threshold for detection of repetition under conditions of simultaneous masking is shown in Figure 2. The data for these two thresholds were compared in a two-factor analysis of variance which yielded significant main effects of threshold type ($F(1, 3) = 16.40$, $p < .05$) and repetition frequency ($F(7, 21) = 38.53$, $p < .0001$) and a significant interaction ($F(7, 21) = 4.20$, $p < .005$). Subsequent Neuman-Keuls tests indicated that thresholds for repetition detection under simultaneous masking were higher ($p < .05$ or better) than the continuity/discontinuity transition at corresponding repetition frequencies of 200 Hz and above, but these two measures did not differ reliably at repetition frequencies of 100 Hz and below. Thus, noise was a relatively poor inducer of continuity for purely tonal sounds, in keeping with the data reported by Warren et al. (1972) for 300 ms sinusoidal tones alternated with 300 ms noises.

III. DISCUSSION

Three types of temporal induction have been described: Homophonic, contextual catenation and heterophonic (Warren, 1984). Homophonic induction is the simplest, and its characteristics can facilitate understanding of the others. It occurs when two intensity levels of otherwise identical sounds are alternated, and consists of the apparent continuity of the fainter level. The sounds producing homophonic induction can be periodic (such as two levels of a sinusoidal tone) or non-periodic (such as two levels of a noise)--in each case induction of the fainter occurs at all audible differences for all audible levels. It seems that the segments of the weaker sound occurring before and after each segment of the louder sound cause it (the louder sound) to be factored into two portions. One of these portions corresponds to the level of the fainter sound and provides the bridging continuity, while the residue (the original louder level minus the fainter level) appears as a pulsed addition to the continuous sound. A simple demonstration of this subtractive factoring is provided by the observation that when 80 dB and 82 dB levels of the same noise are alternated and listeners perceive the 80 dB level as continuous, they paradoxically hear the 82 dB level as a pulsed fainter sound (Warren, 1982, p. 141). While the subtractive nature of induction is not as obvious when the inducer and inducee are qualitatively different, there is evidence that the two other types of temporal induction also involve a subtractive processing which may reverse the effects of masking (Warren, 1984).

Contextual catenation occurs when a time-varying signal such as speech is interrupted by a louder extraneous sound. When the peripheral neural overlap requirements (discussed earlier) are met, the contextual information provided by the intact segments can lead to perceptual synthesis of fragments differing from the preceding and following portions of the signal. Listeners hear the signal as uninterrupted, and cannot distinguish the restored segments from those physically present. In addition to phonemic restorations (Warren, 1970; Bashford & Warren, 1987), contextual catenation can restore missing notes of a melody played on the piano (Sasaki, 1980) and can synthesize obliterated segments of tonal frequency glides (Dannenbring, 1976; Ciocca & Bregman, 1987).

Heterophonic continuity refers to the apparent lack of interruption of a particular sound when replaced by a qualitatively different louder sound which meets the specifications of the peripheral overlap rule. Tones are often employed as the fainter sound, but other periodic sounds can be employed. The iteration of nonsinusoidal waveforms can be detected at infratonal frequencies (below 20 Hz), and the present study has examined the induction of tonal and infratonal repeated frozen noises over a range extending from 10 Hz through 2000 Hz.

Let us compare the auditory mechanisms employed for the detection of repetition in the tonal and infratonal ranges, and their relevance to the observations made in the present study. In the infratonal range, perception of frozen noise repetition is based upon the iteration of neural response patterns. This temporal information is available at all loci on the basilar membrane, for when a 1/3-octave band-pass filter (approximating a critical band) is swept through the audible range, then an infrapitch repetition (attributable to the interaction of unresolved harmonics within a critical band) can be heard at all center frequencies of the filter (Warren & Bashford, 1981). As the RFN frequency is raised into the tonal range, then individual lower harmonics can be resolved along the basilar membrane (see Plomp, 1964 for the limits of spectral resolution), and two additional neural correlates of RFN repetition appear along with the iterated neural patterns corresponding to the unresolved higher harmonics (Warren, 1982, pp. 82-85). The resolved harmonics can provide spectral information concerning RFN repetition frequency through the positioning of stimulation maxima on the basilar membrane, and may also provide temporal information based upon the phase-locking of nerve fiber responses (for a discussion of place cues and phase-locked cues to the pitch of complex tones, see de Boer, 1976, and Evans, 1978). It appears that once the peripheral overlap rule is satisfied, then gaps in the cues to repetition do not interfere with the apparent continuity of repeated frozen noises: The perceptual synthesis of RFNs restores all of the qualitative attributes of repetition.

As can be seen in Figure 1, illusory continuity of iterance occurred at higher sensation levels for repetition frequencies from 10 through 100 Hz than for frequencies from

200 through 2000 Hz. This change in induction limits was both monotonic and gradual, and not related in any direct fashion to the pitch/infrapitch transition at 20 Hz. The lower pulsation thresholds at higher frequencies may be attributable to the increase in spacing between harmonic components. This greater frequency separation enhances spectral resolution, and concentrates stimulation at those neurons with characteristic frequencies close to the resolved harmonics. The concentration of spectral power at discrete loci would necessitate a drop in level of a tonal RFN in order for the 80 dB noise bursts to satisfy induction's peripheral overlap rule.

Figure 2 shows that the transition from induction to pulsation of an RFN remained close to the simultaneous masking threshold for infratonal and low tonal frequencies. The transition of RFNs from noisy tones to smooth, homogeneous tones occurs at about 100 Hz (Warren & Bashford, 1981), and it can be seen that the pulsation thresholds diverged from masked thresholds above that repetition frequency. A similar separation of masking and pulsation limits for pure tones induced by noise has been reported by Warren, Obusek and Ackroff (1972). They alternated 300 ms bursts of tones and noises of various spectral compositions and found that induction limits were 10 dB or more below the simultaneous masking limits.

Why do pulsation thresholds diverge from simultaneous masking thresholds for tonal induction by noise? One possible explanation starts by considering that pulsation thresholds represent the lower limit for detecting signal <u>absence</u> in noise. Noises are characterized by rapid changes in amplitude which produce rapidly fluctuating levels of neuronal stimulation. If, when the noise is present, neurons that had been responding previously to a steady tone exhibit momentary dips below the activity levels corresponding to the tone, then absence of that tone is signaled and induction is blocked. Hence, pulsation thresholds of the tones were not determined by the average SPL of the noise (as appears to be the case for the simultaneous masking of tone by noise), but rather by transitory minima in the noise power spectrum. When the inducee was a periodic sound which itself had a noise-like quality (RFNs up to 100 Hz), it appears that the brief dips in the amplitude of the on-line noise inducer did not block induction, and that the average sound pressure levels of the two fluctuating sounds determined both induction limits and simultaneous masking limits.

References

Aldrich, W. M., & Barry, S. J. (1980) "Critical bandwidths measured by the pulsation-threshold technique," J. Aud. Res. 20, 137-141.

Bashford, J. A., Jr., & Warren, R. M. (1987) "Multiple phonemic restorations follow the rules of auditory induction," Perception & Psychophysics, in press.

Ciocca, V., & Bregman, A. S. (1987) "Perceived continuity of gliding and steady-state tones through interrupting noise," Perception & Psychophysics 42, 476-484.

Dannenbring, G. L. (1976) "Perceived auditory continuity with alternately rising and falling frequency transitions," Can. J. Psychol. 30, 99-114.

de Boer, E. (1976) "On the 'residue' and auditory pitch perception," in Handbook of Sensory Physiology, Vol. V, Auditory System, Part 3: Clinical and Special Topics, edited by W. D. Keidel and W. D. Neff, (Springer-Verlag, Berlin), pp. 479-583.

Elfner, L. F. (1969) "Continuity in alternately sounded tone and noise signals in a free field," J. Acoust. Soc. Am. 46, 914-917.

Elfner, L. F. (1971) "Continuity in alternately sounded tonal signals in a free field," J. Acoust. Soc. Am. 49, 447-449.

Elfner, L. F., & Caskey, W. E. (1965) "Continuity effects with alternately sounded noise and tone signals as a function of manner of presentation," J. Acoust. Soc. Am. 38, 543-547.

Elfner, L. F., & Homick, J. L. (1966) "Some factors affecting the perception of continuity in alternately sounded tone and noise signals," J. Acoust. Soc. Am. 40, 27-31.

Elfner, L. F., & Homick, J. L. (1967) "Auditory continuity effects as a function of the duration and temporal location of the interpolated signal," J. Acoust. Soc. Am. 42, 576-579.

Elliot, L. L. (1971) "Backward and forward masking," Audiology 10, 65-76.

Evans, E. F. (1978) "Place and time coding of frequency in the peripheral auditory system: Some physiological pros and cons," Audiology 17, 369-420.

Guttman, N., & Julesz, B. (1963) "Lower limit of auditory periodicity analysis," J. Acoust. Soc. Am. 35, 610.

Houtgast, T. (1972) "Psychophysical evidence for lateral inhibition in hearing," J. Acoust. Soc. Am. 51, 1885-1894.

