UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2a. SECURITY CLASSIFICATION AUTHORITY

AD-A229 472

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 90 1163

6a. NAME OF PERFORMING ORGANIZATION
Richard M. Warren
University of Wis.-Milwaukee

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
AFOSR

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 [illegible]448

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

8b. OFFICE SYMBOL (If applicable)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR-88-0320

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

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	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual Progress Rep.	FROM 11/1/89 TO 9/30/90	1990, November, 9	9 + Appendices

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	auditory perception, complex sounds

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

Five studies were completed: (1) It was found that, following repetition, long period (500 ms) random waveforms excised from Gaussian noise could be identified when embedded in longer segments of Gaussian noise even when the inter-stimulus interval exceeded the limits of echoic memory; (2) It was demonstrated that some spectral regions of these long-period random waveforms could be recognized with greater accuracy than others; (3) Experiments with three consecutive odd-numbered harmonics demonstrated that triads with low harmonic numbers have a pitch corresponding to the fundamental of the harmonic series, but triads centered at the 9th or 11th harmonic had pitches roughly one octave higher. Deviations from the octave were consistent with the waveform pseudoperiodicities. These pitch judgments have implications for theories concerning the bases for the dominant region of complex tones. Two series of experiments involving (4) the vowel conversion effect and (5) dichotic verbal transformations, which compared the rules governing perceptual organization of speech in Japanese and English, were carried out by the principal investigator during May and June at the Basic Research Laboratories of the Nippon Telegraph and Telephone Co., Tokyo.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dr. Genevieve Haddad	(202) 767-5021	AFOSR/NL

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

UNCLASSIFIED

DTIC ELECTE NOV 30 1990 S D

93 [illegible]

Report AFOSR-88-0320

MECHANISMS MEDIATING THE PERCEPTION OF COMPLEX ACOUSTIC PATTERNS

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

12 November 1990

Annual Progress Report For Period November 1, 1989 - September 30, 1990

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

Annual Progress Report AFOSR

SUMMARY

Work with long-period (500 ms) random waveforms excised from Gaussian noise demonstrated that, following repetition, those complex patterns could be identified when embedded in longer segments of Gaussian noise, even when inter-stimulus intervals exceeded the limits of echoic memory. The investigators also found that some spectral regions of these randomly derived waveforms could be recognized with greater accuracy than others. Experiments with tonal triads consisting of consecutive odd-numbered harmonics demonstrated that low harmonic numbers produce pitch judgments corresponding to the fundamental of the harmonic series, whereas triads centered at the ninth or eleventh harmonics produce pitches approximately 1 octave higher, but deviating from exact doubling by amounts consistent with waveform pseudoperiodicities. These results have implications for theories concerning the dominant region for the pitch of complex tones. Finally, the Principal Investigator carried out studies at the Basic Research Laboratories of the Nippon Telegraph and Telephone Company which demonstrated both similarities and differences in rules governing the perceptual organization of phonetic sequences in English and Japanese.

STATEMENT OF WORK

Five experimental studies have been completed by the investigators during the second year of the current grant. In addition, a chapter has been written attempting to show how auditory mechanisms employed for the processing of complex nonverbal patterns have been modified for the perception of speech.

1. Listeners attempted to identify long period (500 ms) noise segments that were heard first for about 30 s as recycling frozen noises (RFNs), and then presented as components within longer period (1.5 s) RFNs. It was found that listeners could identify those patterns which contained the 500 ms noise segments after delays well beyond the limit of echoic memory, even though presentation of the pattern was changed from looped to linear format. Also, listeners were able to tap in synchrony with that portion of the longer pattern corresponding to the previously heard segment. It appears that recycling can provide an efficient procedure for rapidly establishing a stable memory for complex patterns.

2. Listeners attempted to identify individual RFNs (repetition periods 62.5 ms to 1 s) that were presented first under bandpass filtering, and then presented broadband. A second experiment reversed the order of presentation, and required listeners to identify bandpass derivatives of previously heard broadband RFNs, without knowing in advance which of three frequency bands would be presented: low (150-600 Hz), midrange (600-2400 Hz) or high (2.4 kHz - 9.6 Hz). In both experiments, recognition accuracy was good for the medium band and poor for the high frequency band. For the low range band, accuracy was good only when the isolated band was presented first.

3. Work done the previous year with perception of repeated sequences of steady-state vowels was extended this summer in a study carried out by the principal investigator and Shigeaki Amano at the Basic Research Laboratories of NTT (Nippon Telegraph and Telephone Company), Musashino, Tokyo. Work described in the previous annual report had shown that when the duration of vowels was between 30 and 100 ms, sequences consisting of from 3 to 10 English vowels were subject to the "vowel conversion effect" and

were heard as words, with permuted arrangements producing different verbal forms. Japanese has only five vowels, and repeated sequences of these vowels were presented to a trained panel of monolingual Japanese listeners. When vowel durations were 60 or 90 ms, listeners reported hearing two or sometimes three simultaneous verbal forms consisting of Japanese consonant-vowel syllables (consonant clustering does not occur in Japanese). Deletion of portions of each vowel while keeping the vowel-to-vowel onset time fixed resulted in the perception of individual vowels and inhibited conversion into Japanese syllables. Data obtained with the temporally contiguous vowels were consistent with a "temporal compound" model originally proposed for nonverbal sequences (which considers that the temporal contiguity of a series of brief sounds results in a loss of identity of the component elements and the formation of distinctive compounds). Perception of phonetic sequences follows this model, with the additional application of linguistic rules and restrictions characteristic of the listener's language.

4. An additional study was carried out at NTT with Makio Kashino, dealing with dichotic verbal transformations. Response differences for the right and left ears were observed which appear to differ in some ways from those obtained in earlier experiments in the USA. Data analysis is currently under way both in Tokyo and Milwaukee.

5. Previous studies have shown that tones consisting of only odd-numbered harmonics can produce two types of low pitch, one matching the fundamental frequency and another approximately one octave higher. The present study examined the bases for these judgments. It was found that triads consisting of consecutive odd-harmonic numbers produce a pitch corresponding to the fundamental when the harmonic numbers are low, whereas high-order triads produce pitches approximately one octave higher, but deviating from

exact doubling of the fundamental by amounts consistent with the waveform pseudo-periodicities. Interestingly, the change from fundamental to pseudoperiod matching occurred with triads centered at the ninth or eleventh harmonic, which had distributions within a critical band approximating those of the dominant partials in "all-harmonic" signals. These results are consistent with the report that the ninth and eleventh harmonics played together produce pseudoperiodicities at the level of the cochlear nucleus, and suggest that the dominant region is a transition zone where temporal and spectral modes of pitch analysis overlap.

6. The chapter in press relates speech perception to the perception of nonverbal complex patterns, drawing largely upon work in the principal investigator's laboratory. Among the topics covered are the relation of phonemic restoration to nonverbal auditory induction, the perception of linguistic and nonlinguistic sequences, and the perceptual changes associated with unchanging patterns of stimulation.

PUBLICATIONS

1. Warren, R.M., Bashford, J.A., Jr., & Gardner, D.A. Tweaking the lexicon: Organization of vowel sequences into words. Perception & Psychophysics, 1990, 47, 423-432.

2. Warren, R.M, Gardner, D.A., Brubaker, B.S., & Bashford, J.A., Jr. Melodic and nonmelodic sequences of tones: Effects of duration on perception. Music Perception (In Press).

3. Warren, R.M. Auditory illusions and the perceptual processing of speech. In N.J. Lass (Ed.) Principles of Experimental Phonetics. Philadelphia, B.C. Decker (In Press).

4. Bashford, J.A., Jr., & Warren, R.M. Pattern recognition within spectrally isolated regions of broadband sounds. Journal of the Acoustical Society of America, 1990, 87, S24 (Abstract).

5. Brubaker, B.S., & Warren, R.M. Auditory memory for long-period random waveforms. Journal of the Acoustical Society of America, 1990, 87, S24 (Abstract).

6. Chalikia, M.H., & Warren, R.M. Mapping the organization of vowel sequences into words. Journal of the Acoustical Society of America, 1990, 87, S160 (Abstract).

7. Riener, K.R., Bashford, J.A., Jr., & Warren, R.M. Increasing the intelligibility of speech through multiple phonemic restorations. Journal of the Acoustical Society of America, 1990, 87, S71 (Abstract).

8. Chalikia, M.H., & Warren, R.M. Spectral factors in the organization of vowel sequences into words. Journal of the Acoustical Society of America. 1990. 88. S54 (Abstract).

9. Riener, K.R., & Warren, R.M. Verbal organization of vowel sequences: Effects of repetition rate and stimulus complexity. Journal of the Acoustical Society of America, 1990, 88, S55 (Abstract).

10. Bashford, J.A., Jr., & Warren, R.M. The pitch of odd-harmonic tones: Evidence of temporal analysis in the dominance region. Journal of the Acoustical Society of America, 1990, 88, S48 (Abstract).

PROFESSIONAL PERSONNEL

In addition to the Principal Investigator, Dr. R.M. Warren, Dr. J.A. Bashford, Jr. is participating in the project in the capacity of Associate Researcher. Graduate students who have been assisting in this project this past year are Bradley S. Brubaker, Keri, R. Riener, and Eric Healy.

PROFESSIONAL INTERACTIONS

1. The Principal Investigator was Visiting Senior Scientist at the Basic Research Laboratories of Nippon Telegraph and Telephone Company at Musashino, Tokyo during May and June, 1990. He presented a series of lectures dealing with his research, served as consultant for ongoing research programs, and initiated two experimental studies, one on perceptual organization of sequences of steady-state Japanese vowels (conducted with Shigeaki Amano), and the other on dichotic verbal transformations in volving both English and Japanese words (conducted with Makio Kashino). Papers based upon these studies were presented by Kashino-san and Amano-san at a meeting of the Acoustical Society of Japan, September, 1990.

2. Four papers were presented at the May, 1990 meeting, and three papers will be presented shortly at the November, 1990 meeting of the Acoustical Society of America.

3. I will be a speaker at a Tutorial Workshop entitled "Cognitive Aspects of Human Audition" organized by the Hearing Group of the French Acoustical Society and IRCAM to be held at Paris, 21-22 February 1991. My contribution on "Perception of Global Properties of Sound Sequences" will be published subsequently as a chapter in a book to appear in both English and French versions (English title "Cognitive Aspects of Human Audition", S. McAdams & E. Bigand, Eds.).

APPENDICES

1. Warren, R.M., Bashford, J.A., Jr., & Gardner, D.A. Tweaking the lexicon: Organization of vowel sequences into words. Perception & Psychophysics, 1990, 47, 423-432.

2. Warren, R.M, Gardner, D.A., Brubaker, B.S., & Bashford, J.A., Jr. Melodic and non-melodic sequences of tones: Effects of duration on perception. Music Perception (In Press).

3. Warren, R.M. Auditory illusions and the perceptual processing of speech. In N.J. Lass (Ed.) Principles of Experimental Phonetics. Philadelphia, B.C. Decker (In Press).

4. Text of Poster "Auditory Memory for Long-Period Random Waveforms" presented at the May 1990 Meeting of the Acoustical Society of America.

5. Text of Poster "Increasing the Intelligibility of Speech through Multiple Phonemic Restorations" presented at the May 1990 Meeting of the Acoustical Society of America.

6. Text of Poster "Pattern Recognition Within Spectrally Isolated Regions of Broadband Complex Sounds" presented at the May 1990 Meeting of the Acoustical Society of America.

7. Text of Poster "Mapping the Organization of Vowel Sequences Into Words" presented at the May 1990 Meeting of the Acoustical Society of America.

APPENDIX 1

Perception & Psychophysics
1990, 47 (5), 423-432

Tweaking the lexicon: Organization of vowel sequences into words

RICHARD M. WARREN, JAMES A. BASHFORD, JR., and DANIEL A. GARDNER
University of Wisconsin-Milwaukee, Milwaukee, Wisconsin

The ability of listeners to distinguish between different arrangements of the same three vowels was investigated for repeating sequences having item durations ranging from 10 msec (single glottal pulses) up to several seconds/vowel. Discrimination was accomplished with ease by untrained subjects at all item durations. From 30 through 100 msec/vowel, an especially interesting phenomenon was encountered: The sequences of steady-state vowels were organized into words, with different words heard for the different arrangements of items. In a second experiment, repeating sequences of random arrangements of 10 40-msec vowels were employed. When sets of four such sequences were presented to listeners, distinctive words were heard, which permitted each arrangement to be discriminated from the others. In addition, minimal differences (reversing the order of a single contiguous pair of vowels) in the 10-item sequences could be detected via verbal mediation. Hypotheses are offered concerning mechanisms responsible for these results.

A succession of steady-state vowels presented loudly and clearly can be heard as a word. This unusual verbal organization can help us understand how acoustic components are processed in speech perception.

In the experiments reported here, repeated or recycled sequences were employed. These iterated stimuli were first used in the 1960s as a means of allowing a limited number of sounds (usually 3 or 4) to be presented for extended periods (Warren, 1968; Warren, Obusek, Farmer, & Warren, 1969). Repetition also helps eliminate the special cues to the order of items that are provided by the first and last items of a sequence (Divenyi & Hirsh, 1978; Warren, 1972). In studies with recycled vowel sequences, it has been shown that the naming or identification of order is accomplished readily at 200 msec per item, but that it is not possible at item durations below 100 msec (Cole & Scott, 1973; Cullinan, Erdos, Schaefer, & Tekieli, 1977; Dorman, Cutting, & Raphael, 1975; Thomas, Cetti, & Chase, 1971; Thomas, Hill, Carroll, & Garcia, 1970; Warren, 1968; Warren et al., 1969; Warren & Warren, 1970). In none of these studies have observations involving vowel durations below the threshold for identification of order been reported. However, in preliminary observations, we found that when three steady-state vowels (A, B, and C) were presented as recycled sequences, the two possible arrangements (...ABCABCA... and ...ACBACBA...) could be discriminated readily at item durations much briefer than the limit for naming of order.

Our first experiment confirmed that discrimination between different orders of the same speech sounds does not require the ability to identify the order of the phonemes, or indeed even the ability to identify the components within the sequences. Listeners were required to judge whether alternately presented recycled sequences of three vowels (which could be presented in identical or permuted item orders) were *same* or *different*. The vowels spanned the range from 10 msec (single glottal pulses) through 5 sec (500 glottal pulses), with no acoustic mixing or transitional stages in going from one vowel to the next. When vowel durations were above 100 msec, listeners could name the phonemes in the proper order, and they could then use the difference in named order to distinguish the two arrangements. When vowel durations were below 30 msec, resemblance to speech was absent, and differences in quality or timbre made it easy to discriminate between the two arrangements (for example, a listener might report that one sequence sounded "rougher" than the other). Between these values (30-100 msec), listeners could hear different words (usually lexical, sometimes nonsense) for each of the arrangements. The words heard differed across individuals, and they normally bore little resemblance to the actual phonemes.

In other studies too, verbal organization of repeated sequences of vowels has been observed. Dorman et al. (1975) used these stimuli to investigate the limits for identification of temporal order and noted in passing that verbal organizations interfered with the listeners' task when vowel durations approached the lower limit for order identification. Skinner (1936) used repeated sequences of barely audible vowels with durations of several hundred milliseconds. He reported that his listeners heard words and sentences. When the levels of vowels were raised well above threshold (as were the vowels in each of our experiments), "imitative" responses occurred, and the sequences were identified as a succession of vowels (in keep-

This study was supported in part by grants awarded to R. M. Warren from the National Institutes of Health (DC00208) and the Air Force Office of Scientific Research (88-0320). We thank Bradley S. Brubaker for his valuable help. Correspondence may be sent to Richard M. Warren, Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI 53201.

ing with observations made in our Experiment 1 for vowels with durations of a few hundred milliseconds). Skinner attributed the verbal organization of his faint syllabic-length vowels to a "summation" of originally subliminal responses.

In our second experiment, we employed 48 recycled sequences, each consisting of a different random arrangement of a set of 10 steady-state 40-msec vowels played loudly and clearly. Individual listeners heard characteristic words or pseudowords corresponding to each of the orders, and they could identify a particular sequence among several on the basis of its verbal correlate. Interestingly, listeners often heard a particular arrangement as two concurrent words that differed in timbre and/or pitch. As we shall see, this splitting of the stimulus provides a clue to the mechanisms employed for perceptual syntheses.

In another part of the second experiment, we also employed recycled sequences of 10 different 40-msec vowels. The stimuli consisted of pairs of sequences with minimal differences in structure (the orders of two contiguous vowels were interchanged). Listeners were again able to use verbal mediation to distinguish members of the pairs.

EXPERIMENT 1: DISCRIMINATION BETWEEN DIFFERENT ORDERS OF THREE-ITEM VOWEL SEQUENCES

This first experiment was designed in part to test the hypothesis that discrimination between different orders of the same speech sounds does not require the ability to identify the order of the phonemes, or indeed even the ability to identify the components within the sequences. Our listeners were required to judge whether alternately presented recycled sequences of three vowels (which could be presented in identical or permuted item orders) were *same* or *different*. The vowel durations extended from 10 msec to 5 sec, permitting a comparison of discrimination accuracy and cognitive strategies employed for durations corresponding to, as well as briefer and longer than, those occurring in speech.

Method

Subjects. Participants were recruited from introductory psychology courses; they received either course credit or cash for their participation. The students who passed the audiometric screening procedure described below were assigned randomly to one of two experimental groups, each containing 36 subjects.

Audiometric screening. All subjects participating in the experiments passed an audiometric test designed to eliminate individuals with hearing deficits, as well as anyone who failed to follow the standard instructions used with the Békésy threshold tracking procedure. Following presentation of instructions and familiarization with the task, a pure tone presented diotically was swept up from 400 through 9000 Hz, and then down from 9000 through 400 Hz, at a rate of one octave/minute while subjects tracked their thresholds. Tracking was accomplished using a hand-held button switch (depressing the button decreased the intensity at a rate of 2.5 dB/sec, and releasing the button increased the intensity at the same rate). An X-Y plotter produced audiograms consisting of continuous threshold tracings. Subjects were excluded from further participation if either directional sweep resulted in audiograms that differed from the 1964 ISO standards by more than 22.5 dB at any frequency for the portion of the audiogram extending from 500 through 8000 Hz.

Stimuli. The first stage in the preparation of the recycled sequences of three vowels used as stimuli involved production of extended steady-state recordings of three vowels (/ʌ/, /æ/, /i/) on parallel tracks of a multitrack recorder (16-track Ampex Model MM 1200). These steady-state vowels were derived from recorded statements of syllables containing these vowels ("hud" for /ʌ/, "had" for /æ/, and "heed" for /i/) produced by a male speaker at a vowel fundamental frequency of 120 Hz (the speaker matched the pitch of the vowel to that of a complex tone of 120 Hz heard through headphones). A complete single glottal pulse was excised from the central portion of each consonant-vowel-consonant statement. The waveforms of the glottal pulses were monitored and the period measured by a Nicolet Model 3091 digital storage oscilloscope used in conjunction with a programmable digital delay line (modified Eventide Model BD955) capable of repeating or "looping" stored input corresponding to a single glottal pulse. The repetition period of the delay line was set at 8.33 msec for each of the vowels (which corresponded to a repetition rate of 120 Hz), and recordings of the steady-state vowels were made on parallel tracks.

Two types of series were recorded for each duration employed (10, 12, 30, 100, 300, 1,000, 3,000, and 5,000 msec). A *different* series with successive sequence bursts consisting of /ʌ/, /æ/, /i/, /ʌ/, /æ/, /i/, ... /ʌ/, alternating with the permuted order /ʌ/, /i/, /æ/, /ʌ/, /i/, /æ/, ... /ʌ/, and a *same* series with all bursts consisting of /ʌ/, /æ/, /i/, /ʌ/, /æ/, /i/, ... /ʌ/. Note that because of the special ease of identifying the first and last items of a sequence (Warren, 1972), each sequence began and ended with the same item. Table 1 lists the parameters for the stimuli employed, giving item durations, the number of items (vowel statements) in each sequence burst, the interburst interval separating successive bursts (which were either identical or alternating in item order), and the number of bursts constituting a stimulus set.