Houtgast, T. (1974) "Masking patterns and lateral inhibition," in Facts and Models in Hearing, edited by E. Zwicker and E. Terhardt (Springer-Verlag, Berlin), pp. 258-265.

Kirk, R. E. (1968) Experimental Design: Procedures for the Behavioral Sciences, (Brooks/Cole, Monterey, California).

Miller, G. A., & Licklider, J. C. R. (1950) "The intelligibility of interrupted speech," J. Acoust. Soc. Am. 22, 167-173.

Plomp, R. (1964) "The ear as a frequency analyzer," J. Acoust. Soc. Am. 36, 1628-1636.

Sasaki, T. (1980) "Sound restoration and temporal localization of noise in speech and music sounds," Tohoku Psychologica Folia 39, 79-88.

Shannon, R. V., & Houtgast, T. (1986) "Growth of pulsation threshold of a suppressed tone as a function of its level," Hear. Res. 21, 251-255.

Thurlow, W. R. (1957) "An auditory figure-ground effect," Am. J. Psychol. 70, 653-654.

Thurlow, W. R., & Elfner, L. F. (1959) "Continuity effects with alternately sounding tones," J. Acoust. Soc. Am. 31, 1337-1339.

Thurlow, W. R., & Marten, A. E. (1962) "Perception of steady and intermittent sound with alternating noise-burst stimuli," J. Acoust. Soc. Am. 34, 1853-1858.

Warren, R. M. (1970) "Perceptual restoration of missing speech sounds," Science 167, 392-393.

Warren, R. M. (1982) Auditory Perception: A New Synthesis, (Pergamon, Elmsford, New York).

Warren, R. M. (1984) "Perceptual restoration of obliterated sounds," Psychol. Bull. 96, 371-383.

Warren, R. M., & Bashford, J. A., Jr. (1981) "Perception of acoustic iterance: Pitch and infrapitch," Perception & Psychophysics 29, 395-402.

Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972) "Auditory induction: Perceptual synthesis of absent sounds," Science 176, 1149-1151.

Wilson, R. H., & Carhart, R. (1971) "Forward and backward masking: Interactions and additivity," J. Acoust. Soc. Am. 49, 1254-1263.

Acknowledgment

The support of the Air Force Office of Scientific Research, Air Force Systems Command, USAF (grant agreement number AFOSR 85-0260) is gratefully acknowledged. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. This manuscript is submitted for publication with the understanding that the US Government is authorized to reproduce and distribute reprints for Governmental purposes.

We thank James A. Bashford, Jr. and Bradley S. Brubaker for their valuable contributions.

Table Caption

Table 1. Different types of thresholds for signals consisting of 300 ms bursts of iterated noise segments when alternated with, or superimposed upon, on-line noise at 80 dB SPL. Means and Standard Error (SE) of means are in dB SPL and represent 24 judgments (6 from each of 4 subjects). For further details, see text.

REPETITION FREQUENCY OF ITERATED NOISE SEGMENT (Hz)

	10	20	50	100	200	500	1000	2000
Continuity/Discontinuity Transition (Alternation) [SPL]								
Mean	73.38	71.79	68.08	66.00	59.79	57.58	53.67	53.63
SE	0.49	0.93	0.85	0.89	1.65	1.88	1.49	1.78
Detection of Signal Presence (Alternation) [SPL]								
Mean	31.21	30.50	29.50	30.25	28.46	28.88	28.00	30.46
SE	0.80	0.90	0.81	0.70	0.75	0.85	0.73	0.98
Detection of Signal Pitch or "Motorboating" (Alternation) [SPL]								
Mean	36.58	34.88	32.29	31.71	29.29	29.71	28.13	31.13
SE	1.34	1.28	1.08	0.82	0.79	0.83	0.65	0.95
Detection of Signal Presence (Simultaneous) [SPL]								
Mean	70.54	70.58	69.67	69.46	68.21	66.63	65.08	64.75
SE	0.44	0.55	0.51	0.56	0.47	0.55	0.81	0.57
Detection of Signal Pitch or "Motorboating" (Simultaneous) [SPL]								
Mean	75.58	75.29	72.21	70.92	68.63	66.79	65.50	65.54
SE	0.45	0.57	0.88	0.56	0.60	0.66	0.82	0.66

Figure Captions

Figure 1. The discontinuity/continuity boundary in sensation level (dB above repetition detection threshold) for frozen noise segments with different repetition frequencies when alternated each 300 ms with 80 dB SPL on-line noise. For further description, see text.

Figure 2. Comparison of the continuity/discontinuity boundary (upper limit of temporal induction) with the signal threshold under simultaneous masking. The continuity/discontinuity threshold is for the iterated signal when alternated each 300 ms with an 80 dB broadband noise, and the masked threshold is for the detection of signal repetition when added intermittently (on 300 ms and off 300 ms) to a continuous 80 dB broadband noise. All values are in dB SPL. For further details, see text.

APPENDIX D

AFOSR-TR- 88-1160

Illusory continuity of interrupted speech:
Speech rate determines durational limits

James A. Bashford, Jr., Mark D. Meyers,
Bradley S. Brubaker and Richard M. Warren
University of Wisconsin-Milwaukee
Department of Psychology
Milwaukee, Wisconsin 53201

(received

RUNNING TITLE: Multiple Phonemic Restorations

ABSTRACT: Deleted segments of speech can be restored perceptually if they are replaced by a louder noise. An earlier study of this "phonemic restoration effect" found that when recorded discourse was interrupted periodically by noise, the durational limit for illusory continuity corresponded to the average word duration. The present study employed a different passage of discourse recorded by a different speaker. Durational limits for apparent continuity of discourse interrupted by noise were measured at the normal (original) playback speed as well as at rates which were 15% greater and 15% less. At the normal playback rate, once again the limit of continuity approximated the average word duration--but of especial interest was the finding that changes in playback rate produced proportional changes in continuity limits. These results, together with other evidence, suggest that phonemic restorations represent a special linguistic application of a general auditory mechanism (auditory induction) producing appropriate syntheses of obliterated sounds, and that for discourse the limits of illusory continuity correspond to a fixed amount of verbal information, and not a fixed temporal value.

INTRODUCTION

When portions of an acoustic signal (either verbal or nonverbal) are removed and replaced by a louder extraneous sound, the fragments are restored perceptually if certain conditions are met. This "continuity effect," also known as "auditory induction," requires that deleted portions of the signal be replaced by a potential masker (Bashford & Warren, 1987; Houtgast, 1972; Verschuure, 1978; Warren, Obusek & Ackroff, 1972). Under these conditions, illusory continuity may persist through noise filled gaps lasting several hundreds of ms or more, and may involve either the simple continuation of steady-state signals such as tones, or the reconstruction of portions of time-varying signals such as the "phonemic restoration" of interrupted speech (Warren, & Obusek, 1971). For a review of the literature and theory encompassing both verbal and nonverbal auditory induction see Warren (1984).

Bashford and Warren (1987) conducted two experiments in which listeners were presented with speech interrupted by louder noise and were required to adjust the duration of periodic gaps to their thresholds for detecting speech deletion (the upper limit of phonemic restoration). In one experiment, recorded discourse (a passage from an article in a popular magazine) was band-pass filtered (remaining intelligible) and then interrupted by silence or by a band-pass filtered noise. When the discourse was interrupted by silence, the threshold for detection of gaps averaged about 75 ms. However, when the speech band was interrupted by a louder band of noise having the same center frequency (1.5 kHz) and a slightly greater bandwidth, the threshold gap duration increased dramatically to 304 ms, a value almost exactly equal to the average word duration in the passage (306 ms discounting pause time). Further, the differential efficacy in producing phonemic restorations for other noise bands having different center frequencies paralleled their potential for masking the speech signal.

In a second experiment, Bashford and Warren presented listeners with broadband speech of three types: 1) An unfiltered version of the discourse passage used in the first experiment, 2) the same discourse passage presented at the same word rate but read with the order of words reversed (intonation and phrasing approximated that of discourse), and 3) lists of isolated monosyllabic words. Threshold gap durations were equivalent (about 50 ms) for each of the three types of speech when the signals were interrupted by silence. However, there was a differential increase in threshold durations when gaps in the stimuli were filled with a broadband noise matching the spectra of the speech signals and having a greater amplitude. Threshold gap durations for isolated monosyllables and for the discourse passage read with backward word order both increased by about 100 ms when gaps in these stimuli were filled with noise. In marked contrast, when noise was added to gaps in the normal reading of the discourse passage, continuity threshold durations increased about 250 ms above the value found for silence, and once again (as in the first experiment using filtered discourse with spectrally matched interpolated noise), the threshold gap duration was almost exactly equal to the average word duration.