All sequences (except those with a 12-msec item duration) were generated by gating the output from the three parallel tracks containing extended steady-state single vowels prepared as described above. The output of these tracks was passed through three Coulbourn electronic switches set for a rise/fall time of 2 msec. A series of timers (Grason-Stadler Model 1216A) controlled passage of each vowel through its gate, and introduced a 1-msec separation between the waveforms corresponding to the ending of one vowel and the beginning of the next as seen on the digital storage oscilloscope (this separation minimized the acoustic interaction of items). Another timer regulated the silent interburst interval separating successive bursts. The path of the signals through the equipment was identical in both the *same* and the *different* order series, with the relative timing of the opening and closing of the gates producing

Table 1
Description of the Stimuli Consisting of Three Recycled Vowels in Experiment 1

Item Duration (in msec)	Items per Sequence Burst	Interburst Interval (in msec)	Bursts per Stimulus Set	Stimulus Set Duration (in sec)
10	301	300	10	32.8
12*	301	300	10	38.8
30	91	300	10	30.0
100	31	300	8	26.9
300	10	300	8	26.1
1,000	10	1,000	4	43.0
3,000	7	1,000	4	87.0
5,000	7	1,000	4	143.0

*Items with locked waveforms.

the permuted orders. The number of vowel statements in the burst was controlled by Coulbourn predetermined counters (Model S43-30). The outputs of the gates were combined in a Yamaha audio mixer (Model PM-430) and band-passed from 100 through 8000 Hz by a Wavetek/Rockland filter (Model 751A) having slopes of 115 dB/octave before recording on one of the tracks of the multitrack recorder reserved for the experimental stimuli. The input level of each of the vowels as delivered to the tape recorder was adjusted separately to produce equal intensity (dBA) for all vowels on subsequent playback through the headphones used by the subjects. Sets of sequences were recorded for each item duration with the track for a *same* set parallel to the track of the corresponding *different* set, so that the experimenter could present either *same* or *different* stimuli at the same tape positions.

In sequences consisting of 10-msec items, only a single statement of each vowel's waveform was gated before switching to the next. Since this programmed switching was not exactly synchronous with the waveform repetition period of the recorded vowels (the recorder had a frequency stability of $\pm 0.1\%$), the repeated sequences underwent slow drifts in their waveforms and perceptual qualities. Sequences with longer item durations consisted of multiple identical statements of each vowel's waveform before switching to the next, and perceptual quality was more stable.

Sequences with 12-msec item durations were constructed with *locked* waveforms, so that switching always occurred at a fixed position in the waveform of each vowel and drifting did not take place. This stimulus was prepared as follows: Three separate delay lines (two modified Model BD955 and one modified Model 1745M Eventides) were driven by a common clock, and each repeated a single glottal pulse of a different vowel (the glottal pulses were derived from the same extended statements of the three vowels used for preparing the other sequences). The manner of capturing and repeating single glottal pulses on the delay lines was similar to that already described, except that the programming equipment introduced a 3.67-msec silent interval between successive statements of glottal pulses. The splice point of each single-vowel digital loop was at the center of this silent interval. The repetition period of all delay lines was set at 12 msec (measured by a common clock), the vowel statements were aligned so that each of the three vowels began and ended synchronously, and the vowels were then recorded simultaneously on separate tracks of the multitrack recorder. A timing signal (a unipolar pulse) generated by one of the delay lines at the splice point of its digital loop was recorded on a fourth track at the same time as the vowels. This recorded timing signal permitted the programming equipment to gate single glottal pulses of each recorded vowel in the desired order. Following gating and mixing, the locked three-item vowel sequences were recorded on an additional channel of the multitrack recorder, as described for the other sequences.

Procedure. The subjects passing the audiometric screening test were recalled for their single experimental session lasting about 40 min. They were tested individually while seated in an audiometric room along with the experimenter. The stimuli were presented diotically through matched headphones at a level of 70 dBA as measured by a sound-level meter with a 6 cc coupler. The experimenter operated the tape recorder (located outside the chamber), using a remote control unit equipped with a preset multipoint rapid search-to-cue device. Switches on an audio mixer permitted delivery of the output from the desired track of the recorder.

Half of the 72 subjects served in the main experiment, which included all the sequence pairs listed in Table 1 except for the sequences with 10-msec vowel durations (that is, item durations of 12, 30, 100, 300, 1,000, 3,000 and 5,000 msec) presented in the order listed. The order of increasing item durations was employed to avoid the possibility (discussed by Warren, 1974) that with a series of decreasing item durations, the naming of orders at brief item durations could be accomplished through recognition of qualitative similarities to the previous sequences having longer item durations with directly identifiable orders. As described earlier, the 12-msec sequences had switching from one vowel to the next vowel locked, so that each restatement of a particular vowel was a single intact glottal pulse. All other sequences were *nonlocked*, with successive statements of each vowel starting and stopping at different waveform positions. The separate group of 36 subjects serving in the supplemental experiment received only the stimulus consisting of 10-msec vowels.

The subjects were told that they would be hearing patterns of sounds separated by brief silent intervals, and that their task was to determine if all patterns were identical or if alternate patterns differed in any way. They were instructed to call out "same" or "different" at any time during the stimulus presentation. They were informed that the occurrence of *same* and *different* groupings would be randomly determined. The subjects were encouraged to ask questions if any part of the instructions was unclear. After both the subject and experimenter were satisfied that the instructions were understood, the sequences were presented in an order of increasing item duration for the 36 subjects in the main experiment (the 36 subjects serving in the supplemental experiment received only the 10-msec items).

Before presentation of unknown *same* or *different* sequences at each item duration, each subject was given sample sequences (first a *different* set, then a *same* set) which were identified by the experimenter as *same* or *different*. They were told that they could hear either of the known samples again, if they wished, before hearing the unknowns. When a subject indicated readiness, he or she was given three unknowns at that item duration. The *same* and *different* unknowns were presented in a pseudorandom order, with the constraint that all three of the unknowns presented to a subject at any item duration could not all be of a single type. No feedback was given concerning the accuracy of judgments with the unknowns. In the main experiment, of the total of 21 unknowns presented to each subject, 10 were the same and 11 were different for 18 subjects, and 11 were the same and 10 were different for the other 18 subjects. Half the subjects received orders of *same* and *different* unknowns that were "mirror images" of the other half, with *same* and *different* unknowns being interchanged to maintain symmetry of unknown groupings. This symmetry was also maintained for the supplemental group receiving the 10-msec vowel durations.

Results

Table 2 shows that the accuracy of discriminating between permuted orders ranged from 78% correct to 99% correct, and that it was significantly better than chance

Table 2
Accuracy of Same/Different Judgments for Recycled Sequences of Three Vowels in Experiment 1

Stimulus Duration	Responses: No. Correct (out of 108)	Responses: % Correct	Z Scores*
10†	84	78	5.82
12‡	98	91	8.52
30	90	83	6.86
100	98	91	8.52
300	102	94	9.15
1,000	103	95	9.35
3,000	106	98	9.98
5,000	107	99	10.18

Note—Stimulus duration is given in milliseconds for each stimulus item. *All $ps < .0001$. †Judgments made by separate groups. ‡Items with locked waveforms.

for all of the item durations used ($Z \geq 5.77$, $p < .0001$). The sequences consisting of 10-msec items (with slowly drifting waveforms and perceptual qualities) had 78% correct responses, while the 12-msec locked sequences (with switching occurring at fixed points corresponding to the beginning and end of the single glottal pulse representing each vowel) had 91% correct responses. This difference was significant ($Z = 2.62$, $p < .01$).

Questioning of listeners after completion of the formal experiment suggested that two basic ways of discriminating between the different arrangements of items were used: (1) naming of components in their proper order for vowel durations greater than 100 msec; and (2) a holistic recognition of patterns without the ability to identify the order of components (or even the components themselves) for vowel durations from 100 msec down to 10 msec. The range from 100 to 10 msec consisted of two regions: (2a) From 100 to 30 msec, the sequences of three vowels could be heard as words rather than steady-state vowels, with different words heard for the different arrangements; (2b) below 30 msec, the vowel sequences were heard as nonlinguistic sounds, with different qualities associated with the different arrangements. These perceptual categories (1, 2a, 2b) reported by untrained listeners agreed with observations made by laboratory personnel.

Discussion

Limits for the naming of order. The earliest experiments with recycled sequences of sounds measured thresholds for identifying the order of component items (Warren, 1968; Warren et al., 1969; Warren & Warren, 1970). When four 200-msec sounds were used, listeners instructed to name the order of items performed at chance level with unrelated sounds consisting of noises, tones, and buzzes, but they could accurately name the order of vowels. In subsequent studies, it was established that the threshold for identifying the order of unrelated sounds is 300 msec or more (Warren & Obusek, 1972), and that the threshold for correctly ordering the pitches associated with sequences of four sinusoidal tones is between 125 and 200 msec/item (Nickerson & Freeman, 1974; Thomas & Fitzgibbons, 1971; Warren & Byrnes, 1975). The lowest thresholds for four-item sequences (about 100 msec/item) were obtained with vowels (Dorman et al., 1975; Thomas et al., 1971).

There seems to be general agreement that vowel order can be named at briefer durations than is possible with other sounds. Why this difference? Using evidence from several sources, it was proposed by Warren (1974; also suggested independently by Teranishi, 1977) that the time required for verbal labeling or naming of components in extended sequences was the threshold-determining stage in the identification of order. Since vowels have a name that is the same as the sound itself, no recoding is necessary (naming order can be accomplished through a simple echoic restatement of the stimulus items), and the time required for identifying order is minimal. Nevertheless, as discussed below, the threshold value of 100 msec seems too high for agreement with models that consider identification of phonemic order to be necessary for the comprehension of speech.

Normal conversation has an average duration of speech sounds of about 80–100 msec; this duration drops to about 70 msec for oral reading, and some comprehension of "compressed speech" is possible at average phonetic durations of only 30 msec (for a brief summary of this literature, see Warren, 1982, pp. 119–120). Recognizing that there was a discrepancy between the rate of phoneme occurrence within intelligible speech and the ability to perceive order in a sequence of independently generated speech sounds, Wickelgren (1969) suggested that context-sensitive allophones facilitated temporal ordering. Coarticulation is, at least in part, an acoustic consequence of inertial and neuromuscular constraints on the movement of the tongue and other articulatory organs from one position to the next, and Wickelgren considered that recognition of particular allophonic forms might make it possible to identify more than one phoneme in a single speech sound. Thus, order could be identified at much briefer durations than would be possible for a succession of independent sounds. A number of experiments have demonstrated that coarticulation (and other cues increasing the resemblance of phonetic sequences to normal speech) does indeed facilitate the task of naming components and their orders (Cole & Scott, 1973; Cullinan et al., 1977; Dorman et al., 1975; Warren, 1968; Warren & Warren, 1970). But in no case, even with coarticulation cues, could orders be identified at item durations below 100 msec. However, listeners can comprehend speech consisting of phonemes with average durations of considerably less than 100 msec. One explanation for this discrepancy is that phonetic order is determined at some early level of linguistic processing that is not accessible for the naming of this order. Another hypothesis (which we favor) is that a determination of the order of component speech sounds is not necessary at any level of analysis for the recognition of words or for the comprehension of discourse. It is to be suggested that acoustic sequences need not function as perceptual sequences (that is, a succession of discrete sounds). Patterns formed by particular arrangements of speech sounds may be recognized as *temporal compounds* without any need for identification of constituents. As discussed below, this concept of temporal compound formation was formulated initially on the basis of experiments with nonverbal sounds.

Nonphonetic temporal compounds. In earlier studies involving arbitrarily selected sounds (noises, sinusoidal tones, and complex tones), listeners attempted to distinguish between different arrangements of repeated sequences consisting of the same three items, which were presented without any acoustic interactions or transitions involving contiguous sounds (Warren, 1974; Warren & Ackroff, 1976). These studies demonstrated that the different arrangements of nonverbal sounds could be discriminated with ease for item durations from 5 through

100 msec—yet the naming of orders was not possible within this range. It was suggested that permuted orders of brief items could be distinguished through the bonding of components to form temporal compounds possessing characteristic qualities, even though the component acoustic elements and their arrangements could not be identified. Thus, for an isomeric pair of temporal compounds consisting of identical components arranged in different orders, a listener might describe one arrangement of the nonverbal sounds as "bubbly" and the other as "shrill." These qualitative differences served as the basis for accurate differentiation between different acoustic orders.

Vowel sequences and their verbal temporal compounds. Our subjects in Experiment 1 indicated that discrimination of permuted orders was accomplished in different ways at different item durations. When the item durations corresponded to single glottal pulses (10- and 12-msec vowels), listeners used nonverbal temporal compounds to distinguish between the permuted vowels. Thus, with these very brief durations, an individual might report, for example, that one order was characterized by a "dull" quality while the other order sounded "crisp." However, perceptual organization into syllables and words (verbal temporal compounds) occurred at vowel durations from 30 through 100 msec. Within this durational range, a listener might say that one arrangement of vowels resembled or brought to mind repetitions of the word "kettle," whereas the other arrangement sounded more like repetitions of "puddle"—this, despite the great phonetic differences between the actual stimuli and their lexical correlates. The specific word corresponding to a particular temporal arrangement varied from listener to listener.

It appeared desirable to study further the verbal organization of a succession of steady-state vowels into words, and Experiment 2 was undertaken in accordance with this purpose.

EXPERIMENT 2A: IDENTIFYING DIFFERENT ARRANGEMENTS OF 10-ITEM VOWEL SEQUENCES

Experiment 1 has shown that recycled sequences of steady-state vowels played loudly and clearly can be heard as coherent verbal utterances, and that different arrangements of the same vowels can be discriminated on the basis of their distinctive verbal organizations. Further informal observations indicated that roughly 30–80 msec/vowel was the optimal duration for hearing words. Experiment 2A was designed so that the characteristics of this vowel-word illusion could be examined using recycled sequences consisting of 10 40-msec vowels. The 400-msec duration of these sequences corresponded to that of words in normal conversation. During the experiment, listeners were presented with four recycled sequences, each having a different randomly determined vowel order, and they were instructed to use verbal organizations as a means of identifying the different patterns on second presentation.

Method

Subjects. Thirty-two auditometrically screened listeners (14 male and 18 female) were recruited from introductory psychology courses; they received either course credit or cash for participating. The screening procedure was the same as that described for Experiment 1.

Stimuli. For synthesis of the 10 vowel components, a Data Precision Co. Model 6100 Universal Waveform Analyzer, operating at a sampling rate of 40 kHz with 14-bit resolution, was used to excise single 5-msec glottal pulses from a male speaker's sustained productions (200-Hz voicing frequency) of 10 vowels (those in *heed, hid, head, had, hod, hawd, hood, hud, hoot,* and *herd*). The digitized glottal pulses were then iterated eight times, to produce 40-msec bursts that were judged by a panel of four trained listeners to be identifiable as the parent vowels. Linear ramps of 2.5 msec (0 dB minimum) were imposed upon the onset and offset of each vowel burst for suppression of transients, and the amplitude envelopes of the bursts were rescaled so that each would play back at the same level.

The 10 vowel bursts were sampled randomly without replacement and concatenated in digital form to create 48 10-item sequences (out of a total of factorial nine possible orderings). Digital-to-analog conversion and playback of the 400-msec sequences in recirculating form was accomplished using a Data Precision Co. Polynomial Waveform Synthesizer Model 2020-100 (40-kHz sampling frequency with 12-bit resolution). The analog playback of the recycling sequences was recorded on an Otari Model MTR 90-II 16-track recorder, with sequences to be presented on the same trial (4 sequences for each of the 12 trials) recorded in parallel on separate tracks. During the experiment, the output of the recorder was amplified by a Neotek Series I audio mixer and band-pass filtered from 50 to 8000 Hz with slopes of 115 dB per octave (Wavetek/Rockland Model 751A Brickwall Filter).

Procedure. The listeners were tested individually in an audiometric room, with the stimuli delivered at 70 dBA SPL through diotically wired TDH-49P headphones mounted in MX 41/AR cushions. The experimenter operated the Otari recorder (located outside the chamber) with a remote preset search-to-cue device. Switches on the audio mixer located inside the chamber permitted delivery of the output from the desired tracks of the recorder.

The listeners participated in 2 practice trials and 10 formal trials, with the 12 sets of sequences presented in the same order to all listeners. Each trial consisted of a learning phase and a test phase. During the learning phase, the listeners were presented successively with four sequences, and they were required to listen to the recycling vowel patterns until they could write down what the voice seemed to be saying. (For their transcriptions, the listeners used a response booklet with separate pages for each experimental trial.) It was explained that their written descriptions would provide a means of identifying the sequences during the test phase of the trial. Once the listener had provided written responses for each of the four sequences, the listener began the test phase, using a control box with buttons labeled A, B, C, and D. Each of the buttons could be used to deliver one of the four sequences presented during the learning phase of the trial, and the listener's task was to match the letter of each button with the previous verbal organization for that sequence. The listener did this by placing appropriately lettered cards beside the previous transcription.

The listeners were permitted to switch at will from one sequence to another during a trial's test phase, and they were given as much time as needed to complete the card-placing task. When matching was complete, the experimenter recorded the listener's response, provided feedback concerning accuracy, and began the next trial.

During the debriefing period that followed the 10th formal trial, the experimenter reviewed the transcriptions to verify pronunciation and asked general questions concerning the listener's responses.

Results

Despite the obvious initial doubt of most listeners that they could accomplish the experimental task, their perceptual organization of the recycling vowel sequences into syllables and words proved nearly effortless with little practice. The time required for initial verbal organization (that is, for writing down a description for a particular vowel sequence) decreased from an average of about 35 sec for the first practice sequence to an average of only 8 sec per sequence across the 10 formal trials. Furthermore, once formed during the learning phase of a trial, these perceptual organizations proved sufficiently distinct and stable to permit rapid and highly accurate identification of the different vowel orderings during the test phase. On the average, the listeners completed the four matches of the test phase in about 15 sec, and a majority of their responses were accurate even for the practice trials. Table 3 lists each trial separately and gives the numbers of listeners who identified correctly each of the four sequences for the individual trials.

The chance likelihood of correctly identifying all four sequences on a trial was 1/24, so each fully correct series of responses by a listener exceeded chance at the .05 level. As can be seen, listeners identified all four sequences with above-chance accuracy on most (better than 94%) of their attempts across the 10 formal trials, with no evidence of fatigue or interference due to earlier sequences.

For the 40 sequences presented in the formal trials, 35% of the listeners' responses were nonlexical syllables (which always followed the rules for phoneme clustering of English), and the remaining 65% were words and phrases. Interestingly, most listeners also reported that certain sequences were organized as two different words (e.g., "Frankie" and "go animal") that sounded as though they were produced simultaneously by voices differing in quality. Despite the fact that the sequences were presented in the same order to all listeners, there was very little intersubject agreement in the forms reported for specific vowel orderings. Thus, although the verbal organizations were formed rapidly and were sufficiently stable to permit later recognition of sequences, they were also highly idiosyncratic—perhaps due in part to the fact that the sequences were played as endless loops with no initial and terminal components.

Table 3
Numbers of Listeners (out of 32) with Perfect Scores (Correct Identification of Each of the Four 10-Item Vowel Sequences in a Trial) in Experiment 2A

	Practice Trials		Formal Trials									
	1	2	1	2	3	4	5	6	7	8	9	10
No. perfect scores*	22	27	28	31	30	31	28	31	32	32	27	32

*$p < .05$ for each perfect score.

Experiments 1 and 2A have shown that verbal mediation in the discrimination of random vowel sequences can be very robust when differences in order are substantial. Experiment 2B was designed so that we might determine whether lexical matching could be extended to the discrimination of minimal differences in order.