The manipulation of linguistic context in the study by Bashford and Warren produced substantial variations in thresholds for discontinuity. These findings led the investigators to suggest that the durational limits for induction may provide a sensitive measure of the effect of context upon the size of linguistic chunks employed in the perceptual organization of speech. Of especial interest to the present study was the observation that the upper limit for induction with discourse was equivalent to the average word duration. However, even though it appears clear that context does influence the size of speech fragments subject to restoration, it is possible that the close correspondence found between the upper limit of auditory induction and the average duration of words was fortuitous, and the consequence of a general durational limit for illusory continuity of discourse. The present study was designed to determine whether the upper limit for continuity of discourse has fixed temporal constraints or varies with the rate of delivery (and hence the duration of linguistic segments). A different discourse passage was recorded by a different speaker and played back at three rates. In one condition, the average word duration was approximately the same as in the earlier study. In the remaining two conditions, durations of components within the passage were expanded or compressed by 15% to determine whether the durational limit of induction would covary with signal rate.

I. METHOD

A. Subjects

The forty subjects (18 men and 22 women) were enrolled in the Introductory Psychology course at the University of Wisconsin-Milwaukee and were either given course credit or paid for their participation in the study. They were selected from a larger pool of listeners on the basis of an audiometric screening task described in the procedure section.

B. Stimuli

The speech stimulus was a passage from the United States Constitution. The reading was produced in a sound attenuating chamber (IAC Series 400 A) by a male speaker having a General American dialect. The passage was initially recorded using a Sony model F-98 cardioid microphone and a Sony Model TC-40 cassette recorder which was equipped with an automatic gain control. This initial recording was then band-pass filtered from 200 to 5000 Hz with slopes of 48 dB/octave (Rockland model 852 filter) and then rerecorded at three different tape speeds on separate tracks of an Ampex 440-C 8-track recorder equipped with a continuously adjustable speed control. One version of the passage was recorded at the same tape speed used for playback (7½ ips). The two remaining versions of the passage were recorded with tape speed altered so that, upon playback at 7½ ips, one version was heard at a rate 15% greater and the other at a rate 15% less than that of the original recording. The speech sounded normal at each of the three rates, but differences in playback speed produced differences in the spectra of the stimuli. In order to provide spectrally matched noise for each stimulus, pink noise (that is, noise with equal power per octave which approximates the long-term average spectrum of speech) was subjected to the identical band-pass filtering employed for the original recording of the speech (200-5000 Hz) and then recorded on separate tracks of the multitrack recorder at three different speeds (7½ ips, 7½ ips + 15%, and 7½ ips - 15%).

When presented at its original rate, the discourse passage lasted 39 minutes, and had an overall word rate of 185 wpm. The percentage of pause time in the passage was 8.6% as determined through measurements of amplitude-level tracings (Brüel and Kjaer model 2305 graphic level recorder with pen speed of 4 mm/s and paper speed of 1 mm/s). The average word duration (with pause time discounted) was calculated to be 296 ms for the normal version of the passage, 252 ms for the accelerated version of the passage, and 340 ms for the decelerated version. Amplitude fluctuations were also determined graphically for each speech recording (pen speed 4 mm/s, paper speed .3 mm/s) and were found to be equivalent for corresponding portions of the passage at each signal rate.

C. Apparatus

The six signals recorded on the multitrack recorder (three of which were speech and three noise as described above) were fed to separate subchannels of a Yamaha PM-430 8-

channel mixer. The desired speech signal and its matching noise band were passed from separate master outputs of the mixer to individual electronic switches (Grason-Stadler model 1287-B). The two switches were set for 10 ms rise/fall and were triggered alternately, with a 50% duty cycle, by pulses from a Grason-Stadler model 1219 sequence counter. The sequence counter was driven by a Grason-Stadler model 1270 level zone detector which produced logic pulses at a rate determined by the square wave input from a Wavetek model 135 function generator. During the experiment, this generator, with its dial hidden from view, was adjusted by listeners to vary the rate (duration) of speech interruption. The control knob produced a linear change in interruption rate with turning angle over a range of 0.40 to 25 interruptions per second (ips). The corresponding durations of speech off-time and on-time during each cycle of interruption, ranged from 1250 ms to 20 ms as measured with an accuracy of 0.01 ms by a Hewlett-Packard 5321-A frequency counter. The alternately gated signals from the two electronic switches were combined with a Grason-Stadler model 1292 passive mixer, passed through an impedance-matching transformer (Grason-Stadler model E10589A) and finally transduced diotically through a matched pair of Telephonics TDH-49 headphones mounted in MX 41/AR cushions. The stimuli were presented at an average amplitude of 62 dBC for speech and 72 dBC for noise as measured with a Brüel and Kjaer model 2204 sound level meter equipped with a 6 cc earphone coupler and operating in slow response mode.

D. Audiometric Screening

At least one day prior to participation in the formal experiment, listeners were screened individually in an IAC single-walled sound attenuating chamber. A Békésy-type tracking procedure was used with a diotically presented sinusoidal tone changing from 500 Hz to

8 kHz in alternately ascending and descending frequency sweeps of one octave per minute. Subjects tracked their thresholds by pressing and releasing a remote control switch for the audiometer (Grason-Stadler model E-800), which produced a decrease or increase in tonal intensity at a rate of 2.5 dB/s. Listeners having threshold tracings deviating by more than 22.5 dB from normal at any frequency were not included in the formal experiment. Under these criteria for screening (which were chosen to exclude not only listeners with hearing impairments, but also those who failed to follow the standard audiometric instructions for threshold tracking), approximately 50% of the listeners qualified for further participation in the study.

E. Procedure

Subjects were told that they would be listening to passages from the U.S. Constitution, and that their task would be to adjust a dial to the point where interruptions of the voice became clearly detectable. After the experimenter presented them with samples of discourse interrupted by silence at both the longest (1.25 s) and shortest (20 ms) durations available through turning of the control dial, the subjects were allowed to briefly explore the effects of different interruption rates by turning the control dial themselves. They were then permitted to make two practice adjustments for normal rate discourse interrupted by silence and by noise. Prior to each threshold adjustment, the control knob was set to produce the highest interruption rate of 25 ips (interruptions of 20 ms). Each listener made a total of 18 formal threshold adjustments, with six adjustments made at each playback rate in a separate block of trials. The order in which signal rates were presented was original, slow, and fast for half of the listeners, and was original, fast, and slow for the remaining listeners. Within each block, adjustments were made alternately with silence and noise as interrupters, beginning with interpolated silence. Listeners were given as much time as needed to make their threshold adjustments. The average duration of an experimental session, including instruction and debriefing, was approximately 25 min.

II. RESULTS AND DISCUSSION

The median off-time for a listener's three judgments of the lower limit of speech discontinuity was considered the deletion detection threshold for each condition. The means of those median off-times are presented in Table 1 for interruption by silence and by noise for the three speech rates employed. A two-way analysis of variance for repeated measures yielded significant main effects of interrupter ([$F = 61.64$, $p < .001$] and speech rate [$F = 27.06$, $p < .001$], and a significant interaction [$F = 15.68$, $p < .001$]. Subsequent Tukey tests indicated that thresholds were significantly higher ($p < .01$) at each signal rate when speech was interrupted by noise rather than silence. Thresholds also differed across playback rates ($p < .01$ for all comparisons) when the speech stimuli were interrupted by noise. By Tukey tests, thresholds did not differ across playback rates when the speech stimuli were interrupted by silence. However, Dunnett comparisons for the silent gap conditions did indicate ($p < .01$) that interruption thresholds were higher at the decreased playback rate than at the remaining rates.[1]

-- Table 1 about here --

The interpolation of noise rather than silence in the speech-free portions of the switching cycle produced an increase in threshold gap durations which ranged from 137 ms at the most rapid speech rate to about 199 ms at the slowest rate. The resulting upper durational limits of discourse continuity with interpolated noise (ranging from about 230 ms to about 313 ms off-time, depending on signal rate) are similar to the durational limits previously observed by Bashford and Warren (1987) using the same interruption paradigm but a different discourse passage.

In that earlier study, as mentioned briefly in the introduction, deletion detection thresholds were obtained for a recorded excerpt read from an article appearing in a popular magazine. When regularly spaced gaps in the verbal stimulus were filled with spectrally matched noise, threshold gap durations averaged 304 ms, corresponding to 99% of the average duration of words in that passage (306 ms). Similar effects of interpolated noise were obtained with the passage employed in the present study. When the reading of the Constitution was played back at its original recording speed, threshold off-time with interpolated noise corresponded to 94% of the average word duration. When the speed of playback was increased, so as to temporally compress all components of the signal by 15%, threshold off-time decreased by 17.7%, with the duration of periodic gaps equal to 91% of the average word. In contrast, when the speed of playback was decreased to produce a 15% expansion of the speech signal, the threshold off-time increased 12.2% and equaled 92% of the average word duration. The average percentage of shift in threshold durations produced by a 15% rate change, disregarding the direction of change, was 14.95%.

Thus, as measured in the present study and in the earlier experiments of Bashford and Warren (1987), the discontinuity threshold appears to reflect an informational limit rather than a fixed temporal limit for verbal induction of discourse. Because the interruptions employed in these studies were not linked systematically to specific speech components, interpretation of the results in terms of possible sampling limits for perceptual restoration must be considered as statistical. On average, induction through regularly spaced interruptions begins to fail when gaps in the speech signal approximate word length: Under these conditions, listeners would receive only a single fragment of an average length word. However, as mentioned in the introduction, Bashford and Warren found that threshold gap durations dropped from about 300 ms for a normal reading of discourse to about 150 ms (half the average word duration) for the same discourse passage read at the same rate but with the order of words reversed, and a similarly low threshold duration was obtained for isolated monosyllables. Thus, it appears that the possible "world-length" limit for phonemic restoration does not apply when suprasegmental syntactic and semantic context is absent.