EXPERIMENT 2B: DISCRIMINATION OF MINIMAL ORDER DIFFERENCES WITH TEN-ITEM VOWEL SEQUENCES

In the previous experiments, permuted orders of brief vowels produced distinct verbal organizations, but the differences in order were typically quite extensive: In the two contrasting three-item sequences used in Experiment 1 (ABCA... and ACBA...), each of the three pairwise orderings of vowels was reversed (AB vs. BA, BC vs. CB, and CA vs. AC), and in Experiment 2A, the 48 10-item sequences were drawn without constraint from a pool of 362,880 possible recycled orders. In Experiment 2B, listeners made ABX judgments (deciding whether the unknown X was the same as A or B) for 10-item vowel sequences, in which A and B differed only in the ordering of two contiguous vowels. The listeners also reported the basis for their discriminations for each trial.

Method

Subjects. Four subjects participated in the study. Subjects B.B. and J.B. were psychoacoustically trained listeners who had participated in preliminary observations with 10-item sequences. Listeners J.R. and K.R. were not psychoacoustically trained and had no prior experience with the stimuli employed in this study.

Stimuli. Each of the 48 sequences used in Experiment 2A was used as Sequence A of a contrasting pair. The B sequence of each pair was produced by interchanging the order of two randomly selected contiguous vowels of the 10-item A sequence. Analog playback of the B sequences was recorded on the same 16-track tape as had been used for Experiment 2A, with corresponding A and B sequences arranged in parallel. As in Experiments 1 and 2A, the stimuli were amplified, using an audio mixer, and band-pass filtered from 50 to 8000 Hz with slopes of 115 dB/octave.

Procedure. As in the earlier experiments, the listeners were tested individually in an audiometric room, with the stimuli delivered through headphones at 70 dBA SPL. They were provided with a three-button panel, which they used for switching between contrasting A and B stimuli and a third X stimulus that matched the sequence presented in either the A or the B channel. The listeners switched at will between the three signals (each recorded on a separate track) until they were satisfied that they had determined which signal matched X. After calling out either "A" or "B," they attempted to describe the basis for their discrimination. The listeners were aware that their ABX matches were being timed, and they received trial-by-trial feedback concerning their matching accuracy.

The listeners participated in a total of 16 sessions, with each session lasting about 20 min and involving judgments of 6 pairs of contrasting A and B sequences. Across the 16 sessions of each experiment, the 48 sequence pairs were presented twice to each listener, for a total of 96 judgments. Each listener received a different random ordering of stimuli for the first ABX judgments of the sequence

Table 4
Accuracy and Response Times for ABX Judgments of Recycled 10-Vowel Sequences in Experiment 2B (A and B Sequences Differed in the Order of a Single, Contiguous Pair of Vowels)

Listener	No. Correct (out of 96)*	Response Times (in sec)		
		Median	Q_1	Q_3
B.B.	92	34.5	25.0	51.0
J.B.	94	50.5	30.0	107.5
J.R.	94	72.0	41.0	114.0
K.R.	94	42.0	28.0	68.5

*Accuracy scores for all listeners exceeded chance ($Z \geq 8.98$, $p < .0001$).

pairs, and this order was repeated for the listener upon second presentation of the stimuli, so that the two judgments for each contrast were separated by judgments of the remaining 47 sequence pairs.

Results

The number of correct responses (out of 96) and the median response times for judgments of each listener are presented in Table 4. As is shown, overall matching accuracy was well above chance for all listeners, with the percentage of correct responses ranging from about 96% to 98%.

The listeners' trial-by-trial reports concerning the nature of their discriminations indicated that, although they attributed some discriminations to contrasting nonverbal characteristics (typically, differences in rhythmic complexity), most of their judgments were based on differences in verbal organizations. These occasionally corresponded to pseudowords, but more often to real words (e.g., "valuable" vs. "technical"). Most interestingly, although there was little agreement across listeners in the verbal forms evoked by specific vowel sequences, there was substantial consistency within listeners: In 52% of the cases in which listeners reported specific words upon first presentation of a contrasting pair of sequences, they reported the same word or words on second presentation of the sequences. This repetition of responses occurred in spite of the fact that successive judgments of the same stimuli were separated by several days and by interpolated judgments of the remaining 47 sequence pairs. Thus, although the verbal correlates of these monotone vowel patterns were again found to be highly idiosyncratic, they were also remarkably stable.

Discussion

In studies with 10-item sequences of nonverbal sounds, it has also been found that minimal changes in 10-item sequences can be discriminated. Watson and his coworkers employed sequences of 10 or more brief sinusoidal tones in experiments on the ability to make fine discriminations (e.g., the ability to detect a change in the frequency of a single tone) within complex "word-length" patterns (see Watson, 1987, for a review). In these studies, in which contrasting sequences were presented as single statements, it was found that listeners usually required many hours of training before they could accomplish discrimination. However, Bashford and Warren (1988) have reported that when sequences of 10 tones are recycled, the discrimination of fine changes is very much easier and can be accomplished in less than 1 min in an ABX discrimination task. They found performance with minimal changes (inverting the order of 2 of the 10 tones) to be only slightly poorer than that observed with recycled 10-vowel sequences. Hence, although perception in a "speech mode" (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) is employed for sequences of vowels, fine discrimination can be accomplished with nonlinguistic sequences as well. Bashford and Warren (1988) also reported that sequences need not involve successions of discrete sounds for successful discrimination. Noise "sequences" were constructed by sampling from a catalogue of 10 40-msec segments that had previously been excised from white noise. When the segments were abutted to form a loop, the recycled "sequence" resembled a repeated 400-msec segment of noise that lacked the succession of discrete sounds characteristic of sequences of tones and vowels. Interchanging the order of two contiguous 40-msec noise segments resulted in a discriminable change—although ABX judgments did take about twice as long as those with the recycled sequences consisting of 10 40-msec tones or vowels.

Let us look more closely at the linguistic organization occurring within sequences of vowels in Experiments 1, 2A, and 2B. How is it that syllables and words are heard with a succession of steady-state vowels, despite the great differences between the phonetic compositions of the stimuli and the forms reported?

We hypothesize that the organization of our sequences of loud and clear vowels into syllables and words reflects shifts in perceptual criteria produced by repetition. The *criterion shift rule*, which has been proposed for judgmental processing in general, considers that the criteria used for evaluating stimuli and events are displaced in the direction of simultaneous or recently experienced values (Warren, 1985). When applied to psycholinguistics, this effect can produce changes in the perceptual boundaries of phonemes following exposure to repeated syllables. While there is considerable controversy concerning the processes responsible for these boundary shifts (for discussion, see Diehl, Kluender, & Parker, 1985), there is agreement that the changes that do occur move the acoustic boundaries delineating particular phonemes toward a closer correspondence with the iterated stimulus. This shifting of criteria may be considerably greater when repetition is continuing (as in the present experiments) than after repetition ceases (as is typically the case in studies measuring the extent of category boundary shifts). In the present experiments, it appears that the continuing repetition of a loud and clear sequence of steady-state vowels changed the acoustic requirements for recognition of a syllable or word to the point at which the stimulus itself could be perceived as a particular utterance by a speaker.

The perceptual matching of a repeated vowel sequence to a particular verbal form may be facilitated not only by a criterion shift, but also by the splitting of the stimulus into two simultaneous percepts. Recall that typically an iterated vowel sequence splits into two concurrent forms—usually, two voices with different pitches or qualities, which repeat different things at the same time (although sometimes a single verbal form is heard, accompanied by a nonverbal sound). We suggest that matching of the auditory input to the particular patterns (or templates) required for perception of syllables or words involves separation of the signal into two fractions. One fraction is matched to the template corresponding to a syllable or word (as modified by a repetition-induced criterion shift). The other fraction corresponds to the residue remaining after subtraction of the components of the auditory input that are used for this match. This residue can appear as a nonlinguistic noise, or it may be matched to a second linguistic template and thus heard as a different voice repeating some other utterance. The process of synthesizing an auditory signal through subtraction of the appropriate components from a louder sound has been called *auditory induction* (Warren, 1984; Warren, Obusek, & Ackroff, 1972). In conjunction with repetition-induced shifts in acoustic criteria defining linguistic templates, auditory induction could facilitate the matching of vowel sequences to syllables and words, by permitting the segregation of spectral components corresponding to these modified templates.[1]

It would be of interest to determine the correspondence of individual speech sounds forming the illusory words to the vowels actually present at that time. Preliminary experiments have shown that the mapping of perceptual phonemes to acoustic phonemes can be accomplished, but not through methods that might appear to be the most obvious. Placing an acoustic marker such as a click in one of the vowels does not work, since clicks (and other extraneous sounds) are mislocalized in speech (Ladefoged, 1959; Warren & Obusek, 1971). Increasing the intensity of a vowel appreciably and then listening for a corresponding increase in the level of speech sounds in the illusory word does not work, because the illusory word usually continues to be heard, and the increased intensity results in one's hearing the vowel veridically—but as an extraneous sound that cannot be localized in the word. Deleting a vowel and listening for the disappearance of a portion of the illusory word does not work, because the illusory word can change to another form. However, a method for phoneme mapping of recorded speech employed in earlier studies (Warren, 1971; Warren & Sherman, 1974) does appear to work quite well. When the repeated sequence of vowels is abruptly terminated, the illusory word (or words) also stops suddenly, and it is easy to perceive the last speech sound heard. By systematically changing the point of termination of the sequence of vowels, one can map the perceptual phonemes to the acoustic phonemes. Further work employing this procedure is in progress.

SUMMARY AND CONCLUSIONS

Experiment 1 shows that repeated sequences consisting of different arrangements of the same three vowels can be distinguished either through naming the order of components (for item durations greater than 100 msec) or by means of recognition of patterns through temporal compound identification (for durations from 10 through 100 msec). Perception in a speech mode occurred for items from 30 through 100 msec, allowing permuted orders to be discriminated through perception of different verbal organizations for the different arrangements. Nonverbal temporal compounds permitted the discrimination of different arrangements for vowels briefer than 30 msec. In Experiments 2A and 2B, we examined the speech mode of perception further, by employing complex repeated sequences of 10 40-msec vowels. The recognition of different arrangements was accomplished readily through verbal mediation, even for the minimal changes in order produced by interchanging the position of two contiguous items. The vowel sequences were heard as a single utterance plus a noise, or as two concurrent utterances produced by distinctly different voices.

It was hypothesized that two mechanisms are involved in the illusory perception of words with repeated sequences. The syllabic or lexical templates employed for verbal recognition were temporarily warped into a closer resemblance to the repeated stimulus through repetition-induced criterion shifts, and matching of the stimulus to the template was then completed by extracting components needed for the match from the auditory input. This perceptual splitting of the stimulus (which also occurs during phonemic restoration) produced a residue that was either perceived as an extraneous sound accompanying the illusory verbal organization or organized into a second verbal form heard along with the first.

It is of interest that studies with animals other than humans have shown that, although the animals can discriminate between different arrangements of brief sounds, they fail when the task requires the remembering of sounds for more than a few seconds. As is discussed below, this difference between the performance of humans and other animals has suggested how speech perception might have evolved from auditory skills possessed by our prelinguistic ancestors.

On the basis of a literature survey of studies demonstrating that cats, chinchillas, and monkeys can be taught to recognize not only isolated phonemes, but also monosyllables, the suggestion has been made that the mechanisms employed by humans for speech perception have evolved through the elaboration of an ability to recognize overall patterns (or temporal compounds) that we share with other animals (Warren, 1982, 1988). In addition to the animal studies involving sequences of speech sounds, other experiments involving periodic sounds and noises have shown that dolphins (Thompson, 1976) and monkeys (Dewson & Cowey, 1969) can be taught to discriminate between pairs of brief sounds ar-

ranged in different orders. However, successful discrimination could be accomplished only when the sequences were brief; when the task required that these animals remember the identity of the first sound for 2 sec or more before hearing the second sound, the task became impossible (for further discussion, see Warren, 1982, pp. 137-138). It seems that discrimination between sequences with long separation between items requires a mechanism that is lacking in other animals but available to humans. This mechanism appears to involve verbal encoding, so that, for items separated by more than a few seconds, linguistic characterizations (rather than the memory of the sounds themselves) are stored to serve as the basis of discrimination.

For recognition of sequences with brief item durations (such as speech), neither humans nor other animals need identify the order of components or even the components themselves. Only temporal compounds need be recognized. Although listeners may be able to name the ordered series of phonemes corresponding to a word, this analytical description does not necessarily imply that the components themselves are perceived. Thus, Brubaker and Warren (1988) have demonstrated that listeners can readily learn to name the order of acoustic phonemes corresponding to words that are perceived, even when these words have phonetic transcriptions that do not correspond to the acoustic-phonetic components. They used recycled sequences of three vowels (as in Experiment 1). Their subjects were first presented with the two possible arrangements of the vowels at item durations of a few hundred milliseconds (permitting easy identification of order). They then heard these sequences at item durations that were decreased in a regular fashion down to values well below the threshold of 100 msec reported for identification of order with recycled sequences of vowels (Dorman et al., 1975; Thomas et al., 1971). At no time were the subjects ever told the actual phonemes or their orders. Through a series of successive generalizations, the subjects continued to identify accurately the constituent vowels in their proper orders, even though, as in the present study, the words heard at brief item durations did not have phonetic transcriptions corresponding to the acoustic phonemes actually present in the stimulus. It was concluded that the perception of syllables and words did not involve a "bottom up" or prior identification of an ordered arrangement of phonetic components. Rather, the identification of the acoustic phonemes and their orders required the mediation of a prior verbal organization.[2]

The recognition of lexical items in connected discourse, of course, consists of more than just the factors described above. Syntactic, semantic, and pragmatic rules come into play with lexical aggregates, and these emergent higher level processes can in turn influence word recognition. However, experiments involving perception of isolated words and phrases (as in the present study) can provide information concerning some of the flexible and opportunistic mechanisms used for the early stages of verbal processing.

REFERENCES

BASHFORD, J. A., JR., & WARREN, R. M. (1988). Discrimination of recycled word-length sequences. *Journal of the Acoustical Society of America*, **84**(Suppl. 1), S154.

BRUBAKER, B. S., & WARREN, R. M. (1988). Learning to identify phonemic orders. *Journal of the Acoustical Society of America*, **84**(Suppl. 1), S154.

COLE, R. A., & SCOTT, B. (1973). Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology*, **27**, 441-449.

CULLINAN, W. L., ERDOS, E., SCHAEFER, R., & TEKIELI, M. E. (1977). Perception of temporal order of vowels and consonant-vowel syllables. *Journal of Speech & Hearing Research*, **20**, 742-751.

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

DIEHL, R. L., KLUENDER, K. R., & PARKER, E. M. (1985). Are selective adaptation and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 209-220.

DIVENYI, P. L., & HIRSH, I. J. (1978). Some figural properties of auditory patterns. *Journal of the Acoustical Society of America*, **64**, 1369-1385.

DORMAN, M. F., CUTTING, J. E., & RAPHAEL, L. J. (1975). Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception & Performance*, **104**, 121-129.

LADEFOGED, P. (1959). The perception of speech. In *National Physical Laboratory Symposium No. 10: Mechanisation of thought processes* (pp. 309-417). London: Her Majesty's Stationery Office.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.

NICKERSON, R. S., & FREEMAN, B. (1974). Discrimination of the order of the components of repeating tone sequences: Effects of frequency separation and extensive practice. *Perception & Psychophysics*, **16**, 471-477.

REPP, B. H. (1989). Phone restoration. *Journal of the Acoustical Society of America*, **85**(Suppl. 1), S137. (Abstract No. DDD9)

SKINNER, B. F. (1936). The verbal summator and a method for the study of latent speech. *Journal of Psychology*, **2**, 71-107.

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

THOMAS, I. B., CETTI, R. P., & CHASE, P. W. (1971). Effect of silent intervals on the perception of temporal order for vowels. *Journal of the Acoustical Society of America*, **49**, 84.

THOMAS, I. B., & FITZGIBBONS, P. J. (1971). Temporal order and perceptual classes. *Journal of the Acoustical Society of America*, **50**, 86-87.

THOMAS, I. B., HILL, P. B., CARROLL, F. S., & GARCIA, B. (1970). Temporal order in the perception of vowels. *Journal of the Acoustical Society of America*, **48**, 1010-1013.

THOMPSON, R. K. R. (1976). *Performance of the bottlenose dolphin (*Tursiops truncatus*) on delayed auditory sequences and delayed auditory successive discriminations*. Unpublished doctoral dissertation, University of Hawaii.

WARREN, R. M. (1968). Relation of verbal transformations to other perceptual phenomena. *Conference Publication No. 42: Institution of Electrical Engineers* (Suppl. 1), 1-8.

WARREN, R. M. (1971). Identification time for phonemic components of graded complexity and for spelling of speech. *Perception & Psychophysics*, **9 (4)**, 345-349.

WARREN, R. M. (1972). Perception of temporal order: Special rules for initial and terminal sounds of sequences. *Journal of the Acoustical Society of America*, **52**, 67.

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

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. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin*, **96**, 371-383.

WARREN, R. M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review*, **92**, 574-584.

WARREN, R. M. (1988). Perceptual bases for the evolution of speech. In M. E. Landsberg (Ed.), *The genesis of language* (pp. 101-110). Berlin: Mouton de Gruyter.

WARREN, R. M., & ACKROFF, J. M. (1976). Two types of auditory sequence perception. *Perception & Psychophysics*, **20**, 387-394.

WARREN, R. M., & BYRNES, D. L. (1975). Temporal discrimination of recycled tonal sequences: Pattern matching and naming of order by untrained listeners. *Perception & Psychophysics*, **18**, 273-280.

WARREN, R. M., & OBUSEK, C. J. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics*, **9 (3B)**, 358-362.

WARREN, R. M., & OBUSEK, C. J. (1972). Identification of temporal order within auditory sequences. *Perception & Psychophysics*, **12**, 86-90.

WARREN, R. M., OBUSEK, C. J., & ACKROFF, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, **176**, 1149-1151.

WARREN, R. M., OBUSEK, C. J., FARMER, R. M., & WARREN, R. P. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science*, **164**, 586-587.

WARREN, R. M., & SHERMAN, G. L. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics*, **16**, 150-156.

WARREN, R. M., & WARREN, R. P. (1970). Auditory illusions and confusions. *Scientific American*, **223** (December), 30-36.

WATSON, C. S. (1987). Uncertainty, informational masking, and the capacity of immediate memory. In W. A. Yost & C. S. Watson (Eds.), *Auditory processing of complex sounds* (pp. 267-277). Hillsdale, NJ: Erlbaum.

WICKELGREN, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, **76**, 1-15.

NOTES

1. Another example of linguistic auditory induction is given by the *phonemic restoration effect*, in which contextually appropriate fragments of speech are synthesized from the substrate furnished by a louder sound of appropriate spectral characteristics (for a detailed discussion, see Warren, 1984). In keeping with induction theory, Repp (1989) has reported that spectral components corresponding to the restored phoneme are subtracted from an interpolated noise.

2. It has been suggested by Warren (1983) that there are four rather different uses of the term *phoneme* (acoustic, articulatory, graphemic, and perceptual), and that confusion has resulted from employing the same term for different entities. Warren argued that the existence of phonemes as units entering into the perceptual processing of discourse lacks direct experimental support, and that the treatment of perceptual phonemes in the literature is often confounded with acoustically based phonemes and with articulation-based phonemes.

(Manuscript received September 18, 1989;
revision accepted for publication November 20, 1989.)

APPENDIX 2
To be published in:
Music Perception

Melodic and Nonmelodic Sequences of Tones: Effects of Duration on Perception

by

Richard M. Warren, Daniel A. Gardner,
Bradley S. Brubaker, & James A. Bashford, Jr.

Department of Psychology
University of Wisconsin-Milwaukee
Milwaukee, Wisconsin 53201

ABSTRACT

Familiar melodic phrases were played repetitively with note durations ranging from 40 milliseconds to 3.6 seconds. Recognition required note durations approximating those normally employed for playing melodic themes (roughly 150 ms to 1 sec per note). Additional experiments with nonmelodic sequences of tones indicated that different rules applied for nonmelodic patterns: Permuted orders of the same items could be distinguished from each other at all durations employed (10 ms to 5 sec per item). Recognition of different arrangements occurred not only when each tone differed in pitch, but also when all tones had the same pitch but differed in timbre. It was concluded that the durational limits for melodic recognition are not based upon perceptual limits applicable to tonal patterns in general, but rather reflect special rules governing melodic organization. Hypotheses concerning the bases for these rules are suggested.