Acknowledgment

The support of the Air Force Office of Scientific Research, Air Force Systems Command, USAF (grant agreement number AFOSR 85-0260) is gratefully acknowledged. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. This manuscript is submitted for publication with the understanding that the US Government is authorized to reproduce and distribute reprints for Governmental purposes.

References

Bashford, J. A., Jr. & Warren, R. M. (1987). Multiple phonemic restorations follow the rules for auditory induction. _Perception & Psychophysics_, _42_, 114-121.

Houtgast, T. (1972). Psychophysical evidence for lateral inhibition in hearing. _Journal of the Acoustical Society of America_, _51_, 1885-1894.

Miller, G. A. & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. _Journal of the Acoustical Society of America_, _22_, 167-173.

Verschuure, J. (1978). _Auditory excitation patterns_ (Doctoral dissertation, Erasmus University Rotterdam). Delft: W. D. Meinema.

Warren, R. M., & Obusek, C. J. (1971). Speech perception and phonemic restorations. _Perception & Psychophysics_, _9_, 358-362.

Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. _Science_, _176_, 1149-1151.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. _Psychological Bulletin_, _96_, 371-383.

FOOTNOTE

[1] In the earlier experiments of Bashford and Warren (1987), interruption thresholds varied dramatically for different types of speech when gaps were filled with noise, but were equivalent across stimuli (threshold gap duration about 50 ms) when interpolated silence was employed. Interruption thresholds in the silent gap conditions of the present study were higher than in those previous experiments and also appear to have been influenced to some extent by temporal properties of the speech signals. This difference in results is probably attributable to a change in the instructions given listeners in the present study. Participants in the earlier experiments experienced greater difficulty making threshold judgments with interpolated silence than with interpolated noise. Speech signals interrupted by noise typically appear to be both continuous and natural up to a listener's threshold off-time, and beyond that threshold duration, detectable gaps appear relatively large. In contrast, speech interrupted by silence appears unnatural at all interruption rates. Even with very brief silent gaps, the speech signal has a "rough" or "bubbly" quality and listeners may spend considerable time attempting to make judgments within the fairly wide range of rapid interruption rates producing that effect. In the present study, an attempt was made to simplify the judgments required of listeners: They were instructed to base their adjustments upon the production of detectable gaps, and to avoid judgments based on roughness. As anticipated, threshold gap durations with interpolated silence were greater, while thresholds for interruption with interpolated noise appear to have been unaffected by this change of instructions.

Table 1. Deletion detection thresholds (in ms off-time) for connected discourse at three playback rates. Changes in thresholds from values obtained at the normal rate are given as Δ%.

	Playback rate				
Interrupter	Normal	Increased 15%		Decreased 15%	
	Mean	Mean	(Δ%)	Mean	(Δ%)
Noise	278.7	229.4	(-17.7)	312.7	(+12.2)
Silence	99.3	92.0	(-7.4)	114.0	(+14.8)

APPENDIX E

AFOSR-TR- 88-1160

Ear Advantages for Monaural Periodicity Detection

by

Richard M. Warren and James A. Bashford, Jr.

Department of Psychology
University of Wisconsin-Milwaukee
Milwaukee, Wisconsin 53201 U.S.A.

ABSTRACT

Monaural asymmetries were found for periodicity detection using repeated 200 msec segments of Gaussian noise (repetition frequency of 5 Hz). An overall left ear advantage was found for monaural delivery with contralateral silence. When lateralization of the monaural signal was abolished by sinultaneous presentation of on-line noise to the opposite ear (contralateral induction caused the signal to be heard as centered on the medial plane), ear advantages were still obtained despite the elimination of the possibility of attentional biases favoring one of the sides. Evidence is presented suggesting that asymmetries in active subcortical processing of periodicity information may be responsible for the ear advantages observed.

Key Words: Ear Advantages
Periodicity Detection

Ear advantages in the identification of sounds have been studied using speech, musical phrases, familiar environmental signals, and pure tones (for a listing and brief summary of the vast literature see [1]. Typically, two clear competing sounds of the same type (i.e., both speech or both music) are presented to opposite ears, and it is determined which can be identified with greater accuracy. The ear advantages obtained usually have been attributed either to favored access to the hemisphere with major responsibility for processing the particular type of signal [8,9,17,19],or to attentional biases or processing strategies favoring sounds heard on one side [3,10,12].

The present study reports ear advantages involving a novel task (detection of infratonal periodicity) using novel stimuli (200 msec segments of Gaussian noise repeated at 5 Hz). Unlike ear differences for identification of familiar sounds, the ear advantages observed were quite pronounced both with and without contralateral competition. As we shall see, there is some evidence that ear advantages in the perception of periodicity might reflect lateral asymmetries in subcortical processing.

Let us look more closely at the nature of the task and the stimuli employed in this study. Guttman and Julesz [6] had reported that iteration of segments of Gaussian noise having durations as long as 1 sec (repetition frequencies of 1 Hz) could be detected readily by listeners. Such infratonal repetition was described as sounding like "whooshing" from 1 Hz through 4 Hz, and like "motorboating" from 4 Hz through 20 Hz. The detection of this infratonal repetition is of necessity based upon the ability to recognize iteration of temporal patterns of neural stimulation since, as will be discussed later, place cues to repetition frequency are unavailable at these frequencies.

In preliminary experiments we presented listeners having normal audiograms for each ear with a monaural 5 Hz iterated noise segment (silence at the opposite ear). When the signal was switched from one ear to the other, some listeners heard the repetition much more clearly on one side than the other, and would not believe that the same repeated sound was delivered to each ear. It was necessary to keep the same 5 Hz signal playing through the same headphone and allow them to turn the headset around and listen to that headphone with each ear in order to convince them that they were observing ear differences, and not signal differences.

Part A of the present study attempted to determine the extent and nature of ear advantages for the perception of repetition of a randomly generated waveform repeated at 5 Hz when present monaurally with silence at the contralateral ear. In Part B, noise (nonrepeated) was used rather than silence at the ear opposite the one receiving the repeated signal. This noise not only produced contralateral competition, but also caused the iterated sound to be heard at a position centered on the medial plane through a process called "contralateral induction" [20]. If ear differences could be observed when the monaural 5 Hz signal was perceived to be at a central rather then lateral position, then (contrary to some suggestions in the literature) attentional biases favoring one side over the other are not required for ear advantages.

METHOD

While initial observations had indicated that the iteration of a 200 msec segment of broadband noise repeated at 5 Hz could be heard much more clearly by some listeners when delivered to one ear than the other, it still was possible for these listeners to detect repetition at either ear. By mixing on-line broadband noise with the repeated signal and presenting the mixture monaurally, the difficulty of detecting repetition could be increased so that it could be heard only by the favored ear. After preliminary experiments with a pan-pot (which kept sound intensity constant while varying proportions of signal and noise) we decided to measure ear advantages in the ability to detect repetition using a mixture in which 85% of the power corresponded to a random waveform repeated at 5 Hz and 15% corresponded to on-line (nonrepeated) noise.

Stimuli. Bandpassed Gaussian (white) noise (100-8000 Hz, 48/octave cut-off slopes) was delivered to an Eventide Model 1745A Digital Delay Line which had a nominally flat response from 50 Hz through 16,000 Hz and a signal-to-internal-noise ratio of 60 dB. When the delay line was placed in recycled mode, a 200 msec segment of noise stored in the shift registers was repeated over and over at 5 Hz. This periodic sound was mixed with the bandpassed white noise (also 100-8000 Hz, 48 dB cut-off slopes) using a Technical Laboratories Pan-Pot (Type DTA 811-PP) which was adjusted so that 85% of the output power consisted of the 5 Hz sound, with the other 15% consisting of noise. A total of 50 signals, each of 5 sec duration, were prepared in this fashion and recorded successively on an Ampex 440C 8-Track Recorder. The 5 Hz signals for use in Part A (monaural stimulation, silence in the contralateral ear) were each recorded on one or the other of two tracks (25 stimuli on each track), so that on playback, one ear received the sound from the recorded track, and the other ear silence from the nonrecorded track. The stimuli for use in Part B were the same as those used for Part A, except that monaural bandpassed noise described above was substituted for the monaural silence of Part A. The sounds for Part B were recorded at the same time as those for Part A using two additional tracks. Ten "catch trials" were used for both Part A and Part B in which bandpassed noise was substituted for the iterated signal. These catch trials were distributed throughout the sequence of experimental stimuli in pseudorandom order, with the restriction that one substitution of the 5 Hz signal by noise occurred on each of the two lateral channels during each block of ten successive stimuli. In Part B, the monaural noise replacing the monaural silence of Part A always was uncorrelated with any noise presented to the opposite ear. The 60 stimuli in Part A (fifty 5 Hz signals, 10 catch trials) were arranged so that the left and right ears were stimulated alternately; Part B employed the identical stimuli except that, in addition, noise at the same SPL as the signal was always present contralaterally.

Subjects. Sixty-four students (36 male and 28 female) from introductory psychology classes served as subjects. All subjects were classified as right-handed on the basis of a questionaire, and had no left-handed siblings or parents. In a preliminary screening session, Békésy audiograms were obtained for each ear, both in order of increasing and in order of decreasing frequency sweeps. Individuals were not used as subjects if they had a tracing for either ear differing by more than 20 dB from audiometric normal for any frequency, or tracings for any frequency differing by more than 15 dB between successive ascending and decending frequency sweeps for the same ear. Those completing the initial screening successfully were invited to return for the second session. As a final criterion for eligibility, individuals participating in the second session were required to have no more than three incorrect responses for the 20 catch trials in Parts A and B in which no iterated signal was present (9 individuals were rejected on this basis before the group of 64 subjects meeting criteria was completed).

Procedure. Stimuli (whether monaural or dichotic) were presented at 80 dB SPL through matched TDH-39 headphones (frequency response flat from 100-8000 Hz) while seated in an audiometric room. Half the subjects started with Part A (monaural stimulation) and half with Part B (dichotic stimulation). All subjects read the same typewritten instructions informing them that we were interested in their perception of a hiss-like noise repeated several times a second. They were told that the repetition would be clearer for some samples than others, and that sometimes there would be no repetition at all. They were instructed to answer yes only if they were certain that they heard repetition, and to answer no otherwise. Subjects were then given 10 practice stimuli to acquaint them with the nature of both the 5 Hz signal and noise: before Part A, the practice stimuli were presented alternately to left and right ears with silence at the other ear; before Part B, the same practice stimuli were used, but in addition, a simultaneous uncorrelated noise was presented contralaterally. All sounds presented as practice stimuli were at the same SPL (80 dB) as sounds in Parts A and B of the formal experiment.

After the samples were heard, the experimenter answered any questions concerning the procedure, and then presented the 60 experimental stimuli. Responses were given during a 5 sec interval separating the successive recorded stimuli. All subjects heard the same sequence of stimuli, but for half, the first signal was delivered to the left ear and for half to the right ear. Systematic reversal of the headphones' positions across subjects prevented any asymmetrical laterality effects within the group attributable to any slight (i.e., unmeasurable) differences in responses of the matched headphones.

RESULTS

The data for the 64 listeners (the number of correct detections of repetition out of the 25 presentations to each ear under each of the two experimental conditions) were collapsed across order of left/right alternation and order of silence/contralateral noise conditions and then subjected to an Ear x Competition x Subjects analysis of variance. This analysis indicated a left ear superiority [$F(1,63) = 10.30$, $p < .005$] and an adverse effect of competition [$F(1,63) = 61.21$, $p < .001$] which did not interact [$F(1,63) = .11$, $p > .05$].

Table 1 summarizes the results of the analysis of variance and shows that, for the combined scores of the group, there is a significant left ear advantage for both the monaural (no competition) condition of Part A, and the dichotic (competition) condition of Part B.

---TABLE 1 ABOUT HERE---

Figure 1 shows that, while the distributions of individual scores in both Parts A and B have modes corresponding to a weak left ear advantage, the distribution in Part B (contralateral noise) is less sharply peaked, with a greater number of both strong left ear and strong right ear advantages, despite the inability of listeners to tell which ear received the repeated monaural signal in this part.

It should be noted that while the pooled data showed highly significant right ear advantages in Part A and Part B, some individual listeners in our group of 64 right-handed subjects showed significant right ear advantages and some showed significant left ear advantages. Analyses of scores of individuals using a Z-test for the significance of a difference between proportions showed that for Part A (contralateral silence), 8 listeners had a significant left ear advantage and 2 a significant right ear advantage; for Part B (competition with contralateral noise), the corresponding numbers were increased to 10 for the left ear and 6 for the right ear. Individuals with significant ear advantage in one of the experimental parts favored the same ear in the other part.

---Fig. 3 About Here---

DISCUSSION

Our study uses stimuli and procedures differing from those employed in other investigations of ear advantages, and the results provide new information concerning lateral differences in auditory processing.

Let us consider first the nature of stimulation by 200 msec segments excised from noise and repeated at 5 Hz. The stimuli had line spectra extending from 100 through 8,000 Hz, with a 5 Hz separation between successive harmonics. The amplitude of individual harmonics was determined randomly, and the phase spectrum was flat. The harmonics were too closely spaced to permit resolution along the basilar membrane [16], and their interactions within critical bands produced patterns of amplitude modulation repeated every 200 msec. While all critical bands had patterns iterated at 5 Hz, the temporal structure of amplitude fluctuation within each critical band was different. As we shall see, there are reasons to believe that the temporal processing necessary to detect such low frequency periodicity involved a low level of the auditory pathway.

Experiments dealing with ear advantages in recognition of sounds usually employ familiar patterns such as speech, melodies, or identifiable environmental sounds. Stimulation is usually dichotic, since it is very difficult to observe ear advantages without competition between different signals of the same type.

The literature dealing with ear advantages is considerable -- Berlin and McNeil [1] list over 300 articles published between 1954 and 1975 in their heroic attempt to summarize and extract essential features of these studies. Differences reported (for example, right ear advantage for speech, left ear advantage for music) have been attributed to asymmetries in cortical processing of stimuli and comparison with the long-term memory traces of related sounds stored at particular hemispheric loci, coupled with a transmission advantage (enhanced by dichotic competition) of contralateral over ipsilateral pathways from ear to cortex [8,9,17,19]. Attempts also have been made to relate ear advantages to laterally directed attention and reporting strategies [3,10,12]. Neither of these approaches considers differences in active subcortical processing to be involved in ear differences -- it is assumed that either information reaching cortical processing centers is degraded less for the input to one of the ears, or that cortically based biases direct greater attention towards inputs heard on one side.

However, there is evidence that processing of considerable complexity occurs along the auditory pathways to the cortex. The 30,000 fibers of the auditory nerve from one cochlea feed into approximately one million subcortical neurons, of which about 90,000 are located in the cochlear nucleus [21]. Since the fibers of the cochlear branch of the auditory nerve terminate at the ipsilateral cochlear nucleus, Møller [15] suggested that it would be an obvious place for decoding of periodicity: he pointed out that the fine structure of temporal information necessary for periodicity analyses would deteriorate through the course of passage through the several neurons constituting the ascending pathways to the cortex. Indeed, he had reported earlier that single units in the cochlear nucleus of rats ipsilateral to a stimulated ear had firing patterns synchronous with a very low frequency amplitude modulation of tones and noise [14]. Similar observations have been made for cats by Hirsch and Gibson [7] who demonstrated that single units in the cochlear nucleus responded synchronously to infrapitch amplitude modulation frequencies of 5 Hz and 10 Hz. If, as this evidence indicates, a temporal analysis of very low frequency patterns of envelope repetition occurs at the cochlear nucleus ipsilateral to the stimulated ear, than lateral asymmetries in processing efficiency at this level of the auditory pathway could be involved in the ear advantages observed in the present study.

The explanation of ear advantages in terms of favored access to the dominant hemisphere considers information flowing in only one direction, from periphery to cortex. But there are feedback loops from the cortex to lower centers extending through the olivocochlear bundle all the way down to the cochlear hair cells [5,18]. Neurophysiological studies with animals other than man have suggested that crossed fibers of the clivocochlear bundle make it possible for activity at higher centers to change the nature of neural processing of clicks and of periodic stimuli at more peripheral levels [2,4], and recent work by Lucas [13] involving auditory brainstem potentials recorded from the vertex in humans has provided evidence suggesting that the olivocochlear bundle in man allows centripetal modulation of peripheral activity. This efferent control of afferent activity blurs the concept of processing hierarchies, and makes the auditory system an integrated unit. It is possible that lateral asymmetries in cortical processing could be associated with asymmetry of efferent control. Thus, while neural information necessary for initial periodicity analysis may not be available at the cortex due to transmission-related loss of temporal fine structure as discussed earlier, cortical asymmetries in final processing and in centripetal control of auditory pathways could give rise to peripheral asymmetries in periodicity analysis, possibly through dominance of crossed efferent pathways. In addition, there may of course, be intrinsic asymmetries in processing of periodicity along the auditory pathways which are not directly dependent upon feedback loops.

Kinsbourne [11] has preferred to explain ear advantages in terms of attentional biases favoring sounds heard on one side over the other. There can be little doubt that attention can be directed to one side or the other when desired, so that when presented with competing speech messages, a voice at one side can be followed while a voice at the other is ignored. Kinsbourne maintained that lateral biases occur even without deliberate direction of attention to one side or the other, and that it is not possible for experiments to eliminate such biases as likely bases for ear advantages. Nevertheless, in Part B, we did rule out the possibility of lateral attentional biases: significant ear advantages were observed despite the loss of lateralization of the monaural signals (the iterated sound delivered to either ear was always perceived as centered on the medial plane, as was the broadband noise delivered to the contralateral ear).