Why is it that melodies become unrecognizable when played very quickly? Winckel (1967) proposed a seemingly reasonable explanation, suggesting that rapid playing causes a perceptual metathesis of the component notes, so that different orders become indistinguishable and melodic organization is lost. But, as we shall see, listeners in the present study could distinguish between different orders of tones played at rates well above the limit for melodic recognition. At the other extreme, when the durations of the individual notes were increased to several seconds, listeners who lacked formal musical training could not recognize the same melodies they had identified with ease at normal tempos.

Preliminary observations suggested that the range of note durations over which repeated melodies could be recognized corresponded roughly to the range (150 ms to 900 ms) which Fraisse (1963, p. 89) gave as the normal rate of occurrence of notes used for melodic themes. These observations led to experiment 1 which measured the temporal limits for recognition of familiar tunes consisting of from six to nine notes which were presented at note durations ranging from 40 ms to 3.6 sec. Experiments 2a and 2b were undertaken to determine whether or not the temporal limits found for melodic recognition also applied to sequences of tones which did not form melodies. Listeners attempted to distinguish between permuted arrangements of the same tones at item durations ranging from 10 ms through 5 sec when the tones differed in pitch (experiment 2a) and also when each tone had the same pitch but differed in timbre (experiment 2b).

EXPERIMENT I: Temporal limits for melodic recognition

METHOD

Subjects. Participants without musical training were recruited from introductory psychology courses, and received either course credit or cash for their participation.

Stimuli. The eight melodies listed and described in Table 1 were synthesized using sinusoidal tones covering the frequency range from 311 Hz (D#4) to Hz (D#5). The rise/fall time for each note was 5 ms (which prevented audible clicks), and 20 ms of silence occurred between each note. The durations of notes given in Table 1 extend from the onset of one note to that of the next note. All melodies were prerecorded and presented diotically through headphones at a sound pressure level of 75dB (A scale weighting).

---Table 1 About Here---

Procedure. Subjects served in a single session lasting about 45 minutes. The session consisted of three phases: Training, screening, and the formal experiment.
Training. Listeners were informed that they would be hearing the first eight notes or so of a song, and that they were to identify

which song was being played. They were given a list consisting of the names of the eight melodies, and told that each song they would hear was on the list. They were then presented with the melodies played at 320/ms note in the following order: Camptown Races, Yankee Doodle, Rockabye Baby, Camptown Races, God Rest Ye Merry Gentlemen, Happy Birthday, Twinkle Twinkle, Little Star, Skip to My Lou, Love Me Tender. Note that Camptown Races was presented twice (as the first and the fourth item) to inhibit use of the process of elimination as an aid to the identification of melodies (the subjects had been informed that any of the eight melodies could be present at any time, and that melodies already heard could be presented again). The melodies started with the first note and terminated at the end of the first phrase, with two seconds separating repetitions of the melody. Subjects were allowed to listen for as long as they wished, and if their answers were incorrect, they were told the correct name and allowed to listen again. A list of the eight melodies was kept in view throughout the training phase. The list was removed before starting the screening phase.

Screening. After completion of the training exercise, subjects tried to identify each of the eight melodies played at 320 ms/note and repeated over and over without pausing between repetitions. Because listeners can sometimes guess the melody after hearing the initial notes (especially at slow and moderate tempos) the repeated melodies always started with either the third or fourth note. They were allowed to listen to each of the melodies for as long as they wished before attempting to name the tune. The order of melodies was determined randomly. Subjects were given no feedback concerning the accuracy of their responses. Only those who correctly identified six or more of the melodies were allowed to continue to the next part of the study. Six subjects were eliminated on this basis, leaving 30 participants in the formal experimental study.

Formal Experiment. The durational limits for melodic recognition were determined using the psychophysical method of limits. Half the subjects received each of the eight melodies at increasing durations (fast to slow) for their first set of judgments followed by decreasing durations of the melodies (slow to fast) for their second set of judgments -- and the other half of the subjects received the decreasing set followed by the increasing set of note durations. Different randomly determined orders for presentations of the melodies were used for the first and the second set of judgments. All melodies started on the third or fourth note and were recycled until the listener responded. They were instructed not to guess, but to name the song only if they were sure. If the subject indicated that the melody could not be recognized, or if an incorrect response was given, the next duration in the series was presented. This process continued until the correct response was obtained, or until the duration of 320 ms/note was reached. The durations used in the increasing order of presentation were: 40, 57, 80, 113, 160, 226, and 320 ms (steps increased by a factor of $\sqrt{2}$, so that two steps doubled duration). For the decreasing

order of presentation, durations were: 3600, 2560, 1810, 1280, 905, 640, 453, and 320 ms (two steps halved the duration). At the longest durations it took an appreciable time for a complete melody to be played, and subjects were required to listen to at least two complete statements before responding.

RESULTS

The experimental results are summarized in Table 2, and the temporal limits for melodic recognition for each tune is shown in Figure 1. The overall or grand median for the lower limit of melodic recognition was 160 ms and the upper limit was 1280 ms. These values approximate the limits of the duration of notes customarily employed for melodies (from roughly 150 to 900 ms) as cited by Fraisse (1963).

---Table 2 followed by Figure 1 about here---

DISCUSSION

At the longest item durations, the sequence of notes lost melodic cohesion and were heard as independent sounds. Reconstruction of the melodies was difficult for our listeners. As mentioned earlier, when the notes in a melody are played starting with the first note, it is sometimes possible to infer the identity of a tune played at moderate to slow tempos. This inferential recognition works best when based upon the initial notes of the melody, much as it is easier to complete a word if we are given the first two or three letters rather than the same number of letters found elsewhere in the word. Since our recycled melodies started in the middle, listeners did not know where to begin melodic construction. Some listeners who performed better than others at slow tempos reported that they remembered the intervals from one note to the next as they occurred, and then played them back from memory at a faster rate. We tested some additional people with training in musical notation who came from the same pool of introductory psychology students as those in the formal experiment. Most of them could reconstruct the melodies when played at the slowest rate, apparently using notation for musical intervals to facilitate recognition rather than relying upon direct memory of the sounds themselves. When the duration of notes fell within the range normally used for melodic themes, recognition was effortless for all listeners familiar with the tune, even when the recycled melody started in the middle. The initial melodic fragment could be ignored, and the complete melodic phrase was heard whether or not the listener had formal musical training.

At very brief note durations it was not possible to identify recycled melodies, even with musical training. However, the rapid series of notes appeared to form a distinctive pattern, although the pattern was not melodic. This observation suggested that metathesis of brief notes does not occur, and that the lower temporal limits for melodic recognition may be based upon special or-

ganizational rules for musical perception rather than general rules governing the perception of tonal sequences. Experiment 2 was designed to study the effects of item duration on the perception of nonmelodic sequences of tones.

EXPERIMENT 2: Perception of nonmelodic sequences of tones

Recycled nonmelodic sequences consisting of three to six tones have been used to determine the temporal limits for identification of the order of components (Bregman & Campbell, 1971; Nickerson & Freeman, 1974; Thomas & Fitzgibbons, 1971; Warren & Byrnes, 1975). These studies have found that the briefest durations permitting identification (that is, naming) of the order of component pitches along the continuum from high to low was between 100 and 200 ms per tone. However, it has been reported that permuted arrangements of recycled sequences consisting of qualitatively different types of sounds (e.g., noises, pulse trains, square waves) can be discriminated down to 5 or 10 ms per item (Warren, 1974a, Warren & Ackroff, 1976). At these brief item durations, it not only is impossible to identify the order of components, but the components themselves cannot be recognized. Nevertheless, these studies found that different arrangements of the same sounds were heard as qualitatively different, forming what have been called "temporal compounds" (Warren, 1974a), each compound having a characteristic quality.

Experiment 2 examined tonal sequence discrimination over a wide range of item durations (10 ms through 5 sec) using three tones which we call A, B, and C. There are two possible temporal arrangements of three tones starting and stopping with the same item: (ABCABC...A, and ACBACB...A). Listeners heard a succession of sequence bursts separated by brief silence and were asked whether alternate bursts were different (corresponding to ABCABC...A alternating with ACBACB...A) or the same (all bursts were ABCABC...A). Preliminary observations using three sinusoidal tones spaced at one semitone intervals had indicated that at item durations from 5 ms through 100 ms (which were below the threshold for identification of order), discrimination involved comparison of qualitative differences in temporal compounds formed by the component tones. At longer item durations, discriminating between the different arrangements was accomplished through naming the relative pitches of three components in their proper order (for example -- high, medium, low contrasted with high, low, medium). The range of item durations employed in the formal experiments extended both well below and well above the 100 ms threshold for naming of order, permitting a comparison of the effectiveness of temporal compound formation with the identification of components in their proper order as a means of discriminating between different arrangements of the same tones. In experiment 2a, the sinusoidal tones forming the sequences had frequencies separated by 9 semitones (the three items spanned a total of 18 semitones). This frequency separation is considerably greater than the two to three semitones approximating the width of a critical band (Dowling & Harwood, 1986, p. 83), so that each of these sinusoidal

tones stimulated separate populations of receptors situated in non-overlapping portions of the basilar membrane.

Dowling & Harwood (1986, pp. 155-156) have reported an interesting limit found in virtually all of the musical systems employed by different cultures. They stated the contiguous notes of melodies in these systems are separated by no more than about four or five semitones. In keeping with this generalization, Ortmann (1926) reported that use of large pitch ranges can result in perceptual grouping of tones by pitch contiguity rather than temporal contiguity. Thus, the interleaving of melodic lines in different registers has been employed by Baroque composers such as Bach and Telemann to produce the effect of two simultaneous melodies while using only a single instrument. This melodic segregation has been called "implied polyphony" by Bukofzer (1947), "compound melodic line" by Piston (1947), and "melodic fission" by Dowling (1973). As we shall see, such fission or "auditory stream segregation" of tones with their own restatements (see Bregman & Campbell, 1971) did not occur with our nonmelodic tone sequences.

In experiment 2b, each of the three recycled tones had the same waveform repetition period (and hence the same pitch), but different spectral compositions (and hence different timbres). Distinguishing different arrangements of these tones could only be based upon tonal quality or timbre, not tonal pitch. As we shall see, timbre appears as effective as pitch in permitting listeners to distinguish between permuted orders.

METHOD

Subjects. Participants were recruited from introductory psychology courses and received either course credit or cash for their participation. Students who passed the audiometric screening procedure described below were assigned to one of two experimental groups, each containing 36 subjects.

Audiometric Screening. All subjects participating in experiment 2 passed an audiometric test designed to eliminate individuals having hearing deficits, as well as listeners failing to follow the standard instructions for a Békésy threshold tracking procedure. The screening employed a tone presented diotically through a pair of matched headphones. Following presentation of instructions and a familiarization with the task, the tone was swept up from 400 Hz through 9000 Hz, and then down from 9000 Hz through 400 Hz at a rate of 1 octave/min, while subjects tracked their thresholds with a hand-held button switch (depressing the button decreased intensity at a rate of 2.5 dB/sec; releasing the button increased intensity at the same rate). An X-Y plotter produced audiograms consisting of continuous threshold tracings. Subjects were excluded from further participation if either directional sweep produced an audiogram differing from the 1964 International Standards Organization values by more than 22.5 dB at any frequency for the portion of the audiogram extending from 500

through 8000 Hz.

Stimuli. Only sinusoidal tones were used in experiment 2a: Tone A was 2794 Hz; Tone B was 1661 Hz; Tone C was 988 Hz. In experiment 2b, each of the three tones had the same pitch but they differed in harmonic composition and timbre. As in experiment 2a, each tone was produced by a separate signal generator. Tone A was an 800 Hz pulse train (pulse width 0.1 ms) with the spectral fundamental removed, Tone B was an 800 Hz sinusoidal tone, and Tone C was an 800 Hz square wave with the spectral fundamental removed. In both experiments 2a and 2b, "different" and "same" stimulus series were prepared for each note duration employed (10, 30, 100, 300, 1000, 3000, 5000 ms). The "different" series had bursts consisting of the tonal order [ABCABC...A] alternating with order [ACBACB...A], and the "same" series had successive sequence bursts all consisting of the tonal order [ABCABC...A]. All items within bursts had a rise/fall time of 2 ms with 1 ms of nominal silence separating the successive tones. Note that because of the special ease of identifying the first and last items of the sequence, sequence bursts always began and ended with the same item (A). Table 3 lists the parameters for the stimuli employed, giving item durations, the number of items (tones) in each sequence burst, and the interburst interval separating successive bursts (which were either identical or alternating in item order). The path of the signals through the equipment was identical for both the "same" and "different" order series (the order of items was determined by programs controlling the opening and closing of gates). Stimuli were bandpass filtered from 100 through 8000 Hz (filter slopes of 115 dB octave). Before recording for subsequent stimulus presentation, the input level of the individual components was adjusted so that each tone would have the same level (dBA) upon playback through headphones.

---Table 3 About Here---

Experimental Procedure. Listeners who passed the audiometric screening test were scheduled for their single experimental session. They were randomly assigned to serve in either experiment 2a or 2b, and were tested singly while seated in an audiometric room with the experimenter.

Subjects were told that they would be hearing patterns of sounds separated by brief silent intervals, and that their task was to determine if all bursts were identical or if successive patterns differed in any way. They were informed that the occurrence of same and different groupings would be randomly determined and were instructed to call out "same" or "different" at any time during the stimulus presentation. Subjects were encouraged to ask questions if any part of the instructions was unclear, and after both subject and experimenter were satisfied that the task was understood, the stimuli were presented in order of increasing item durations. This order was employed to avoid the possibility (discussed by Warren, 1974b) that identification of the order of components at longer item durations might be transferred to

briefer item durations through a recognition of qualitative similarities in successive sequences. Stimuli were presented diotically through matched headphones at a level of 70 dBA, as measured by a sound level meter with a 6 cc coupler. The experimenter operated the multitrack recorder (located outside the chamber) using a remote control unit equipped with a preset multipoint rapid search-to-cue device.

Before presentation of the experimental stimuli at each item duration, the subjects were given sample stimuli (first a "different" series, and then a "same" series), which were identified by the experimenter as same or different. They were told that if they wished, they could hear either of the known samples again before hearing the unknowns. When subjects indicated readiness, they were given a succession of three unknown sets at that item duration. The same and different unknowns were presented in a pseudorandom order with the constraint that all three of the unknowns presented to a subject at any item duration could not all be of a single type. Each subject was matched with another for whom the order of same and different presentations was reversed, so that equal numbers of "same" and "different" stimuli were presented for each experiment at each of the durations across subjects.

RESULTS

The number of correct responses at each duration was compared with the value expected by chance based on binomial expansion. There were 36 subjects and 108 judgments at each item duration. Tables 4 and 5 list the number of correct responses and the proportion of correct responses, as well as the Z scores showing the differences between the proportion correct and the proportion expected by chance at each item duration for experiments 2a and 2b, respectively. It can be seen that for both experiments response accuracy was beyond what could be attributed to chance performance at each of the item durations.

---Tables 4 and 5 About Here---

DISCUSSION

Experiment 2 demonstrated that untrained listeners can readily discriminate between different arrangements of three tones having item durations from 10 ms through 5 s. Accurate performance was observed both when pitch was varied and timbre was held constant (experiment 2a), and when pitch was constant and timbre varied (experiment 2b).

Let us deal first with discrimination of permuted orders based upon pitch. For durations above 100 ms/item, the literature indicates that different arrangements within recycled sequences of sinusoidal tones can be discriminated through the direct identification of the individual pitches and their orders (Bregman & Campbell, 1971; Nickerson & Freeman, 1974; Thomas & Fitzgibbons, 1971; Warren & Byrnes, 1975). Our sequences consisting of 100 ms

sinusoidal tones had durations slightly below the threshold for order identification, and although each of the individual tones could be identified, the order in which they occurred remained frustratingly elusive. Nevertheless, each of the arrangements had a distinctive quality, and discriminating between the two was easy. At the shortest item durations (10 ms), the individual tones could not be heard, and discriminating between the two arrangements was accomplished on the basis of differences in quality. These results are in agreement with the informal observations made in experiment 1 which suggested that, at durations below the threshold for melody recognition, listeners could still perceive a distinctive pattern associated with the rapid notes.

When the order of items are changed in three-component recycled sequences, then all of the transitions are changed: [...A→B→C→A...] is changed to [...A→C→B→A...]. A recent study (Bashford & Warren, 1988) employed complex sequences of ten 40-ms sinusoidal tones with minimal changes in temporal order -- only the positions of two contiguous tones were interchanged. Listeners were able to distinguish between the two arrangements of these complex tonal patterns within roughly a minute with an accuracy (90% or better) approximating that found in experiments 2a and 2b.

In experiment 2b, pitch was held constant and the items differed only in their timbre (the sinusoidal tone, the filtered pulse train, and the filtered square wave each had the same pitch). Examination of Tables 4 & 5 shows that discriminating different arrangements of pitches (experiment 2a) and different arrangements of timbre (experiment 2b) were accomplished with comparable accuracy. While sequences with differences in pitch are encountered in music and furnish a basis for melody, sequences differing in timbre are encountered in speech, and furnish a basis for distinguishing between different vowels. The ability to distinguish between different arrangements of 10 ms sinusoidal tones demonstrates that we can recognize particular pitch patterns far below the threshold for melodic recognition as measured in experiment 1. The discrimination of different orders of 10-ms tones differing in timbre demonstrates that we can recognize changes in overtones which are more rapid than those associated with speech -- the average duration of speech sounds in normal discourse is about 70 to 80 ms (see Warren, 1982, pp. 119-120).

SUMMARY AND CONCLUSIONS

Experiment 1 demonstrated that familiar melodies are recognized when the duration of notes is within the range normally used for the playing of melodies (roughly 150 ms through 1 sec). Experiment 2a demonstrated that each of the two possible arrangements of three recycled tones differing in pitch could be distinguished from the other for item durations from 10 ms through 5 sec. Experiment 2b employed three recycled tones each having the same pitch but differing in timbre. Again listeners were able to recognize same orders and distinguish different orders for item

durations from 10 ms through 5 sec.

These observations have implications for broader issues in hearing. These issues are discussed briefly below as a series of questions along with their possible answers.

1. Since listeners can recognize same orders and distinguish permuted orders of tones having durations as brief as 10 ms (see experiment 2a), why is it not possible to perceive that a pattern of notes having durations below the range of melodic recognition (say 100 ms/note) has the same pattern as a recognizable melody played at say 200 ms/note? There is evidence that at item durations below the threshold for naming order, the ability to recognize the equivalence of temporal order when a sequence is played at different tempos is limited by a "temporal template." Thus, if the rates at which a particular arrangement of notes are played differ by more than a critical value, then it is not possible to distinguish whether temporally mismatched sequences have their components in the same or different orders (see Warren, 1974b, for the durational limits of a temporal template).

2. How can listeners distinguish between different arrangements of the same tones when the items in a sequence occur too rapidly for their order to be recognized? We share with other animals the ability to recognize sequences of brief items as unresolved patterns or "temporal compounds" rather than as a succession of separate sounds (see Warren, 1988). The identification of component sounds having durations less than 100 to 200 msec is not required for comprehension of speech or the recognition of sequences of brief sounds in general (for further discussion, see Warren, 1982, Chapter 5). Melodies appear to have a special status, as discussed below.

3. Why is the time separating the onset of successive notes in melodic themes limited to a range extending roughly from 150 ms to 1 sec (corresponding to a rate of about seven per second to one per second)? A possible answer lies in the relationship between melodies and speech. Tunes are frequently associated with lyrics, so that familiar melodic phrases are often yoked to familiar linguistic phrases. Consider the eight tunes of experiment 1 -- if you know the melodies, you probably know the accompanying words as well. The syllable is the linguistic unit which usually corresponds to the individual notes of songs. The average duration of syllables in speech (radio newscasters) was measured to be about 170 ms (Lenenberg, 1967). It should be noted that lyrics are usually sung at a rate somewhat slower than normal speech, so that the melodic recognition range in experiment 1 is in keeping with a linguistic-melodic linkage.