Perhaps there is no single mechanism underlying ear advantages, but rather three: (1) hemispheric asymmetries coupled with transmission advantages for information carried by the pathways from the contralateral ear; (2) laterally directed attentional biases; (3) lateral asymmetries in active processing within subcortical centers and nuclei. Particular experimental tasks and stimuli would determine which of these mechanisms operate.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institutes of Health (grant #2 R01 NS19295-04A3), and by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under grant or cooperative agreement number, AFOSR 85-0260. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. This manuscript is submitted for publication with the understanding that the US Government is authorized to reproduce and distribute reprints for Governmental purposes.

REFERENCES

[1] Berlin, C. I. and McNeil, M. R. (1976): Dichotic listening. In: Contemporary Issues in Experimental Phonetics, pp. 327-387. Editor: N. J. Lass. Academic Press, New York.

[2] Borg, E. (1971): Efferent inhibition of afferent acoustic activity in the unanesthesized rabbit. Experimental Neurology, 31, 301-312.

[3] Broadbent, D. E. (1954): The role of auditory localization in attention and memory span. Journal of Experimental Psychology, 47, 191-196.

[4] Dewson, J. H., III. (1967): Efferent olivocochlear bundle: some relationships to noise masking and to stimulus attenuation. Journal of Neurophysiology, 30, 817-832.

[5] Galambos, R. (1956): Suppression of auditory nerve activity by stimulation of efferent fibers to cochlea. Journal of Neurophysiology, 19, 424-437.

[6] Guttman, N. and Julesz B. (1963): Lower limits of periodicity analysis. Journal of the Acoustical Society of America, 35, 610.

[7] Hirsch, H. R. and Gibson, M. M. (1976): Responses of single units in the cat cochlear nucleus to sinusoidal amplitude modulation of tones and noise: linearity and relation to speech perception. Journal of Neuroscience Research, 2, 337-356.

[8] Kimura, D. (1961): Cerebral dominance and the perception of verbal stimuli. Canadian Journal of Psychology, 15, 166-171.

[9] Kimura, D. (1967): Functional asymmetry of the brain in dichotic listening. Cortex, 3, 163-178.

[10] Kinsbourne, M. (1970): The cerebral basis of lateral asymmetries in attention. In: Attention and Performance III, pp. 192-201. Editor: A. F.

Sanders. North-Holland, Amsterdam.

[11] Kinsbourne, M. (1973): The control of interaction between the cerebral hemispheres. In: Attention and Performance, IV, pp. 239-256. Editor: S. Kornblum. Academic Press, New York.

[12] Lazarus-Mainka, G. and Hörmann, H. (1978): Strategic selection (metacontrol) of hemisphere dominance in normal human subjects. Psychological Research, 40, 15-25.

[13] Lukas, J. H. (1980): Human auditory attention: the olivocochlear bundle may function as a peripheral filter. Psychophysiology, 17, 444-452.

[14] Møller, A. R. (1972): Coding of amplitude and frequency modulated tones in the cochlear nucleus of the rat. Acta Physiologica Scandinavica, 86, 223-238.

[15] Møller, A. R. (1978): Neurophysiological basis of discrimination of speech sounds. Audiology, 17, 1-9.

[16] Plomp, R. (1964): The ear as a frequency analyzer. Journal of the Acoustical Society of America, 36, 1628-1636.

[17] Prohovnik, I. (1978): Cerebral lateralization of psychological processes: a literature review. Archiv für Psychologie, 130, 161-211.

[18] Rasmussen, G. L. (1955): Descending or "feed-back" connections of the auditory system of the cat. American Journal of Physiology, 183, 653.

[19] Studdert-Kennedy, M. (1975): Dichotic studies II: two questions. Brain and Language, 2, 123-130.

[20] Warren, R. M. and Bashford, J. A., Jr. (1976): Auditory contralateral induction: an early stage in binaural processing. Perception & Psychophysics, 20, 380-386.

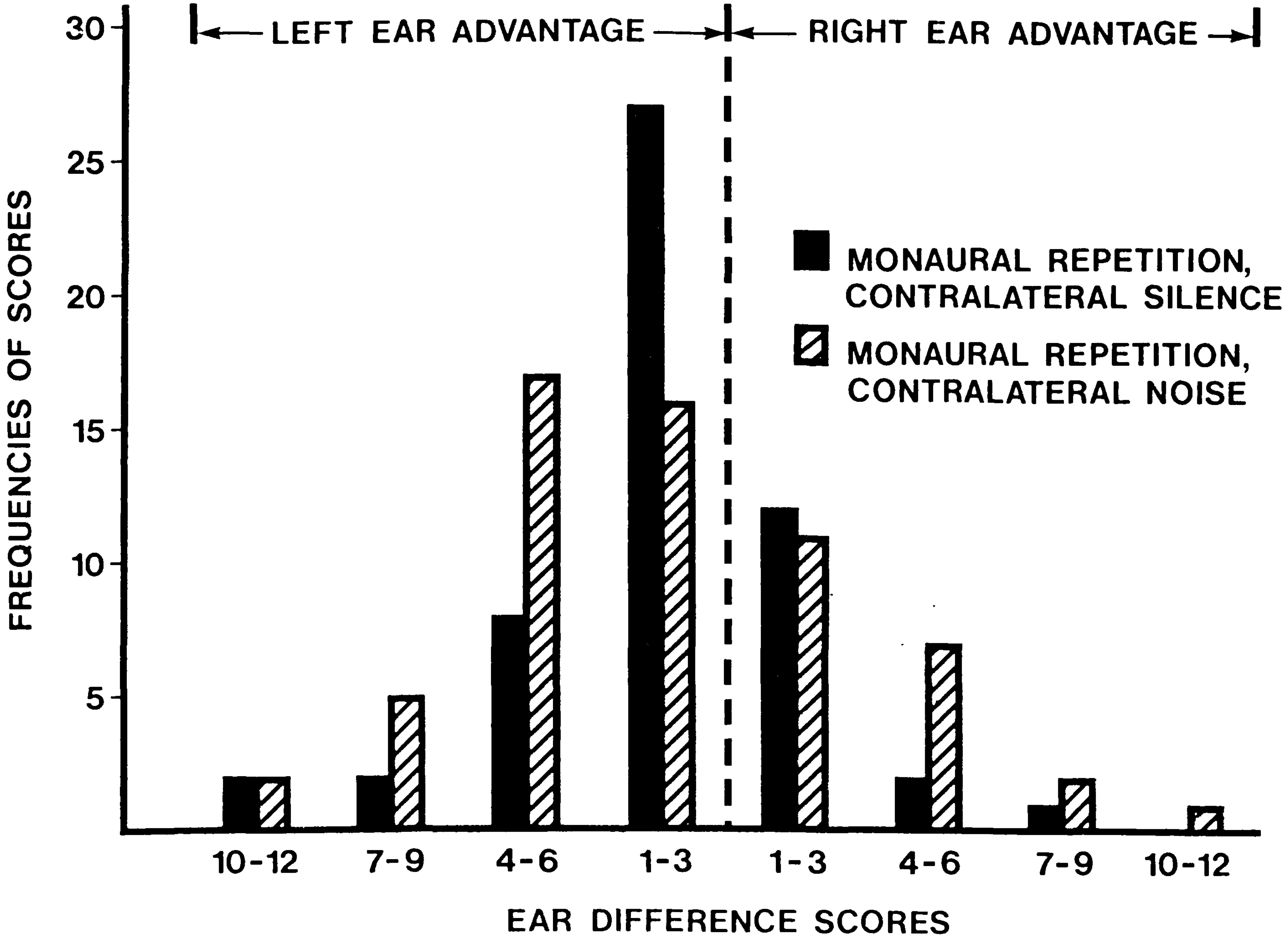
[21] Worden, F. G. (1971): Hearing and the neural detection of acoustic patterns. Behavioral Science, 16, 20-30.

Table 1: Percentage of correct detections of recycling for each ear with and without competing noise

	Left Ear	Right Ear	Left Ear Advantage	p*
No Competition	78.0	70.6	7.4	.005
Competition	60.1	52.9	7.2	.005

*Significance levels based upon an analysis of variance.

Figure 1. Frequency of scores demonstrating ear advantages for individual subjects. Left and right ears were each presented with 20 monaural noise segments repeated at 5 Hz, and the scores represent the number of correct identifications of periodicity for delivery to the ear showing more accurate performance minus the number of correct identifications for the other ear. In the "Contralateral Silence" condition (Part A in the text) there was no dichotic competition. In the "Contralateral Noise" condition (Part B in the text) nonrepeated noise delivered to the opposite ear produced contralateral induction, causing the monaural repetition to be heard as centered on the medial plane.


LEFT EAR ADVANTAGE
RIGHT EAR ADVANTAGE
MONAURAL REPETITION, CONTRALATERAL SILENCE
MONAURAL REPETITION, CONTRALATERAL NOISE
FREQUENCIES OF SCORES
30
25
20
15
10
5
10-12
7-9
4-6
1-3
1-3
4-6
7-9
10-12
EAR DIFFERENCE SCORES

APPENDIX F

AFOSR-TR- 88-1160

Perceptual Bases for the Evolution of Speech

Richard M. Warren
Department of Psychology
University of Wisconsin-Milwaukee
Milwaukee, Wisconsin 53201

*To be published in M. Landsberg (Ed.), The Genesis of Language. Berlin: Walter de Gruyter.

ABSTRACT

It is suggested that speech perception is based upon a holistic recognition of complex acoustic patterns, and does not require the ability to identify individual component sounds. Much confusion in the literature is associated with attempts to consider that speech perception requires the ability to recognize phonemes and their orders at some level of perceptual organization. There is evidence that our ability to recognize acoustic patterns holistically is shared with other animals, and that speech perception evolved from this prelinguistic ability. It appears that identification of component sounds and their orders is a linguistic skill which is the consequence of, not the basis of, speech recognition.

Before we can begin to trace evolution of speech and language, it is necessary to understand the nature of mechanisms used for speech perception. Unfortunately, a pervasive emphasis upon phonemes as linguistic units has impaired our understanding the nature of speech perception and its development from auditory capabilities of our prelinguistic ancestors. This paper will attempt to demonstrate:

1. The concept of phonemes as units of speech can be traced back to the invention of the alphabet.
2. The term "phoneme" as used today has multiple meanings (articulatory, acoustic, perceptual, and graphemic), and the use of the same term for different entities has led to considerable confusion along with inappropriate theories of speech perception.
3. The perception of sound patterns consisting of sequences of several acoustic "phonemes" serves as units of organization in speech perception.
4. Animals other than man are capable of differentiating between complex acoustic sequences including those of speech.
5. The emphasis placed by some theorists upon the lack of speech-producing capabilities of nonhuman primates and other animals may not be directly relevant to an understanding of the differences in linguistic capacity between humans and other creatures.

6. While human languages have evolved as a form of acoustic communication, these languages can readily be extended into nonphonetic acoustic modes, as well as a number of forms employing sensory modalities other than hearing. A cross-modality comparison of the modes of linguistic communication should be useful in understanding the essential characteristics of human language.

Multiple Meanings of the Term "Phoneme"

The use of the same term to describe different entities can impair the development of a science. In a recent paper (Warren, 1983) I have attempted to show that there are four different uses of the term "phoneme": The articulatory phoneme refers to units employed in the production of speech; the acoustic phoneme refers to units employed to classify the sounds of speech; the perceptual phoneme refers to units employed in the auditory organization of heard speech; the graphemic phoneme refers to the written symbol employed to designate any or all of the other three classes of phonemes. As we shall see, the lack of correspondence between entities bearing the same name has caused great confusion concerning the nature of speech, and this confusion has implications for theories concerning the evolution of speech.

The Alphabet and Its Relation to Graphemic and Articulatory Phonemes

The concept that speech can be analyzed into a sequence of phonemes can be traced back to alphabetic writing (for discussion, see Warren, 1983). Unlike other forms of writing, the alphabet seems to have been invented only once, and to have spread rapidly to other cultures. The alphabet is based upon articulatory activities employed in generating speech. It was an insight of considerable utility to consider that there are a limited number of ways of producing sounds used in a particular language, and that by using a separate written symbol for each of these sound-generating activities, it is possible to transcribe speech as a sequence of articulatory gestures. Note that I have described this alphabetic analysis of speech in terms of articulatory activities rather than sounds (the evidence for and the significance of this distinction will emerge shortly). It is possible to analyze and tabulate these activities readily by direct observation involving oneself and others. The positions employed for consonants are in general easiest to observe, and historically consonants were transcribed by graphemes first. The manner of producing vowels is not as readily observable, and early alphabetic writing did not include symbols for vowels. Writing with a full alphabet of consonants plus vowels permits an unfamiliar word to be pronounced, since the string of graphemes not only represents the word but provides instructions for its production. Of course, the graphemes used for languages such as English may diverge considerably from current pronunciation. However, it is still possible for readers to pronounce many unfamiliar printed English words with some degree of accuracy. Other languages have maintained closer correspondence between orthography and pronunciation than English, and as we know, the "phonetic" alphabet employs a series of graphemes designed especially to correspond closely to spoken language.

Differences Between Articulatory Phonemes and Acoustic Phonemes (Speech Sounds)

It is often assumed that every articulatory phoneme has a corresponding acoustic phoneme. However, devices capable of analyzing speech sounds acoustically have indicated that this assumption is false. The lack of correspondence between articulatory phonemes and their acoustic consequences has resulted in what Klatt (1979) has called the "acoustic-phonetic non-invariance problem." To take one example, the acoustical nature of the articulatory phoneme /d/ in /di/ is quite different from the acoustical nature of the /d/ in /du/ (see Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967). The great effects of neighboring speech sounds upon the nature of acoustic "phonemes" are evident when attempts are made to read sound spectrograms which display the results of a spectral analysis in visual form. The sound spectrograph was developed in the 1940s by Bell Laboratories with the hope of enabling the deaf to understand speech through vision (Potter, Kopp, & Kopp, 1947). However, even with considerable practice, it is not possible to use such a display for real-time perception of speech due to the varied acoustic forms of the same "phonemes." Nevertheless, there are those who maintain that although some acoustic characteristics of a speech sound change with context, there may be other invariant cues (not readily apparent through acoustic analysis), which are used by listeners for identification of acoustic phonemes (see Jusczyk, Smith, & Murphy, 1981; Stevens & Blumstein, 1981).

Is There a Perceptual Phoneme?

Most theories of speech perception have assumed the existence of phonetic units at some level of auditory analysis (for discussion, see Warren, 1982; 1983). However, there is now considerable evidence that phonetic analysis is not necessary for speech perception, and probably does not take place as a precursor to comprehension. Many of the persistent and ingenious attempts to demonstrate the invariance of acoustic phonemes result from the need to use such entities for phonetically-based perceptual theories. But if there are no phonetic perceptual units, then this need vanishes.

Let us examine some of the evidence that phonemes are not units for the perception of speech. It has been shown that before children can read, they have great difficulty in segmenting words into speech sounds corresponding to phonemes or graphemes (Calfee, Chapman, & Venezky, 1972; Gibson & Levin, 1975; Gleitman & Rozin, 1973; Savin, 1972). Once children have progressed to reading in school, then division of words into phonemes becomes possible (Liberman, Shankweiler, Fischer, & Carter, 1974). It might be considered that the facilitation of phonetic segmentation results from developmental changes and increased linguistic skills rather than the acquisition of reading ability. However, Morais, Cary, Alegria, and Bertelson (1979) reported that adults who had never learned to read could not recognize, delete, or add phonemes to words, but that other members of the same population of illiterates could perform these tasks involving phonemes following training in special adult reading classes.

Another line of evidence indicating that phonemes are not employed as perceptual units is provided by reaction time studies. It has been shown that the time required to react to phoneme targets in syllables is greater than the time required to react to the syllables themselves (Savin & Bever, 1970). These results suggested to Savin and Bever that the phoneme may be derived from prior identification of the syllable, rather than serving as the unit requiring identification before the syllable can be recognized. Support for this view was afforded in a study by Warren (1971) in which prior syntactic and semantic contexts within sentences were manipulated to vary the probability of occurrence of target words. As anticipated, a more likely word was identified more quickly. But, the point of interest for this discussion is that a contextually facilitated reaction time to a word as measured for one group of subjects was associated with a similar facilitation of the reaction time to an individual phoneme target within that word as measured for a separate group of subjects, in keeping with the hypothesis that phonemes are derived perceptually from words, not the words from phonemes.

Consequences of Non-Phonetic Theories of Speech Perception upon Theories of Speech Evolution

If we rid ourselves of the belief that speech perception rests upon special processing requiring the identification of component phonemes and their orders, then a number of questions suggest themselves, such as whether equivalent rules govern the perception of acoustic patterns in other animals and whether the rules governing speech recognition also govern the recognition of nonverbal patterns in humans. A number of investigators have shown that nonhuman animals could be taught to discriminate between different isolated phonemes and between different syllables. Thus, it has been shown by Dewson (1964) that cats can learn to distinguish between the vowels "ee" and "oo" whether spoken by a woman or a man. Kuhl and Miller (1978) taught chinchillas to discriminate between the voiced and unvoiced consonant pairs represented by "kah" and "gah," "pah" and "bah," and "tah" and "dah." Warfield, Rubin, and Glackin (1966) reported that cats could be taught to discriminate between "cat" and "bat" and that the limit of acoustic distortion permitting discrimination was similar for cats and humans. Since it cannot be argued that these animals have evolved genetically determined mechanisms specialized for human speech sounds, these studies must be tapping some general mechanisms for detection of acoustic sequences. It has been suggested that humans and other animals possess mechanisms for perceiving complex patterns holistically, so that the pattern is recognized as an entity without the need for analysis as a sequence of identifiable items in a particular order (for a discussion see Warren, 1982). Studies of sequences of hisses, tones, and buzzes have helped to demonstrate that we share the ability to recognize complex acoustic sequences holistically with animals, and that this ability serves as the basis for the perception of speech.