Author Notes

1. Requests for reprints may be sent to Richard M. Warren, Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, 53201.

2. This study was supported in part by grants from the National Institute on Deafness and Other Communication Disorders (DC 00208) and the Air Force Office of Scientific Research (88-0320).

REFERENCES

Bashford, J.A., Jr., & Warren, R.M. Discrimination of recycled word-length sequences. Journal of the Acoustical Society of America, 1988, 84, S141 (Abstract).

Bregman, A.S., & Campbell, J. Primary auditory stream segregation and perception of order in rapid sequences of tones. Journal of Experimental Psychology, 1971, 89, 244-249.

Bukofzer, M.F. Music in the baroque era. New York: Norton, 1947.

Dowling, W.J. The perception of interleaved melodies. Cognitive Psychology, 1973, 5, 322-337.

Dowling, W.J., & Harwood, D.L. Music Cognition. New York: Academic Press, 1986.

Fraisse, P. The psychology of time (J. Leith, Trans.). New York: Harper & Row, 1963.

Lenenberg, E.H. Biological foundations of language. New York: Wiley, 1967.

Nickerson, R.S., & Freeman, B. Discrimination of the order of the components of repeating tone sequences: Effects of frequency separation and extensive practice. Perception & Psychophysics, 1974, 16, 471-477.

Ortmann, O. On the melodic relativity of tones. Psychological Monographs, 1926, 35, (1, Whole No. 162).

Piston, W. Counterpoint. New York: Norton, 1947.

Thomas, I.B., & Fitzgibbons, P.J. Temporal order and perceptual classes. Journal of the Acoustical Society of America, 1971, 50, 86-87 (Abstract).

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

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

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

Warren, R.M. Perceptual bases for the perception of speech. In M.E. Landsberg (Ed.), The genesis of language, Berlin: Mouton de Gruyter, 1988, 101-110.

Warren, R.M., & Ackroff, J.M. Two types of auditory sequence perception. Perception & Psychophysics, 1976, 20, 387-394.

Warren, R.M., & Byrnes, D.L. Temporal discrimination of recycled tonal sequences: Pattern matching and naming of order by untrained listeners. Perception & Psychophysics, 1975, 18, 273-280.

Winckel, F. Music, sound and sensation: A modern exposition. New York: Dover, 1967.

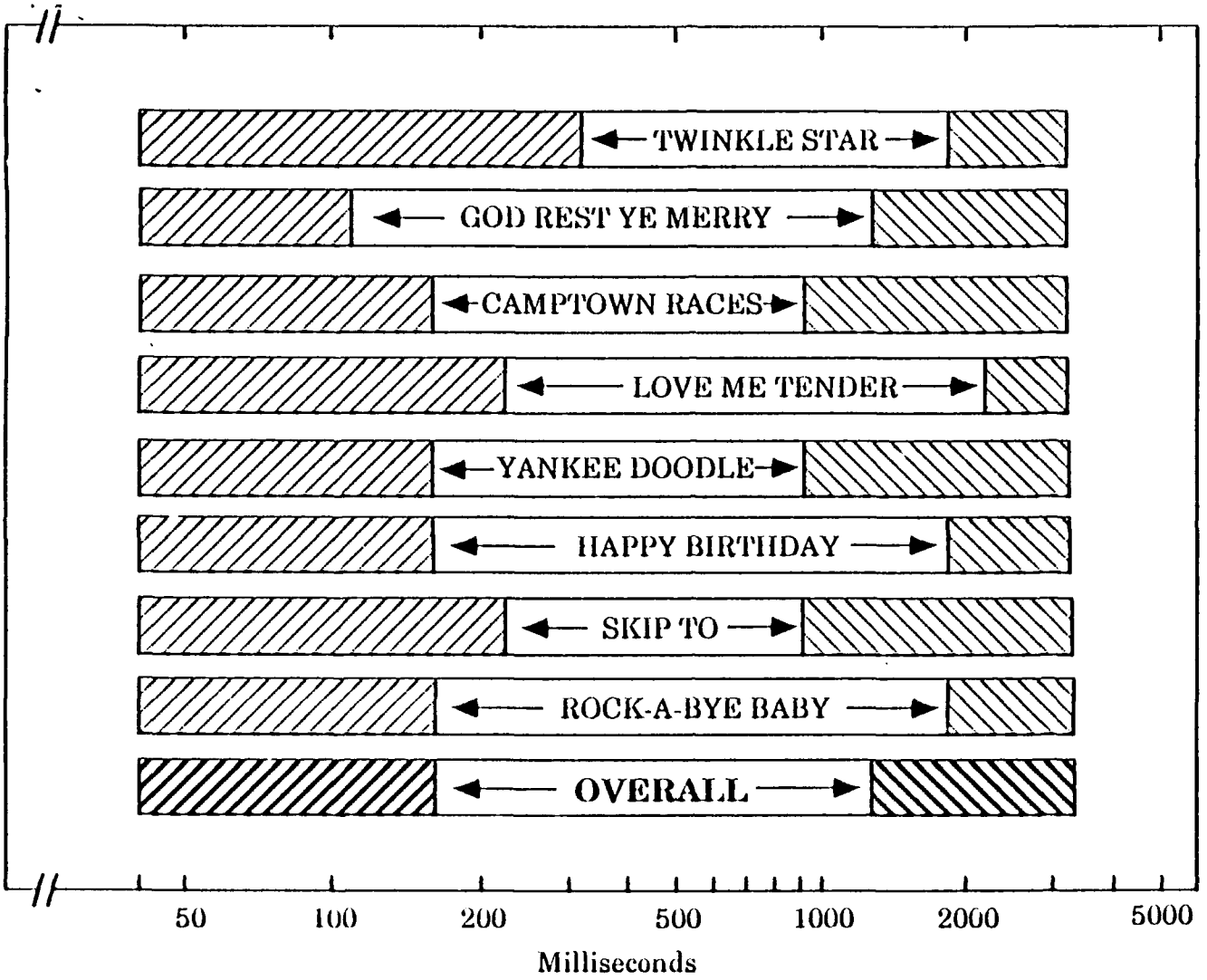


Figure 1. Note durations required for melodic recognition. The range of durations investigated ranged from 40 ms to 3.6 s. The arrows indicate the median limits for recognition of the eight melodies by thirty subjects without formal musical training. For durations below the recognition limit, it was possible to hear distinctive patterns for each of the sequences of notes, but the corresponding melodies could not be identified. For durations above the recognition limit, the notes were heard as individual pitches lacking melodic organization.

APPENDIX 3
To be published in: N.J. Lass (Ed.)
Principles of Experimental Phonetics
Philadelphia, B.C. Decker

Chapter 15: Auditory Illusions and the Perceptual Processing of Speech

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

To be published in: Norman J. Lass (Editor) PRINCIPLES OF EXPERIMENTAL PHONETICS
Philadelphia: B.C. Decker

Highly skilled perceptual processes, such as those employed for the perception of speech, seem to be immediate and direct, and as a consequence, the underlying mechanisms remain hidden during normal veridical perception. Illusions represent a breakdown in veridical functioning, and when used as experimental tools, can reveal these normally inaccessible processes. This idea is not new. In the 19th century Helmholtz stated that illusions provide a "particularly instructive" method for discovering the laws governing perception (see Warren & Warren, 1968, p. 140). An analogous principle is used in medicine, where an old maxim states that pathology lays bare the mechanisms used in health.

The following chapter describes research dealing with several interrelated illusions and their implications for normal perceptual processing. These illusions are (1) perceptual cancellation of masking and the illusory presence of obliterated sounds; (2) confusions in the temporal ordering of sounds; (3) illusory verbal organization of vowel sequences; and (4) illusory changes in repeated words. In addition, there will be some excursions into related areas.

ILLUSORY PRESENCE OF OBLITERATED SOUNDS: PHONEMIC RESTORATIONS AND AUDITORY INDUCTION

Our world is a noisy place, and when we are listening to a message, portions are often masked by louder sounds. However, we posses a mechanism capable of reversing the effects of masking and restoring obliterated segments. The process by which listeners restore portions of masked speech has been called *phonemic restoration* (Warren, 1970; Warren and Warren, 1970).

In the first study of phonemic restorations (Warren, 1970), listeners heard the sentence, *The state governors met with their respective legislators convening in the capital city*, in which the /s/ in *legislators* (the first "s") was deleted and replaced with a variety of sounds (cough, tones, buzzes). In order to minimize transitional cues to the identity of the missing sound, adjacent parts of the preceding and following phonemes were also deleted. Removal of this portion of the sentence would not be expected to have any effect upon intelligibility since the identity of the /s/ is evident from the information both preceding and following the deletion. It would seem reasonable to expect that listeners could, if they tried, readily identify the missing segment. However, it was not possible for them to identify the missing sound even when told that a speech sound was absent: The /s/ was "heard" as clearly as the phonemes actually present, and was perceptually indistinguishable from them. Attempts to identify the missing sound through localization of the extraneous sound were unsuccessful because this sound (whether a cough, tone, or a buzz) could not be localized accurately; it seemed to coexist with the speech sounds in the sentence at some poorly defined position.

The mislocalization of extraneous sounds in sentences had been reported by Ladefoged (1959), and later described in more detail by Ladefoged and Broadbent (1960). They employed brief sounds (clicks and short hisses), and since they were careful not to obliterate any phoneme, phonemic restorations could not arise. Warren and Obusek (1971) compared mislocalizations of a short click within a phoneme with the mislocalizations of a spliced-in cough (and other extraneous sounds) which completely replaced the same phoneme in an otherwise identical recording of the same sentence. We found that the direction of localization errors was rather different: When the sentence was heard for the first time, the longer sounds producing phonemic restoration were judged to occur earlier in the sentence than did the click. Apparently, a delay in perceptual organization occurring with phonemic restoration caused an earlier part of the sentence appear to be selected as simultaneous with the separate event corresponding to the extraneous sound. This delay in processing associated with phonemic restoration within a sentence heard for the first time should be reduced by familiarity with the sentence. As would be anticipated, replaying the stimulus to the listener resulted in a pronounced shift toward a later localization of the long extraneous sound, with judgments of its location having a more nearly symmetrical distribution

about the true position. Phonemic restoration itself was not reduced by replaying, and the listener's certainty that all speech sounds were present remained unchanged after four presentations. However, when extraneous sounds replacing the missing phoneme were eliminated and a silent gap was present, then phonemic restorations did not arise, and listeners localized the gap accurately and recognized which speech sound was missing.

A subsequent experiment showed that the context identifying the phoneme replaced by an extraneous sound need not precede the missing speech sound (Warren and Sherman, 1974). That study employed a variety of sentences each having some of the information required for identification of the deleted segments following their occurrence. To avoid the possibility that cues to the missing sounds were furnished by coarticulation involving the neighboring phonemes, the deleted speech sounds were deliberately mispronounced in each of the seven stimulus sentences as they were recorded initially prior to deletion. Care was taken that the mispronounced phonemes to be deleted matched the durations of the contextually correct items. Under these conditions, the contextually appropriate phoneme was restored according to the rules and restrictions of English, irrespective of any misleading coarticulation cues.

In initial studies of phonemic restorations, it was not recognized that they represent a linguistic adaptation of a more general ability to restore portions of signals obliterated by brief transient sounds. The restoration of obliterated sounds has been called *auditory temporal induction* and it has been suggested that there are three basic types of restoration (Warren, 1984). Homophonic induction was discovered when three levels of broadband noise, each lasting 300 msec, were presented sequentially (60, 70, 80, 60, 70, 80, 60, . . . 80 dB) [Warren, Obusek, and Ackroff, 1972]. Paradoxically, the faintest level (60 dB) appeared to be on all the time, coexisting with the two louder levels of the same sound. Homophonic induction did not require three levels, two levels would do, and the illusory continuity of the fainter sound was found to occur for noises and for tones at all intensity differences when two levels of the same sound were alternated. Induction appeared to be subtractive (that is, the continuous sound was subtracted from the louder sound). Thus, when the difference between the two levels was small (say 2 dB), the physically louder sound was heard as a faint pulsed addition to the louder continuous sound.

Heterophonic continuity consisting of the illusory continuation of a brief sound alternated with a louder sound of different spectral characteristics was discovered independently by Miller and Licklider (1950), Thurlow (1957), Vicario (1960), Houtgast (1972), and Warren et al. (1972) [for a review of this and related literature and issues, see Warren, 1984]. Houtgast as well as Warren et al. suggested that the continuity of one sound alternated with another of different spectral characteristics was related to masking, and that this continuity required that the neural units stimulated by the louder sound include the units stimulated by the fainter sound. Thus, auditory induction corresponds to a rather sophisticated mechanism, leading to the restoration of the signal if the stimulation by an interpolated sound could represent an extraneous sound plus the signal -- if the signal could not be masked by the louder sound, its absence would be detected and continuity would not be heard.

A third type of auditory induction was called contextual catenation. This variety of induction did not involve continuation of a steady-state signal, but rather the perceptual synthesis of a time-varying signal for which the restored sound differed from the portions of the signal preceding and following the interpolated noise. Phonemic restorations are one type of contextual catenation, as is the restoration of segments of tonal glides replaced by noise as reported by Dannenbring (1976), and also the restoration of missing notes of a familiar melody played on a piano which had been replaced by noise (Sasaki, 1980).

A somewhat different task involving both phonemic restoration and masked thresholds has been used by Samuel and his associates (Samuel, 1981a, 1981b, 1987; Samuel and Ressler, 1986). Listeners were required to discriminate between a syllable in which a phoneme is replaced by

noise (resulting in phonemic restoration at an appropriate noise level) and the same syllable in which the phoneme is present along with the noise (resulting in masking at an appropriate noise level). Using a signal detection methodology, a number of factors influencing the upper limit of masking (the phoneme detection threshold) were investigated, including attention, the interaction of bottom-up and top-down influences, and the lexical uniqueness of a word.

There is evidence that phonemic restorations appear to follow the general rules for auditory temporal induction which govern the restoration of other masked sounds. In accordance with the requirements for induction of nonverbal sounds, Layton (1975) found restorations to be enhanced when single phonemes in words and sentences were replaced by broadband sounds such as coughs or noises, but not when the phonemes were replaced by a sinusoidal tone. Samuel (1981a) used a signal detection paradigm, and also found that white noise produced greater restoration of affricative phonemes (consisting of continuous spectra as does noise) than for voiced sounds (consisting of line spectra and resonance-produced maxima or formants). Neither Layton nor Samuel and his coworker recognized the role of auditory temporal induction and masking functions in determining the extent of phonemic restoration.

A number of studies have employed multiple interruptions of speech by noise. As would be anticipated, multiple interruptions of word lists by silent gaps results in a decrease in intelligibility, along with a rough or harsh quality corresponding to the discontinuities. When the gaps were filled with a louder noise, the speech sounded more natural, but intelligibility was not improved. Quite different results were obtained for interrupted connected discourse by Cherry and Wiley (1967) and Wiley (1968), who found a great increase in intelligibility when silent gaps were filled with noise. Cherry and Wiley suggested that the addition of noise prevented a disruptive effect of silence on the natural rhythm of speech (however, it is not evident why noise would not also disrupt this rhythm). Powers and Wilcox (1977) were unaware of the earlier work on intelligibility of regularly interrupted speech described in Wiley's dissertation, and also reported that intelligibility increased when these gaps in speech were filled with noise. They reasoned that noise might improve intelligibility by removing transitions to silence, which could suggest the presence of stop consonants. But this explanation does not handle the observation of Miller and Licklider (1950) that the addition of noise to gaps in word lists did not improve intelligibility. A different explanation for the ability of noise to increase intelligibility was offered by Warren and Obusek (1971). We suggested that phonemic restorations occur when multiple gaps in connected speech are filled with noise, with contextual cues associated with syntactic and sentential context furnishing the information necessary for restoration. Since word lists lack this context, intelligibility is not increased by the addition of noise.

The evidence cited thus far indicates that filling multiple gaps in speech with noise can increase both apparent continuity and intelligibility. However, the relation between these two measures, and the relation of each of these measures to auditory temporal induction of nonlinguistic sounds has not been discussed. Bashford and Warren (1987) investigated the relation of auditory temporal induction to the illusory continuity of speech. One experiment in this study employed a narrowband speech passage (1/3-octave band centered on 1500 Hz). This filtered speech (based on a magazine article) was interrupted by 1/3-octave bands of noise having various center frequencies (375, 750, 1500, 3000 & 6000 Hz). The recorded speech and on-line noise were alternated regularly (equal durations for each), and the stimuli were presented at peak speech amplitude and average noise amplitude of 80 dB SPL. It was found that the longest noise duration for which continuity was observed (304 msec) occurred when the noise and the speech had the same center frequency. The continuity limit was less than half that value when the noise bands with the lowest and the highest center frequencies were used.

The second part of this study by Bashford and Warren examined the effect of contextual information upon the continuity of recorded broadband speech alternated with on-line broadband noise. It was found that the limit for illusory continuity was longest when the speech consisted of

connected discourse: The same passage read with the word order reversed and lists of unrelated monosyllabic words each had continuity limits only half as long as that found for normal discourse. The continuity limit (or discontinuity threshold) for discourse corresponded to the duration of the average word in the passage, while the discontinuity thresholds for backward reading and for the word lists were between the average duration for syllables and for individual phonemes of the stimuli.

The close correspondence between the average duration of a word in connected discourse and the limit of illusory continuity suggested a subsequent experiment by Bashford, Myers, Brubaker, and Warren (1988). A new passage of speech was recorded broadband and was interrupted by broadband noise. In agreement with the results of Bashford and Warren (1987), the discontinuity threshold once again approximated the average word duration. Separate groups of listeners then heard the recorded played at tape speeds which were 15% greater and 15% less than the original recording (the speech appeared normal when heard at each of the three rates). It was found that the faster playback changed the interruption threshold by -17.7%, and the slower playback changed the interruption threshold by +12.2%. The average of these two values for a 15% rate change (ignoring the direction of change) equaled 14.95%.

The most recent study on multiple phonemic restorations in our laboratory examined the effects of contextual constraints on intelligibility (Riener, Bashford and Warren, 1990). Three basic types of stimuli were used: Monosyllabic word lists, sentences having key words which had low contextual probability (e.g.,"Ruth hopes she called about the junk") where the key word scored for intelligibility was "junk", and sentences with high contextual probability (e.g., "Throw out all this useless junk") where the key word scored for intelligibility again was "junk". The scores for intelligibility using speech which alternated with silence of equal duration (200 msec), as would be anticipated, was least for word lists and greatest for high probability key words. When the gaps were filled with noise, there was no change in intelligibility for the word lists, but an increase in intelligibility occurred for the low probability key words in sentences (which had normal syntactic or grammatical context, but lacked semantic contextual clues to the identity of the key words). A still greater increase in intelligibility occurred for the high probability key words in sentences (which had both syntactic and semantic clues to the key word).

We have seen that filling gaps in speech with noise can increase both apparent continuity and intelligibility. When these effects take place, they both follow the rules governing nonlinguistic temporal induction. However, while the increases in continuity and intelligiblity are related, there are important differences. While noise does produce an increase in interruption thresholds for word lists (Miller and Licklider, 1950; Bashford and Warren, 1987), it does not produce an increase in intelligibility (Miller and Licklider, 1950; Riener et al., 1990). However, noise not only increases illusory continuity in sentences, but also increases sentence intelligibility (Cherry and Wiley, 1967; Holloway, 1970; Powers and Wilcox, 1977; Verschuure and Brocaar, 1983; Bashford and Warren, 1979; Riener et al., 1990).