Holistic Pattern Recognition in Humans

Several studies have demonstrated that humans can discriminate between permuted orders in otherwise identical sequences consisting of non-speech sounds, even when the acoustic components are too brief to be identified.

Efron (1973) and Yund and Efron (1974) found that listeners could distinguish between "micropatterns" consisting of permuted orders of two-item sequences (for exampe two tones), when the separation between the sounds was only one or two msec. Listeners appeared to discriminate on the basis of qualitative differences, and could not identify the order of components. These observations were confirmed in essential details by Wier and Green (1975).

Two-item sequences are rather special, and the use of iterated sequences of three or four sounds was introduced by me as a way of studying the perception of continuing sequences consisting of only a few items (Warren, 1968; Warren, Obusek, Farmer & Warren, 1969). It was found that three- or four-item "recycled" sequences of nonverbal sounds require at least 200 msec/item for identification of the order of items, yet it is possible to distinguish readily between different arrangements of the same sounds down to 5 or 10 msec/item whether subjects are trained (Warren, 1974a) or untrained (Warren, 1974b). While discriminating between permuted orders of brief items is accomplished on the basis of qualitative or holistic perceptual differences, the ability to discriminate between different orders of items having durations longer than a few hundred milliseconds appears to rest upon the linguistic skill of naming items in their appropriate order and remembering this sequence of names (Warren, 1974a, Teranishi, 1977). We shall return to the use of verbal mechanisms for discriminating between different arrangements of long-duration items later when we discuss sequence perception in animals other than humans. At this point, it should be noted that there is no upper limit for item durations permitting discrimination of permuted orders in humans.

Holistic Pattern Recognition in Nonhuman Animals

A few studies have examined the ability of nonhuman mammals to distinguish between permuted orders of discrete sounds. While each of these studies has found that the animals employed could discriminate between permuted orders of sounds having brief durations, it was observed that a breakdown in the ability to distinguish between different orders of the same items occurred when the item durations exceeded more than a few seconds.

Dewson and Cowey (1969) taught monkeys to discriminate between the four possible pairs of sounds which can be generated using a tone and a hiss (tone-hiss, hiss-tone, hiss-hiss, tone-tone) when items had durations of less than about 1.5 sec. At item durations of 3 sec and greater, the monkeys could not perform the task, and it appeared that they were unable to remember the first item after the second item ended (they were not permitted to respond until the sequence was completed). Monkeys are primarily visual rather than auditory, and their failure to master the task at long-item durations might be attributed to a general difficulty with auditory tasks. However, a similar experiment was carried out using the dolphin (Thompson, 1976), a creature generally considered both highly intelligent and primarily auditory rather than visual in its normal activities. Four sounds, which can be designated as A, B, C, and D, were used to construct sequences of two sounds which were presented through hydrophones. The dolphin was rewarded if it pressed one paddle following the sequences AC or BD, or if it pressed a different paddle following the sequences AD or BC. The sounds had a fixed duration, and a silent period of variable length was inserted between the first and second sounds of the pairs. In order to respond appropriately, the dolphin needed to remember the first sound until the second sound occurred. Thompson reported that nearly perfect performance was obtained when the interval separating the sounds was less than 2 or 3 seconds. At longer temporal separations, performance was at chance levels. He concluded that the ability to hear the overall pattern ceased at the upper limit of behavioral discrimination, and that the perception of the overall pattern was required for a correct response.

The evidence which has been summarized suggests that speech perception is based upon the ability to recognize patterns of sounds holistically, and that we share this ability with other animals. Our perception of speech does not require the identification of component sounds and their orders--rather the identification of components and their orders within acoustic sequences is itself a linguistic skill.

What is Special About Human Linguistic Skills?

There seems little doubt that human language originated and evolved as an acoustically based method of communication employing sounds generated by our vocal tract. However, our use of language today does not require conventional speech sounds--whistled languages which remain intelligible over great distances have been developed as an ancillary method of communication in a number of mountainous areas (Busnel & Classe, 1976). Language does not even require acoustic signals: Reading is every bit as rapid and accurate in transmitting linguistic information, and sign languages are used with fluency by the deaf. Languages using visual signs need not correspond directly to a spoken language (as does signed English), but can develop into uniquely visual forms with quite different rules (as does American Sign Language). Language does not even require use of our special distance senses of hearing and vision: The sense of touch can be used by the blind-deaf in communication, and braille permits tactual reading by the blind.

Hence, although development of special sound-producing systems seems to be associated with the evolution of human language, linguistic communication can now operate without the use of sound when necessary. It seems that our use of language is based upon an ability to manipulate symbols according to learned conventions in an exceedingly complex and versatile fashion. These symbols can consist of auditory, visual, or tactile patterns. It is through the study of this symbol-manipulative ability within and across sensory modalities that we can more fully understand the mechanisms subserving human language and the evolutionary development of speech.

BIBLIOGRAPHY

Busnel, R. G., & Classe, A. (1976). Whistled languages. New York: Springer-Verlag.

Calfee, R., Chapman, R., & Venezky, R. (1972). How a child needs to think to learn to read. In L. W. Gregg (Ed.), Cognition in learning and memory, (pp. 139-182). New York: Wiley.

Dewson, J. H. III (1964). Speech sound discrimination by cats. Science, 144, 555-556.

Dewson, J. H. III, & Cowey, A. (1969). Discrimination of auditory sequences by monkeys. Nature, 222, 695-697.

Efron, R. (1973). Conservation of temporal information by perceptual systems. Perception & Psychophysics, 14, 518-530.

Gibson, E. J., & Levin, H. (1975). The psychology of reading. Cambridge, Massachusetts: MIT Press.

Gleitman, L. R., & Rozin, P. (1973). Teaching reading by use of a syllabary. Reading Research Quarterly, 8, 447-483.

Jusczyk, P. W., Smith, L. B., & Murphy, C. (1981). The perceptual classification of speech. Perception & Psychophysics, 30, 10-23.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. Journal of Phonetics, 7, 279-312.

Kuhl, P., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. Journal of the Acoustical Society of America, 63, 905-917.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Reading and the awareness of lingustic segments. Journal of Experimental Child Psychology, 18, 201-212

Morais, J., Cary, L., Alegria, A., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? Cognition, 7, 323-331.

Potter, R. K., Kopp, G. A., & Kopp, H. G. (1947). Visible speech. New York: Van Nostrand.

Savin, H. B. (1972). What the child knows about speech when he starts to learn to read. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye, (pp. 319-329). Cambridge, Massachusetts: MIT Press.

Savin, H. B., & Bever, T. G. (1970). The nonperceptual reality of the phoneme. Journal of Verbal Learning and Verbal Behavior, 9, 295-302.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), Perspectives on the study of speech, (pp. 1-38). Hillsdale, New Jersey: Erlbaum.

Teranishi, R. (1977). Critical rate for identification and information capacity in hearing system. Journal of the Acoustical Society of Japan, 33, 136-143.

Thompson, R. K. R. (1976). Performance of the bottlenose dolphin (Tursiops truncatus) on delayed auditory sequences and delayed auditory successive discriminations. Doctoral Dissertation, University of Hawaii.

Warfield, D., Ruben, R. J., & Glackin, R. (1966). Word discrimination in cats. Journal of Auditory Research, 6, 97-119.

Warren, R. M. (1968). Relation of verbal transformations to other perceptual phenomena. Conference Publication No. 42, Institution of Electrical Engineers (London), Supplement No. 1, 1-8.

Warren, R. M. (1971). Identification times for phonemic components of graded complexity and for spelling of speech. Perception & Psychophysics, 9, 345-349.

Warren, R. M. (1974). Auditory temporal discrimination by trained listeners. Cognitive Psychology, 6, 237-256. (a)

Warren, R. M. (1974). Auditory pattern discrimination by untrained listeners. Perception & Psychophysics, 15, 495-500. (b)

Warren, R. M. (1982). Auditory perception: A new synthesis. New York: Pergamon.

Warren, R. M. (1983). Multiple meanings of "phoneme" (articulatory, acoustic, perceptual, graphemic) and their confusions. In N. J. Lass (Ed.), Speech and language: Advances in basic research and practice, Vol. 9 (pp. 285-311). New York: Academic Press.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. Auditory sequence: Confusion of patterns other than speech or music. Science, 164, 586-587.

Wier, C. C., & Green, D. M. (1975). Temporal acuity as a function of frequency difference. *Journal of the Acoustical Society of America*, *57*, 1512-1515.

Yund, E. W., & Efron, R. (1974). Dichoptic and dichotic micropattern discrimination. *Perception & Psychophysics*, *15*, 383-390.

ACKNOWLEDGMENTS

The preparation of this manuscript was supported by grants from the National Institutes of Health (2 R01 NS19295-04A3), and by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under grant or cooperative agreement number, AFOSR 85-0260. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. This manuscript is submitted for publication with the understanding that the US Government is authorized to reproduce and distribute reprints for Governmental purposes.