Contralateral Auditory Induction. Auditory temporal induction fills in a missing signal across time, and auditory contralateral induction fills in a missing signal across space. Contralateral induction occurs when a signal at one ear is masked or replaced at the other ear by a louder noise: Under conditions producing complete contralateral induction, the monaural signal becomes completely delateralized, and is heard as centered on the medial plane. As we shall see, this illusion follows the rules for the restoration of masked signals described for temporal induction. In keeping with the rule governing illusions in general which was cited earlier, contralateral induction appears to reflect mechanisms normally leading to veridical perception. In this case induction appears to result from mechanisms employed as early stages in binaural processing.

A partial contralateral induction was reported by Egan (1948) who found that when speech was delivered to one ear and noise to the other, the position of the voice seemed to shift

toward the side receiving the noise. Thurlow and Elfner (1959) reported that a tone delivered to one ear moved toward the opposite side when the contralateral ear received a tone of a different frequency. Related shifts in lateralization were observed by Butler and Naunton (1962, 1964) when their listeners were stimulated by both a monaural headphone and a loudspeaker which was moved to various positions. The qualitative nature of lateralization shifts are easy to observe, but difficult to measure quantitatively for two reasons: Lateralized sounds seem to drift toward the medial plane as one continues to listen, and the exact extent of lateralization is difficult to measure because of the diffuse boundaries of the spatial image. However, there is a way of eliminating these problems.

Warren and Bashford (1976) reported a technique permitting precise measurement of delateralization by a contralateral sound. A signal (either a tone or the recorded voice of someone reading an article) was delivered to one ear and a noise to the other, and the stimuli were made to reverse sides each 500 msec (see Figure 1). This resulted in the signal appearing to be completely stationary at a diffuse location centered on the medial plane as long as the signal-to-noise ratio did not exceed a limiting value. Once this value was exceeded, then an abrupt change in the apparent location of the signal occurred, and it was heard to alternate from side to side as did the contralateral noise. Contralateral induction requires not only appropriate intensity relations, but also appropriate spectral relations between the signal and noise. It was found that the rules governing contralateral induction parallel the rules governing temporal induction; that is, if the monaural noise could mask the signal, then the signal becomes delateralized -- but if the noise cannot mask the signal, then the signal is heard on the side of stimulation. These rules apply to both verbal and nonverbal sounds.

---Figure 1 About Here---

First, let us consider the experiment of Warren and Bashford dealing with tones. As shown in Figure 2, sinusoidal tones of various frequencies were alternated (in a manner described in Figure 1) with three types of noise which were always delivered at 80 dB sound pressure level: (1) broadband (white) noise; (2) 1/3-octave band of noise centered on 1000 Hz; and (3) band-reject noise consisting of white noise with a spectral gap one octave wide centered on 1000 Hz. It can be seen that the upper limit of delateralization of the monaural signal occurred when the tone matched the frequencies present in the noise. The lower set of curves for masking were obtained using the same procedure, but instead of judging the position of the signal, listeners were instructed to adjust the level of the tone until it could just be heard. These threshold measurements showed that the masking of the contralateral signal by the various noises did not exhibit the same spectral effects as contralateral induction, and hence cross-ear masking artifacts could not be responsible for the spectral dependence of contralateral induction. Warren and Bashford than used monaural narrow-band filtered speech centered at either 1000 Hz or 3000 Hz and a contralateral narrow-band noise at 80 dB sound pressure level which was centered at various frequencies as shown in Figure 3. As can be seen, the upper limit for contralateral induction (expressed as Sensation Level or dB above threshold) was highest when the center frequencies of the voice and the noisebands were the same.

---Figures 2 and 3 About Here---

An interesting change from one form of induction to the other was observed when a monaural voice and a contralateral noise switched sides at rates between 30 and 200 msec: Contralateral induction was replaced by a monaural temporal induction at each ear, and listeners heard an unusual effect. The same voice seemed to be saying the same thing at the two ears at the same time without the identical simultaneous images fusing.

Delateralization through contralateral induction is a subtractive process as is temporal induction: That is, the components corresponding to the signal are subtracted from the inducer

when the signal moves to the medial plane. There is a condition involving uncorrelated noises which demonstrates clearly the subtractive nature of contralateral induction. First some background information: When uncorrelated noises (noises from separate generators) having equivalent long-term spectra are delivered to opposite ears at the same level, both noises are delateralized and heard as a single spatial blur centered on the medial plane. This effect has been described as the sound heard when standing under a waterfall, or under a tin roof in the rain (David, Guttman, and van Bergeijk, 1958; Kock, 1950; Warren and Bashford, 1976). Thus, while the two inputs cannot be heard separately, neither do they fuse to form a single compact image as when noise from a single generator is presented simultaneously to the two ears. Now let us return to the demonstration of the subtractive nature of contralateral induction. When, for example, an 80 dB noise is delivered to one ear and an 82 dB uncorrelated noise to the other, and the input to the two ears is switched every 500 msec as shown in Figure 1, listeners perceive two images. One is a diffuse stationary sound centered on the medial plane, and the other a faint lateralized residue corresponding to the difference in intensities between the two lateral inputs. (When 80 dB is subtracted from 82 dB, the remainder is 77.7 dB.) This fainter residue is heard to switch from side to side while the louder noise remains in an unchanging centered position. The moving residue can be made fainter or louder by decreasing or increasing the level of the noise having the greater amplitude. Of course, when the two levels are matched there is no residue, and switching sides of the two noises produces no perceptible effect on the centered diffuse image.

Warren and Bashford suggested that contralateral induction represents a first stage in binaural processing. When appropriate information is available at each ear, further processing can produce accurate estimates of the location of a speaker or other sound source (see Warren, 1982, pp. 31-58 for a discussion of mechanisms for localizing sources in space). Contralateral induction can also result in other types of binaural integration. Fused percepts (sometimes accompanied by lateralized residues) are possible with speech and other complex sounds which are not heard when the input to either ear is presented alone.

PERCEPTION AND CONFUSIONS OF TEMPORAL ORDER

Comprehension of speech, recognition of melodies, and many others aspects of hearing require the ability to recognize and distinguish different temporal arrangements of the same sounds. Some years ago, the literature on perception of acoustic sequences seemed quite tidy. Auditory temporal acuity appeared to analogous in some ways to visual spatial acuity: Resolving power could be measured in milliseconds with one, and in seconds of arc with the other. Hirsh (1959) and Hirsh and Sherrick (1961) demonstrated that detection of order within a pair of sounds such as a tone and a hiss could be accomplished for onset disparities as brief as 15 or 20 msec. These studies were verified in other laboratories, although thresholds were generally a little higher (see Fay, 1966, for a review of this early work). It was generally accepted that these experimental values were quite adequate to permit listeners to perceive the order in which the components of speech and music occurred. However, some subsequent observations concerning confusions and illusions in the identification of temporal order have led to new hypotheses concerning the nature of temporal resolution and its role in the perception of acoustic sequences.

There were a few curious reports in the 1950's which did not seem to fit the concept of a general threshold for the perception of order. Ladefoged (1959) and Ladefoged and Broadbent (1960) inserted brief extraneous sounds (such as clicks) in sentences, and found that listeners could not localize the position of the extraneous sound. (Although Ladefoged and Broadbent were careful not to remove or mask any phonemes, we have seen that errors in localization also occur for extraneous sounds completely replacing phonemes and producing phonemic restorations.) Heise and Miller (1951) described a similar phenomenon with tones -- if all but one of the tones in a series had frequencies close to one central value, then the odd tone seemed to "pop out" of the group, and listeners could not locate its position within the sequence. However, these anomalous observations were considered to reflect special attentional and informational process-

ing mechanisms associated with speech and music. Subsequently, another observation was reported which was quite difficult to reconcile with a concept of a general threshold for the perception of order, and which suggested a new look and reinterpretation ot the literature on sequence perception.

At a conference on pattern recognition, some unusual observations were described based upon a process for iterating or recycling sequences (Warren, 1968a). A loop of tape was constructed by splicing together 200 msec statements of a tone, a hiss, a buzz, and a vowel. When the loop was played back, the four items were presented in a fixed order which was repeated over and over. Subjects could listen for as long as they wished and could start naming the order with which ever sound they chose (there were factorial three or six possible arrangements of the four items). Even when each of the sounds could be heard clearly, the order could not be reported correctly. It was not that a wrong arrangement was perceived by the listener, but rather that no decision could be reached with confidence. Yet the duration of each sound was considerably longer than the accepted values for durations permitting renaming of order.

However, a repeated sequence of four successive items could be ordered correctly when each item was a 200 msec word. As part of the initial study, a tape was prepared which consisted of four spoken digits repeated over and over, each complete statement of the four items taking 800 msec. In order to eliminate any transitional cues to order, each of the digits was recorded separately before cutting and splicing into a loop. Despite the fact that each of the digits was itself complex (consisting of a sequence of phonemes), correct identification was accomplished easily by all of the listeners.

The last part of the initial study dealt with the identification of temporal order for vowels. Sequences of four vowels were constructed by cutting 200 msec segments from longer steady statements of each vowel and splicing the segments into a loop which repeated them without pauses. Although it was fairly difficult to judge the order, performance of a group of 30 subjects was significantly above chance. The task became easier when each vowel was reduced to 150 msec with 50 msec of silence separating items, and it was the easiest of all for single statements of each vowel with natural onset and decay characteristics for each statement, each vowel again lasting about 150 msec with 50 msec silences (Warren, 1968a). Subsequent work by Thomas, Cetti, and Chase (1970) with recycled synthetic vowels led them to conclude that the ease of identifying temporal order provided a possible method for measuring the speech like quality of synthetic speech sounds. Dorman, Cutting, and Raphael (1975) also used recycled synthetic vowels, but they introduced formant transitions resembling coarticulation effects between adjacent vowels for some sequences, and formant transitions resembling stop consonants separating successive vowels for other sequences. They concluded that the more their stimuli resembled sequences which could be produced by a speaker, the easier it became to identify the order of components.

The initial puzzling observations involving recycled sequences of three or four unrelated sounds (hisses, tones, buzzes) were examined further in a series of experiments dealing with the effects of experimental procedure on performance (Warren, Obusek, Farmer, and Warren, 1969; Warren and Warren, 1970; Warren and Obusek, 1972). When threshold durations permitting correct identification of order were measured, profound differences were found depending upon the method used for responding. When listeners called out the order of items (hiss, tone, buzz, and the vowel /i/), the threshold was found to be between 450 and 670 msec/item. But when different groups of subjects responded by arranging cards bearing the names of the sounds in the order of their occurrence, the threshold dropped to between 200 and 300 msec/item. Identification of order was made a little easier when the recycled sequences consisted of only three sounds (for which there are only two possible arrangements compared with six possible arrangements with the four-item sequences). With three item sequences, subjects can choose one sound as the anchor, and make only a single decision concerning the following sound. The ordering is completed with this single decision, since the remaining sound must precede the anchor. When this

process was used for card-ordering with three-item sequences, the threshold was 200 msec/item for sequences of nonverbal sounds.

Why is it not possible to identify the order of nonverbal sequences below 200 msec/item? Even though the sequences are nonverbal, the answer to this puzzle seems to involve linguistic processing time. Both Helmholtz (1887, 1954) and Garner (1951) found that the number of identical acoustic events within extended sequences could not be counted when the rate of occurrence exceeded five or six per second. Counting and the naming of items in the order of occurrence involves fixing distinctive verbal labels to successive events. It was suggested (Warren, 1974a) that the time required for verbal labeling sets the limit for both counting and the naming of order. This suggestion also explains why the threshold drops below 200 msec when recycled sequences consist of vowels or monosyllabic words. Verbal encoding of linguistic sounds would be facilitated not only because listeners are very familiar with these items, but primarily because the sound is the name -- that is, the names of the sounds and the sounds of the names are the same. Teranishi (1977) independently arrived at the same explanation proposed by Warren. Working with four-item recycled sequences consisting of either unrelated nonverbal sounds or Japanese vowels, he concluded that the rate-determining step in identification of order was the time required for naming, and that vowels can be ordered at high presentation rates because the sounds are equivalent to the verbal labels.

If verbal encoding time determines the threshold for identification of order, we would expect similar thresholds to be observed for visual sequences. Terence O'Brien and Anne Treisman in 1970 recycled three visual items (successive geometrical figures or successive colors) in a three-channel tachistoscope, and found the threshold for identifying order to be about 200 msec/item (personal communication). Sperling and Reeves (1980) presented a rapid series of digits on an oscilloscope and reported that although the digits could be recognized, their order could not be named. They concluded that this difficulty was "analogous" to that reported by Warren for sequences of sounds.

When recycled sequences consisting of the same sounds in different orders have item durations below the threshold for identification of order, they are not perceptually equivalent. Warren (1974) demonstrated that this ability to recognize and discriminate permuted orders involved recognition of the overall pattern, and did not require resolution of the pattern into an ordered series of items. In that study, sequences consisted of four sounds (tone, noise, /i/, and buzz). The items in Sequence A were always 200 msec (which was below the threshold for identification of order). Sequence B had one of eight item durations (127, 160, 200, 215, 315, 415, 515, or 600 msec), and the order of items was either identical to that of Sequence A (tone, noise, /i/, buzz, tone...) for "same" pairs, or had the order of the noise and the buzz interchanged for "different" pairs. Separate groups of 30 subjects were used with each item duration and the participants were instructed to indicate whether the order of items in Sequence A and Sequence B were the same or different. The results are shown in Figure 4. It can be seen that the accuracy of judgments was highest at 200, and 215 msec durations for Sequence B, with monotonic

---Figure 4 About Here---

decreases in accuracy at shorter and at longer item durations. Of especial interest is the decrease of accuracy at those longer item durations in Sequence B which permit listeners to identify the order within that sequence. If the accurate same/different order judgments obtained when the two sequences had similar item durations involved recognition of the actual order of components, then accuracy of judgments should not decrease when the duration of items in Sequence B were made longer. Instead, accuracy should be at least as good when identification of order for Sequence B was made easier. The results shown in Figure 4 were interpreted in terms of a *temporal template* which defines the extent of temporal mismatch permitting pattern recognition. This temporal specificity raises an interesting consideration which will become clearer if we turn

a moment to vision. It is possible to recognize a face despite changes in the visual angle subtended, so that the pattern of stimulation invoked by a retinal image need not match a stored template as a casting matches its mold. Thus, a smaller photograph may be placed alongside an enlargement of the same negative, and recognition of identity or difference is easy despite disparity in size. Since the temporal dimension in hearing is often considered as analogous to the spatial dimension in vision, we might expect that same/different judgments could be made over a wide range of temporal differences. However, the observed limits for temporal disparities permitting matching suggest that temporal templates play a role in auditory pattern recognition. Of course, with speech and music, some degree of durational flexibility is tolerated in pattern recognition. Recently, the durational limits for temporal templates have been measured for the melodic organization of notes (Warren, Gardner, Brubaker, and Bashford (in press).

Since same/different judgments were possible for pairs of sequences with item durations of 200 msec, would listeners be able to distinguish identical from permuted orders at briefer item durations? Warren and Ackroff (1976) used sequences of three unrelated sounds. In one of their experiments, they found that same/different judgments of pairs of recycled sequences (tone, square wave, noise) was accomplished with accuracy for separate groups of untrained listeners for matched item durations ranging from 5 msec through 400 msec (at 200 msec and above, sequences could be matched through direct ordering of their components). At brief item durations, it was not only order which could not be identified, the component sounds themselves could not be recognized. It was suggested that the elements of brief sequences formed *temporal compounds* which had properties characteristic of their particular temporal arrangements, and which could not be decomposed perceptually into their constituents (Warren, 1974; Warren and Ackroff, 1976; see Warren, 1982 for a detailed description of this concept and its implications). Recently, Warren, Bashford, and Gardner (1990) reported that listeners could distinguish between recycled sequences of three vowels for all durations employed ranging from 10 msec/vowel (single glottal pulses) through 5 s/vowel. As would be anticipated from studies described earlier, at durations above 100 msec/vowel, the components could be named with ease in their proper order, and different arrangements were discriminated on that basis. From 30 msec to 10 msec/vowel the sequences did not resemble speech, and discrimination was accomplished through differences in quality, such as one order appearing more "crisp" and the other more "dull." However, discriminating between different orders was accomplished through an especially interesting illusion for vowel durations ranging from 30 through 100 msec (corresponding roughly to the range of phonemic durations occurring in speech). The vowel sequences were heard as verbal forms -- that is, as syllables and words --- with different verbal organizations reported for different arrangements of the same vowels. Listeners were unable to hear the sequence veridically as a succession of vowels even when they knew the true nature of the stimulus. This illusion has a number of surprising aspects which are currently under investigation in my laboratory, as will be discussed in the next section. However, at this point I would like to point out that the vowel sequences are perceptually transformed into syllables consisting of consonants and vowels other than those actually present, but always following the clustering rules of English (thus, syllables starting with /ts/ or /sr/ are not heard). It is not possible for listeners to perceive the sequences as a succession of vowels even when they try. Yet if any of these vowels are heard in isolation at the durations resulting in verbal organization (30-100 msec), they can be identified with ease (we have observed that isolated vowels can be identified, with some difficulty, from single glottal pulses lasting 10 msec -- identification becomes quite easy with two or three of these glottal pulses). Apparently the presence of other contiguous vowels resulted in formation of verbal temporal compounds which could not be resolved into the actual components.

Is it possible that phonemes in normal speech also form linguistic temporal compounds? If so, then our ability to recognize phonemic components of syllables and words in normal speech might itself be an illusion. That is, listeners may recognize syllables and words as temporal compounds, and then infer the existence of the ordered series of phonetic components necessary to form these groupings. There is evidence along several lines that this is indeed the case. Savin

and Bever (1970) found that the identification time for a nonsense syllable was always shorter than that for a phoneme within that syllable (*identification time* seems preferable to *reaction time* since it appears that subjects do not react directly to phonetic components.) Warren (1971) confirmed these findings in an independent study completed shortly before their work was published. In addition to carrying out measurements on stimuli consisting of nonsense syllables, as did Savin and Bever, Warren systematically varied the level of intra- and intersyllabic organization of monosyllabic targets. These levels were: (1) a nonsense syllable list, (2) a word list, (3) sentences with the target word having a low contextual probability, and (4) sentences with the target word having a high contextual probability. At all levels, the identification time for a phoneme was always greater than that for the syllable containing the phoneme. The organizational level of a target word stimulus had surprisingly little effect on identification time for the word and for phonemes within the word up to level (3). However, when prior context made the target word's occurence more probable, then identification time for the word decreased. The point of interest to theory is that identification times for a phoneme within that word changed correspondingly, indicating that the phoneme identification was derived from a prior syllabic organization. The identification times for phoneme clusters were also measured and found to be always shorter than that for the entire syllable containing the cluster, and always longer than that found for a phoneme within that cluster.

This study also measured identification times for letter targets in the spelling of auditory stimuli. Spoken words were chosen with irregular spelling so that word recognition was required before target letters could be identified. The results obtained paralleled those obtained with phonemic targets, and suggested that, although letter identification for the spelling of a heard word and phoneme identification represented functionally separate processes, each derived from prior identification of the spoken word. Thus, the identification of words and their constituent phonemes appears to be a serial process with the higher level organization of a syllable or word preceding the phonemic analysis. There have been a number of subsequent studies which have confirmed the basic observation that identification of a spoken syllable or word takes less time than identification of a constituent phoneme (see Massaro, 1979 for a description of this work and a discussion of the implications for theory).

TWEAKING THE LEXICON: THE ILLUSORY ORGANIZATION OF LOUD AND CLEAR VOWEL SEQUENCES

As mentioned in the previous section, Warren et al. (1990) reported that sequences of steady state vowels repeated loudly and clearly can organize themselves into either real words or nonsense words. From roughly 30-100 msec per vowel it was not possible to perceive the sequence as a series of vowels, and an obligitory verbal organization took place. From about 100-150 msec per vowel, two perceptual modes were possible, and listeners could hear the stimulus both as an ordered series of vowels and as an illusory verbal form. Before describing these vowel transformations further, some background information will be given.

In the days before B. F. Skinner became a Skinnerian, he attempted to develop a projective psychological test which he described as "a sort of verbal ink-blot" (Skinner, 1936). He used repeated sequences of barely audible vowels, each having a duration of several hundred milliseconds (at these durations, the vowels and their orders could be perceived readily when the stimulus was made louder). Skinner found that after several repetitions of faint vowels, listeners heard what had originally sounded like three to five syllables of indistinct speech become intelligible, and they generally believed the words and phrases that they reported were accurate representations of the stimulus. He thought that this illusion involved a summation of meaning produced by repetition, and he therefore called his device *the verbal summator*. A rather different sort of illusory organization of vowels into words was observed by Dorman et al. (1975) when using recycled sequences of four vowels in a study designed to investigate thresholds for identifying vowel orders. They noted in passing that verbal organization interfered with their

listeners' task when the vowel durations approached the lower limit for the naming of order (about 100 msec/vowel).

The optimal vowel durations for perceiving the transformations into words is between about 30 msec and 100 msec/vowel (Warren et al. 1990). This study found that below 30 msec, vowel sequences did not sound like speech, and when the vowel durations reached about 100 msec linguistic organization became less pronounced and listeners started hearing the sequence veridically as a succession of steady-state vowels. When recycled sequences of three vowels were used, different words were heard for the two possible arrangements (/ʌæiʌ/... and /ʌiæʌ/...), and this verbal organization permitted easy discrimination between the two orders. A second experiment in this study employed vowel sequences which were much more complex, consisting of recycled sequences of ten steady-state vowels. There are factorial nine or 362,880 possible arrangements of these vowels, and 48 of these were selected randomly for the experimental stimuli. In the first part of this experiment, listeners heard sets of four of these vowel sequences which were labeled A,B,C and D. They were told to write down the words heard with each sequence, so that the sequences could be identified subsequently. The listeners were then presented with the same sequences again as unknowns. Matching of each of the four sequences with the corresponding verbal forms heard earlier was easy, and performance was near perfect. Most of the verbal forms heard (65%) were English words and phrases. The remaining verbal forms were nonlexical, but they all invariably followed the phoneme clustering rules of English. In the next part of this experiment, the same sequences were used to create 48 sequence pairs, members of each pair differing only in the transposition of two contiguous vowels. Listeners made ABX judgments -- that is, one member of the pair was designated as A, the other member designated as B, and they were required to tell whether the vowel sequence "X" was the same as A or B. Listeners again used verbal mediation, and were able to find different words characterizing each of the minimally different arrangements of the same components. Accuracy in identification of sequence X was near 100%. The task was repeated several days later with the same stimulus pairs, and performance was again nearly perfect. While there was only a slight agreement across subjects for the forms heard with a particular vowel sequence, slightly over half the forms reported by individual subjects on their second trial matched the forms they heard earlier for the same sequence.

A curious feature of vowel transformations is that only part of the acoustic signal is organized into a particular word. A residue is heard which occasionally is heard as a nonverbal noise, but more frequently is heard as a second stimulus word. This second word has a different timbre or quality, and appears to consist of different frequency components.

A subsequent study by Chalikia and Warren (1990) mapped the correspondence between the phonemes physically present and the illusory phonemes heard with recycled sequences of six 80 msec vowels. A number of procedures which seem appropriate for matching illusory phonemes with their stimulus counterparts do not work. The reasons they failed are of some interest. It might be thought that if one of the vowels was made louder, the listener could tell which portion of the illusory word increased in level. However, the listeners heard no change in the illusory words -- instead, a vowel having a louder level was heard veridically as a single vowel which could not be localized within the illusory word (the word itself did not change). Removing a vowel, and then listening for the part of the illusory word that seemed to become weaker or disappear failed to reveal that vowel's contribution to the verbal form, since the sequence was reorganized into a different form. Introducing a click as a marker for a particular vowel also failed, since as discussed earlier, Ladefoged (1959) and Ladefoged and Broadbent (1960) have shown that nonspeech sounds cannot be localized accurately within a word. But there was a technique that could be used successfully to determine the stimulus vowels corresponding to illusory speech sounds. This "mapping" of illusory words employed a technique used for locating the perceptual boundaries of phonemes in words by Warren and Sherman (1974). This earlier study had shown that when a sentence was terminated abruptly, the last speech sound could be identified with a precision which was often greater than that achievable using spectrograms.

Chalikia and Warren interrupted their vowel sequences at the beginning, middle and end of each vowel and asked listeners to report the last speech sound heard. An example of the results obtained for matching by a listener is shown in Figure 5. Prior to the experimental mapping of illusory words, each subject mapped a repeating real English word and a repeating nonsense word to insure that they could match the perceptual phonemes with the corresponding

---Figure 5 About Here---

physical phonemes of a repeated word. With these practice words, mapping was quite accurate as shown by comparing responses with the position of the interruption on sounds spectrograms. However, the terminal phoneme of the interrupted vowel sequence always maintained its identity as a portion of the illusory word. It is of interest that even when listeners who knew that the stimulus consisted of a sequence of vowels tried to identify the terminal steady-state vowel, they failed and could identify only its illusory transform within the illusory word. Considering this observation, why is it that listeners can identify accurately the terminal phoneme (vowel or consonant) in repeated meaningful or nonsense words? Does this accurate identification also involve a prior higher-level organization?

Chalikia and Warren also reported results obtained for mapping of the second (less salient) form heard along with the first. Usually, the primary form was of greater phonetic complexity than the secondary form. The two forms more often than not had no phonemes in common, and had their onsets corresponding to different stimulus vowels. As in the earlier report by Warren et al. (1990), the two illusory words seemed to be produced by different speakers, and had quite different timbres. This difference in timbre appears to be attributable to the organization of different spectral regions for the two voices.

Warren et al. (1990) offered some speculations concerning the illusory transformation of vowels into words. They started with the assumption that a listener hearing a sequence of speech sounds has a tendency to interpret the stimulus in terms of an utterance produced by a speaker. But how can a series of steady-state vowels be heard as a word? It was suggested that listeners employ a set of criteria or auditory templates for identifying a syllable or word, and that repetition produces a shift in the criteria employed for these templates. According to the *criterion shift rule* proposed for perceptual processing in general (Warren, 1985), the standards used for evaluating and classifying stimuli are displaced in the direction of simultaneous or recently experienced stimuli. In psycholinguistics, application of this rule would result in shifts in the perceptual boundaries of phonemes after listening to repeated syllables. Such shifts were first reported by Eimas and Corbit (1973), and since then they have been studied extensively. There is considerable controversy concerning the underlying processes responsible for these changes in phonetic boundaries (see Diehl, Klunder and Parker, 1985 for discussion), but there is little doubt that boundaries marking the perceptual existence region for particular speech sounds shift in the direction of the repeated stimulus. This shift in criteria may produce much greater effects if observed when repetition is in progress (as with the repeated vowel sequences) rather than after repetition ceases (as in the typical studies measuring the extent of category boundary shifts). It appears that repetition of vowel sequences can change the acoustic requirements for perception of a particular syllable or word to the point where the sequence is identified as a particular utterance. Similar conclusions were reached by Warren and Myers (1987) in their study of repeated syllables.

The perceptual matching of a vowel sequence to a particular syllable may be facilitated not only by criterion shifts, but also by the splitting of the stimulus into two concurrent forms corresponding to different spectral regions. It is suggested that one fraction is matched to the repetition-modified template corresponding to a syllable or word, while the other fraction corresponds to the residue remaining after the components corresponding to this match are subtracted (the fractionation of acoustic input into two portions has been discussed earlier under the

topics of auditory induction and phonemic restoration). The second fraction, or residue, is heard either as a nonlinguistic noise, or is matched to a second linguistic template and perceived as a different voice repeating a different utterance.

If one continues to listen to repeated vowel sequences after the initial organization into verbal forms takes place, illusory changes in what the voices seem to be saying can occur (Riener and Warren, 1990). Such changes have been studied in considerable detail for repeated words, and this "verbal transformation effect" will be discussed in the following section.

ILLUSORY CHANGES IN REPEATED WORDS

It has been known for a long time that continued stimulation with an unchanging pattern can lead to illusory changes or, under some conditions, to perceptual fading an disappearance. In the 17th century, John Locke noted that "*the mind cannot fix long on one invariable idea*" (Locke, 1690, 1894, p. 244). He concluded that any attempt by an individual to restrict his thoughts to any one concept would fail, and new concepts or modifications of the old "*will constantly succeed one another in his thoughts, let him be as wary as he can*" (p. 245).

Perceptually unstable figures have had a long history as designs. A mosaic floor depicting stacked cubes which reverse in apparent perspective has been uncovered at the Temple of Apollo at Pompeii. This design, as well as more intricate and ingenious reversible figures, can be seen in medieval and renaissance Italian churches. Perhaps the richest collection is on the floor of St. Mark's in Venice, providing a dynamic counterpoint for the nonreversible devotional figures found in the ceiling mosaics. However, these reversible designs do not seem to have provoked any scientific curiosity. But, in the early 19th century, Necker (1832) called attention to the illusory changes in apparent perspective of an outline drawing of a cube-like figure (a rhomboid), and tried to explain the inversions in terms of perceptual processes. Many reversible figures have been constructed and studied since.

All of the visual reversible figures are actually ambiguous; that is, they have plausible alternative interpretations. These interpretations employ the same contours as parts of separate perceptual organizations, so that at any given time, one interpretation precludes the other. While three or more interpretations may be possible (especially with the classical mosaics), most figures have only two. A consideration of these facts led me to look for an auditory analogue of the visual reversible figures.

If a person repeats a word over and over, he generally will experience a lapse of meaning called *semantic* or *verbal satiation* (see Titchener, 1915; Amster, 1964). It seemed that it should be possible to create an ambiguous verbal stimulus by repeating aloud a word such as *ace* over and over without pauses -- the stimulus should be acoustically equivalent to the repeated word *say*. Would perceptual alternation occur between these two plausible interpretations of the stimulus, and so prevent lapses of meaning? When I tried this for myself (as you can for yourself), such alternations seem to occur. These observations suggested the desirability of further work in which articulation by the listener would be avoided. In a preliminary study, Richard Gregory and I prepared short loops of recorded tapes containing single words. When we played these tapes to ourselves and others in the laboratory, we found that changes of the sort anticipated seem to occur. But surprisingly, compelling illusory changes in phonetic structure were observed as well, even though the words were played clearly and listeners knew each iteration was identical. Our bias was such that the note describing our observations was entitled "An auditory analogue of the visual reversible figure" (Warren and Gregory, 1958). However, after subsequent work, I concluded that passive listening to repeated words produces both phonetic and semantic lability. This is in sharp contrast with the effects of restating words to oneself, which produces only verbal satiation (or perceptual alternation with ambiguous words) without illusory changes in the phonetic structure. I came to believe that the auditory illusion based on listening to recorded

repetitions, which I named *the verbal transformation (VT) effect*, was not closely analogous to visual reversals (Warren, 1961a).

To illustrate the sorts of changes observed for VTs in this first detailed study, I will give a few examples obtained from subjects listening for three minutes to a loop of tape containing a clear statement of a single word or phrase repeated over and over. Subjects were instructed to call out what they heard initially and then to call out each change as it occurred, whether the change was to something new or to a form reported previously. The changes generally seemed quite real, and listeners believed that they were simply reporting what the voice was saying. The first example of illusory changes given below is based upon the stimulus *seashore* (since British naval ratings were used as subjects, a voice with standard English pronunciation was employed, and the terminal /r/ was not pronounced). The initial perceptual organization and all of the illusory changes reported by one subject during three minutes are listed in the order of occurrence: *seashore, sea-shove, seashore, she-saw, seesaw, sea-shove, seashore, she-saw-seesaw, seashore, she-saw-seesaw, seashore, she-sawve, seashore-seesaw, she-saw, seashore, seesaw-saw, seashell.* Another subject listening to *ripe* reported: *ripe, right, white, white-light, right, right-light, ripe, right, ripe, bright-light, right, ripe, bright-light, right, bright-light.* As a final example, a third subject listening to *fill-up* experienced somewhat fewer changes and greater phonetic distortion than most: *fill-up, clock, fill-up, build-up, true love, build, broad, lunch, fill-up.* It should be noted that changes which occur in going from one perceptual form to the next are frequently quite complex phonetically and sometimes suggest semantic linkages.

The main distinction between VTs and visual reversible figures revealed by this study seem to be: *(1)* visual reversible figures correspond to relatively few special configurations -- VTs occur with all syllables, words, and phrases; *(2)* reversible figures generally involve reinterpretation without appreciable distortion of the stimulus configuration -- VTs usually involve considerable distortion of clear auditory stimuli; *(3)* each of the reversible figures generally involve the same perceptual forms for different people -- VTs vary greatly with individuals; *(4)* reversible figures generally invoke changes between two (occasionally three or four) forms -- VTs usually involve more than four (sometimes more than a dozen) different forms during a period of two or three minutes. Yet, there is some relation between these two types of illusions. In broad terms, both seem to reflect Locke's principle that any particular "thought" or perceptual organization cannot be maintained without change for any length of time. There is a visual effect which seems to resemble verbal transformations more closely than do reversible figures. If the small eye tremors which occur continuously during normal vision are cancelled optically to produce a fixed or stabilized retinal image, then perception becomes unstable. As Wheatstone (1835) concluded (after considering the rapid perceptual fading of the shadows of retinal blood vessels when their images became stationary), the pattern of sensory input needs to change to maintain perceptual stability. Subsequent studies have shown that stabilized retinal images in general are seen to fade, fragment, disappear and sometimes reappear in a dynamic display of illusory changes (Riggs, Ratliff, Cornsweet, and Cornsweet, 1953; Pritchard, Heron and Hebb, 1960). While the selective suppression of portions of a pattern is found for verbal transformations as well, the auditory illusion produces a greater distortion and synthesis of physically absent elements than do stabilized retinal images.

Subsequent studies have indicated that verbal transformations may be of value as a tool for studying speech perception. No attempt will be made to cover all aspects of VTs which have been reported in the more than 50 experimental papers published on this topic. Interested readers can find reviews of this literature by Warren (1968b, 1976, 1982). The discussion below will deal briefly with acoustic and phonetic factors involved in VTs, and (in somewhat more detail) implications concerning the mechanisms employed normally for perceptual processing of speech sounds. In addition, the relation of VTs to other phenomena in speech perception will be described.

The first detailed phonetic analysis of the verbal transformation effect was that of Barnett (1964). After using a variety of words as stimuli, she concluded that the articulatory positions of both vowels and consonants were relatively labile and subject to frequent illusory changes. Stability was noted for the voiced-voiceless property of consonants and the type of movement characteristic of individual consonants and vowels. Intervowel glides were, in general, stable both in position and type of movement.

A detailed study of the nature of phone-type substitutions by linguists and nonlinguists (each group consisting of native and non native speakers of English) listening to the repeated word *cogitate* was reported at a meeting of the Acoustical Society of America (Naeser and Lilly, 1970). In an unpublished manuscript based upon this paper, they reported that linguistics and nonlinguists gave similar responses. It was stated that consonants generally were substituted by the manner of articulation (not place) so that, for example, stops most often were substituted by other stops. On the other hand, vowels most often were substituted on the basis of similarity of place of articulation. A resemblance was noted to the articulatory feature-type substitution described by Wickelgren (1965, 1966) in his work involving errors in short-term memory. More recently Lass and Gasperini (1973) reported a study comparing verbal transformations for a number of stimulus words presented to phonetically trained and phonetically untrained subjects. They noted some quantitative differences between the two groups, but emphasized that responses were qualitatively similar. The phonetically trained group reported more forms and transitions, and required fewer repetitions of the stimuli to induce the first illusory change.

Clegg (1971) used 18 separate repeating syllables, each consisting of a different consonant followed by the vowel /i/. He ignored the illusory changes of the vowel, which he stated were minimal, and analyzed the transformations of the consonants. His analysis was in rough agreement with Naeser and Lilly's (1970) but considerably more detailed. He concluded that a consonant and its transform tended to share the features of voicing, nasality, and affrication, but not of duration and place of articulation.

Evans and Wilson (1968) used a variety of consonants followed by the same vowel as stimuli for VTs. They analyzed responses only for changes in the initial consonant and reported a surprisingly high frequency of responses involving the aspirated phoneme /h/. Goldstein and Lackner (1974), in a more comprehensive phonetic study, used a variety of nonsense syllables as stimuli and also found a large number of responses involving illusory /h/ as well as /j/ in both the initial and final positions of the syllables reported. However, they were not interested in these intrusions and constructed matrices for illusory transformation of consonants and vowels for which /h/ and /j/ were excluded. Analysis of these matrices in terms of distinctive features revealed a number of "very systematic" types of changes governing vowels and consonants which were summarized by Goldstein and Lackner. However, subsequent work by Lackner, Tuller, and Goldstein (1977) suggested to them that feature detectors were not involved in speech perception. This belief was not shared by Ohde and Sharf (1979) who attributed VTs to adaptation of feature detectors which respond selectively to particular acoustic aspects of the repeated stimulus. More recently, Debigaré (1984) has proposed that VTs are a consequence of changes in cortical "cell assemblies" which had been proposed by Donald Hebb.

Lass and Golden (1971) employed repeating stimuli consisting of short segments of tape excised from recordings of steady-state vowels. Onset and decay characteristics of single utterances of the vowels were lacking, so that the stimuli differed from normal speech productions. A high proportion of nonphonetic alternatives were reported (such as a telephone busy signal), perhaps reflecting the difference of the stimulus from normal speech productions. Changes usually involved illusory consonants, generally plosives, possibly do to the rapid onset and termination of the vowels. No analysis in terms of distinctive features was offered.

The effects of noise upon VTs are curious. When listeners hear a voice which is not quite

intelligible (either because of its faint level or the addition of masking noise to a voice which would otherwise be intelligible) the rate and variety of VTs are quite different from those observed with loud and clear repeated words. The first observations of the effect of repetition upon unintelligible verbal stimuli were reported by B.F. Skinner (1936). In the early days of his long career, he was interested in developing a projective test -- as he put it, "a sort of verbal ink-blot" (p. 71). He used a phonograph recording of a series of faint and indistinct vowels played over and over. After several repetitions, the listener perceived illusory words and phrases. Since Skinner considered that repetition caused meaning to summate, he called the device *the verbal summator*. He turned the device off after the first response (as did later investigators using this technique). Had the stimulus been left on by Skinner, he would have discovered that illusory changes take place. When we allowed repeated indistinct speech to remain on, it was observed that the initial perceptual organization was unstable and VTs took place (Warren, 1961a). Common sense suggests that a word which is heard less clearly should change more readily, but such is not the case. Partially masked speech has a considerably lower rate of VTs than clear speech.

Sadler (1989) found that noise *per se* did not influence VTs. Rather it was the level of voice above threshold that determined the nature of verbal transformations, so that the effects of noise could be duplicated by decreasing the amplitude of the voice in the absence of background noise. Once the voice became intelligible, further increase in level had little effect on either numbers of transitions or forms. However, when the intelligible voice was made still louder, transitions from one form to the next involved fewer phonemes, and (somewhat surprisingly) the ratio of nonsense forms to lexical forms increased.

Is it possible to observe any illusory changes while listening to repeated nonverbal stimuli? A number of experimenters have reported such illusory changes, but they seem rather different in important respects from VTs Repetitions of white noise bursts were used by Lass, West, and Taft (1973); tone bursts by Fenelon and Blayden (1968), Perl (1970), and Lass, West, and Taft (1973); and melodic phrases by Gilford and Nelson (1936), Obusek (1971), and Lass, West, and Taft (1973). The changes reported in these studies generally were slight alterations in loudness, pitch, and tempo, with a rate of such changes sometimes similar to, and sometimes slower than the rates corresponding to VTs. However, experimenters generally have ignored the fact that tape recorders to produce variations in intensity (loudness) and speed (pitch and tempo). Even high quality professional recorders can have a moment-to-moment changes of about 1 dB in intensity and about 0.3% in both record and playback speed, which are at or above the just-noticeable differences for "unchanging" stimuli. Anyone who has listened to a tape recording of an extended tone (e.g., 1000 Hz) rather than on-line output from an oscillator probably has encountered this stimulus instability. This real change should be kept in mind when evaluating the fact that reported changes with nonverbal stimuli all seem to involve relatively subtle changes along the perceptual continuum of loudness, pitch, and tempo, rather than we characteristically gross and categorical suppression, synthesis, and transformation of individual sounds observed with VTs. Of course, it may be that illusory variations are introduced with repetition of nonverbal stimuli which go beyond those corresponding to the instability of recordings. Perhaps a general perceptual lability of unchanging patterns has been specially modified for speech. As we shall see, the verbal lability seems related to processes normally leading to an improved intelligibility. Studies of age differences for VTs have suggested that this illusion employs mechanisms aiding the comprehension of speech, and these mechanisms change systematically from childhood through old age.

The same set of stimulus tapes was used throughout cross-sectional age studies of VTs in our laboratory. Experiments with children (Warren and Warren, 1966) showed that virtually no illusory changes were experienced at five years of age. At the age of six years, almost half the subjects tested heard illusory changes, and those who did, experienced them at the high rate found for older children. At age 8, all subjects heard illusory changes (average rate for all subjects with all words was 34 changes for three minutes). This rate remained about the same (32

changes in three minutes) at the age of ten years. In an earlier study (Warren, 1961b), the average rate of change for young adults (18-25 years of age) was equivalent to that of older children (31 in three minutes). However, this study also revealed that the rate for aged adults (62-86 years) was very much lower (5.6 in three minutes). Using subjects with a median age of 35 years and different stimulus words, Taylor and Henning (1963) reported a rate of illusory changes intermediate between that for young and aged adults reported earlier. This suggested that the decrease in susceptibility during adulthood occurs gradually. The decrease in VTs with older adults does not seem to reflect a decrease in auditory acuity with age, since the aged not only maintain stability in what they hear, but they are more accurate, generally reporting the correct word and staying with it. Continuing to look for an explanation for the reduced rate of VTs, one might think that an increase in so called "neural noise" associated with aging would reduce the effect of signal-to-noise ratio for a given sound pressure level, and that this reduced ratio is responsible for the difference between young and old adults listening to a stimulus delivered at the same intensity. However, as noted earlier, once a signal is intelligible, changes in the signal-to-noise ratio have only small effects upon the rate of VTs. A suggestion concerning the basis for the decrease with aging will be offered after we have considered some additional factors.

In addition to counting the numbers of illusory changes, the groupings of speech sounds were examined to determine the functional rules governing reorganization at different ages. Children respond in terms of the sounds of English, but they may group these sounds in ways not permitted in the language. For example, with a repeated word *tress*, a child may report *sreb*, although the initial /sr/ sequence is not found in English words. Young adults seem to group speech sounds only in clusters permitted in English, but they do report nonsense syllables: Given a stimulus *tress*, they might report *tresh* as one of the forms they hear. However, older people tend to report only meaningful words. Presented with *tress*, they will typically hear *tress* continuously, and when infrequent changes do occur, they usually are to such closely related forms as *dress*. Even when presented with a nonsense syllable such as *flime*, the aged usually did not report the stimulus but distorted it into a meaningful word such as *flying*, frequently hearing the distorted word throughout the 3 minute test. An interesting exception is the incorrect past participle *flyed*, reported more frequently than the actual stimulus by the aged.

When the extent of phonemic differences between successive forms reported by individual subjects was analyzed, a regular change with age was found. For the five age groups (subjects ranging from 6 through 86 years of age), the number of phonemic changes decreased with increasing age. Thus, perceptual reorganization involved finer and finer distinctions with advancing age. Such changes seemed to be associated with normal, healthy aging; the extent of phonemic changes observed for a senile group of subjects approximated that of 10-year old children (Obusek and Warren, 1973a).

These observations of the effect of aging upon VTs, together with other considerations to be dealt with shortly, suggest that: *(1)* this illusion reflects reorganizational processes normally leading to the correction of errors in speech perception; and *(2)* the age differences observed for VTs mirror changes in processing strategies over our lifespan in keeping with changes in functional capacity. Since performance as measured by comprehension may be similar at different ages, it is all too easy to assume that the same perceptual processes are used. However, it may well be that it is necessary to change processing mechanisms to maintain performance accuracy over one's lifespan. In other words, adaptive changes in perceptual processing may be requisite for healthy maturing and aging.

There have been suggestions that repetition of specially selected "reversible" words may produce changes more closely linked to perceptual reversals observed with such visual illusions as the Necker Cube (Esposito, 1985; Roivainen, 1989). Roivainen, working with repeated bisyllables found that changes involving a reversal in the order of syllables predominated, and that the rate of these changes was only slightly slower for his aged subjects (64-86 years of age) then in that

observed for young adults. This is in keeping with observations that changes in visual reversible figures decrease only slightly with age, if at all (Miles, 1934).

Roivainen considered that the great decrease in rate of VTs for nonreversible words in the aged resulted from the lack of some specific speech mechanism in this age group. The matter is at present unresolved, but perhaps changes in the order of syllables in bisyllabic reversible words should be considered as fundamentally different from verbal transformations.

Before dealing further with age differences and processing mechanisms, let us first consider the evidence that reorganizational processes are necessary to correct errors made while listening to speech. Bryan and Harter (1897, 1899) claimed that skilled telegraphers used a "telegraphic language" which was similar to other languages. Mastery required several years of continual application, perhaps ten years for the speed and accuracy required for a press dispatcher. When this peak was obtained, the receiver could work effortlessly and automatically, often transcribing complex messages while thinking about something quite different. The expert usually delayed 6 to 12 words before transcribing the ongoing text in normal messages. If redundancy and linkages between elements was reduced by transmitting in cipher, or by sending stock quotations, the task became much harder for expert receivers. Understanding this, the sender slowed down the transmission rate, and the number of words held in storage by the receiver was reduced by decreasing the delay between receiving and transcribing. It seems as if long storage was used only when context permitted useful interactions between information received at different times, so that errors could be corrected, and ambiguities resolved. Skilled storage similar to that observed for telegraphy has been observed for typewriting (Book, 1925), for reading aloud (Huey, 1968) and for tactile reading by the blind (Moore and Bliss, 1975). It seems probable that storage with continuing processing and revision is important for a listener's comprehension of speech as well. The need for such mechanisms has been noted in the past (although the literature has been rather silent on how to study these covert processes). Brain (1962, p. 209) has pointed out that "The meaning of a word which appears earlier in sentence may depend upon the words which follow it. In such a case, the meaning of the earlier word is held in suspense, as it were, until the later words have made their appearance." Lashley (1951, p. 120) has given a classic example of this process. He spoke of *rapid writing* to an audience and after creating a set for the word *writing*, several sentences later stated "Rapid righting with his uninjured hand saved from loss the contents of the capsized canoe." He pointed out that the context required for the proper attribution of meaning to the sounds corresponding to *righting* were not activated for 3 to 5 seconds after hearing the word.

It is not only meaning that can depend upon future context. Chistovich (1962) has noted that subjects, who repeated speech heard through headphones as quickly as possible, made many phonemic errors. She suggested that these mistakes reflect the temporal course of speech identification, with an appreciable delay being necessary for the correction of such errors. Miller (1962) and Lieberman (1963) have emphasized that skilled speech perception can not be a simple Markovian process, with perception occurring first on phonemic and then on higher levels. Such a process does not take advantage of the redundancy of the message and does not allow for the correction of mistakes. Without such correction, an error would continue to provide incorrect context, producing other errors until perception became completely disrupted.

Returning to VTs, I am suggesting that these illusory changes reflect reorganizations which occur normally when part of a continued message is not confirmed by context. With repeated words there can be no stabilizing grammatical and semantic environment provided by surrounding words. Hence, the repeated word is subject to successive reorganizations, none of which can receive contextual confirmation. It is important to note that the processes involved are quite automatic and under little cognitive control. They are not accessible through introspective search, but only through their perceptual effects.

The age differences observed for VTs are consistent with the explanation just offered. If VTs reflect skilled reorganizational processes, they could not appear in children until language skills had attained a certain level. The requisite level seems to be reached normally by the age of 6 or 7. Certainly, the healthy aged (over 60 years of age) have mastery of language, so why then do they exhibit a reduced susceptibility to VTs? The answer may be that they lack the requisite capacity for short-term storage of verbal information. It is well established that special difficulty is encountered by the aged for complex storage tasks involving intervening activity (Welford, 1958). Concurrent processes of coding, storing, comparing and reorganizing may not be possible, so the optimal strategy for the aged would be to employ only the previous context in organizing current input, and to abandon reorganization contingent upon subsequent context. The fact that more meaningful words are reported by the aged Warren (1961b) is consistent with the fact that they could not risk being locked to a meaningless word. VTs, when they do occur with the healthy aged, usually involve changes of a single phoneme to form another meaningful word. Such a change applied to discourse would involve less complex phonetic rearrangement than the typical changes of the young. The extensive phonetic changes observed for senescent aged in their (infrequent) VTs do not represent a strategy capable of optimizing their capabilities: While it might lead to meaningful organization when applied to discourse, the meaning may not be the same as the speaker's (Obusek and Warren, 1973a).

It has been suggested that the verbal transformation effect and the phonemic restoration effect are linked, each being related to mechanisms employed normally for the correction of errors and the resolution of ambiguities (Warren and Warren, 1970). Obusek and Warren (1973b), acting on this suggestion, combined these two illusions by presenting a repeated word ("magistrate") with a portion deleted and replaced by a louder noise. If illusory changes are indeed corrective, they would be expected to be directed to the phonemically restored segment. When the /s/ of "magistrate" was removed and replaced by a louder noise, 42% of the illusory changes involved the position corresponding to the deleted /s/, compared to 5% when a different group of subjects heard the intact word as the repeated stimulus. When the /s/ was deleted and replaced by silence rather than noise (so that phonemic restorations were inhibited), the lability of phonemes heard in this position was greater than with the intact word, but much less than that associated with the presence of noise: Only 18% of the changes involved the position of the gap. When the noise was present, it was not possible for a listener to detect which phoneme corresponded to the noise bursts anymore than with phonemic restorations within otherwise complete sentences (Warren, 1970). Yet, it appears that, at some level of processing, the distinction between speech-based and noise-based organization was maintained, and the perceptual reorganization was directed to the portion lacking direct acoustical justification.

We have seen the phonemic restorations can be considered as a special linguistic adaptation of the general phenomenon of temporal induction. It is also possible to consider verbal transformation as a special linguistic adaptation of broader perceptual rules. Two general principles appear to be involved in verbal transformations: (1) the loss (or "satiation") of a particular perceptual organization resulting from a continued exposure to a stimulus; (2) the emergence of a new verbal form produced by a shift in auditory criteria corresponding to that form with continued stimulation (see Amster, 1964, for a discussion of verbal satiation, and see Warren, 1984, for a discussion of criterion shifts). As described earlier, shifts in the criteria corresponding to a particular verbal organization also occur with repeated vowel sequences. These shifts move templates employed for syllabic or lexical recognition in the direction of the repeated exemplar up to the point where the stimulus itself meets the changed criteria. To summarize the hypothesized sequence of events leading to VTs: With continued repetition any particular verbal form becomes satiated, shifts in perceptual criteria result in the perceptual emergence of a different form, and then this form in its turn undergoes satiation and replacement.

It is possible to stimulate each ear with the same repeated word without hearing the word as a single fused image. In an experiment by Warren and Ackroff (1976), the stimulus word

"tress" with a repetition period of 492 msec was passed through a digital delay line with 2 outputs, and the delay between these outputs was adjusted so that the asynchrony was exactly half the repetition period (246 msec). Listening was through a pair of headphones wired separately so that the temporally asynchronous but otherwise identical stimuli were heard in each ear. Neither ear could be considered as leading with the half-cycle delay, so that there was a lateral symmetry in the nature of simultaneous contralateral input. Since the interaural asynchrony was a few hundred milliseconds, there was no possibility of binaural fusion, and all listeners initially perceived the stimulus as two identical voices saying the same thing on the right and left sides. The question of interest to theory was whether or not the same forms would always be heard simultaneously on the left and on the right. If changes occurred which were identical and simultaneous, it would indicate that there was a sharing of a single set of neural linguistic processors. If independent changes were to occur at each ear, then it would indicate that two sets of functionally separate processors were used for the acoustically identical verbal stimuli.

It was found that verbal transformations occurred at each side, and that the changes occurred at different times. Also the forms heard at the two sides were independent, so that while the stimulus word "tress" might be perceived accurately at one ear, a word as far removed phonetically as "commence" might be heard at the other. Additional unpublished experiments by Warren and Bashford demonstrated that independent changes of the same word were not limited to only two competing versions. Listeners were presented with 3 asynchronous versions of the same repeated word, each separated from the other two by exactly one-third of the word's duration. Two versions corresponded to monaural inputs (one on the right, and one on the left), and one was diotic (synchronous stimulation at each ear) with its apparent location lying in the medial plane. Listeners heard the three versions of the word (right, left, and center) change independently, and hence three functionally independent sets of linguistic processors were employed for the same stimulus. These observations indicate a degree of equipotentiality for cortical units employed in speech processing, so that if one set of processors is occupied with a particular task, others can be assembled for analyzing identical phonetic sequences, at least up to the level of the word.

SUMMARY

It appears that some basic auditory mechanisms employed for nonverbal patterns have been modified in special ways for the comprehension of speech. Temporal auditory induction permits listeners to restore verbal and nonverbal sounds masked by intermittent louder sounds through a detailed comparison of the spectra and intensities of the fainter and louder sounds. The verbal form of temporal auditory induction (phonemic restoration) employs syntactic and semantic information in reconstructing obliterated fragments. Multiple phonemic restorations can restore intelligibility to speech which would otherwise be unintelligible, and can also produce an illusion of continuous speech having no missing segments.

Another kind of auditory induction occurs when a signal is heard at one ear, but obliterated by a louder sound at the other. Contralateral auditory induction permits restoration of the monaurally masked signal so that mislocalization to the side of the unmasked ear does not occur. The rules concerning spectral and intensive requirements for signal restoration are similar for temporal induction and for contralateral induction. It appears that contralateral induction corresponds to an early stage in binaural interaction. Additional detailed comparison of the input to the two ears can result in appropriate localization of a speaker or other sound source, as well as producing percepts which are qualitatively different than those produced by either of the lateral inputs when heard alone.

Experiments with speech and with sequences of nonspeech sounds such as hisses, tones, and buzzes have indicated that much of what seems to be direct identification of temporal order is illusory, and is actually based upon prior identification of a larger pattern. Two principle

mechanisms permit us to discriminate between different arrangements of the same items, whether the sequences are phonetic groupings, tonal groupings, or groupings of arbitrarily selected sounds: (1) recognition of overall patterns capable of operating down to item durations of a few milliseconds; and (2) naming of items in order of occurrence, allowing direct identification of order within extended sequences, and having a lower limit of resolution ranging from about 100 msec to 600 msec per item depending upon the nature of components and the response procedure.

The sequences of phonemes composing speech occur at rates too rapid to permit the direct identification of order. It is suggested that phonemic patterns are perceived holistically rather than as a succession of individual speech sounds. This principle is illustrated by the perceptual organization of sequences of steady-state vowels into illusory syllables and words when the duration of components is too brief to permit identification of the order of sounds. This organizational tendency is so strong that the individual concatenated vowels cannot be perceived by phonetically trained listeners even though the individual vowels could each be identified with ease when all of the others were deleted from the recycled sequence.

The rules governing the perceptual organization of speech sounds has been studied using the verbal transformation effect. Verbal transformations represent a special linguistic form of perceptual changes associated with unchanging patterns of stimulation. It appears that these verbal changes tap into mechanisms normally employed for correction of errors when listening to discourse, and that this illusion permits us to examine these mechanisms. There is evidence that age differences in frequency and types of transformations reflect adaptive changes in the perceptual processing of speech, enabling listeners to compensate for changes in functional capacity accompanying normal maturation and aging.

Studies of auditory illusions and confusions have provided clues not only to the general mechanisms underlying auditory perception, but also to specialized verbal mechanisms. Helmholtz's statement in the last century that illusions provide a "particularly instructive" way to study perception seems at least as valid today. It is suggested that future research involving verbal illusions will continue to be an exciting and rewarding method for studying the mechanisms employed for the perception of speech.

REFERENCES

Amster, H. (1964). Semantic satiation and generation: Learning? Adaptation? Psychological Bulletin, 62, 273-286.

Barnett, M.R. (1964). Perceived phonetic changes in verbal transformation effect. Doctoral dissertation, Ohio University.

Bashford, J.A., Jr., Meyers, M.D., Brubaker, B.S., & Warren,R.M. (1988). Illusory continuity of interrupted speech: Speech rate determines durational limits. Journal of the Acoustical Society of America, 84, 1635-1638.

Bashford, J.A., Jr., & Warren, R.M. (1987). Multiple phonemic restorations follow the rules for auditory induction. Perception & Psychophysics, 42(2), 114-121.

Bashford, J.A., Jr., & Warren, R.M. (1979). Perceptual synthesis of deleted phonemes. In Wolf, J.J. and Klatt, D.H.(Eds.), Speech Communication Papers. New York: Acoustical Society of America, pp. 423-426.

Book, W.F. (1925). The psychology of skill with special reference to its acquisition in typewriting. New York: Gregg.

Brain, W.R. (1962). Recent work on the physiological basis of speech. Advancement of Science, 19, 207-212.

Bryan, W.L., & Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. Psychological Review, 4, 27-53.

Bryan, W.L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. Psychological Review, 6, 345-375.

Butler, R.A., & Naunton, R.F. (1962). Some effects of unilateral auditory masking upon the localization of sound in space. Journal of the Acoustical Society of America, 34, 1100-1107.

Butler, R.A., & Naunton, R.F. (1964). Role of stimulus frequency and duration in the phenomenon of localization shifts. Journal of the Acoustical Society of America, 36, 917-922.

Chalikia, M.H. and Warren, R.M. (1990) Mapping the organization of vowel sequences into words. Journal of the Acoustical Society of America, 87, S160 (Abstract).

Cherry, C., & Wiley, R. (1967). Speech communications in very noisy environments. Nature, 214, 1164.

Chistovich, L.A. (1962). Temporal course of speech sound perception. In Proceedings of the 4th International Commission on Acoustics (Article H 18). Copenhagen.

Clegg, J.M. (1971). Verbal transformations on repeated listening to some English consonants. British Journal of Psychology, 62, 303-309.

Dannenbring, G.L. (1976). Perceived auditory continuity with alternately rising and falling frequency transitions. Canadian Journal of Psychology, 30, 99-114.

David, E.E., Guttman, N., & van Bergeijk, W.A. (1958). On the mechanism of binaural fusion. Journal of the Acoustical Society of America, 30, 801-802.

Debigaré, J. (1984). Le Phénomène de la transformation verbale et la théorie de l'ensamble-cellules de D.O. Hebb: Un Modèle de fonctionnement. Revue Canadienne de Psychologie, 38, 17-44.

Diehl, R.L., Klunder, K.R., & Parker, E.M. (1985). Are selective adaptation and contrast effects really distinct? Journal of Experimental Psychology: Human Perception & Performance, 11, 209-220.

Dorman, M.F., Cutting, J.E., & Raphael, L.J. (1975). Perception of temporal order in vowel sequences with and without formant transitions. Journal of Experimental Psychology: Human Perception and Performance, 104, 121-129.

Egan, J.P. (1948). The effect of noise in one ear upon the loudness of speech in the other. Journal of the Acoustical Society of America, 20, 58-62.

Eimas, P.D., & Corbitt, J.D. (1973). Selective adaptation of linguistic feature detectors. Cognitive Psychology, 4, 99-109.

Esposito, N.J. (1985). Verbal transformation effect and auditory reversals. Perceptual and Motor Skills, 61, 1019-1022.

Evans, C.R., & Wilson, J. (1968). Subjective changes in the perception of consonants when presented as "stabilized auditory images." Division of Computer Science Publication No. 41, National Physical Laboratory, England.

Fay, W.H. (1966). Temporal sequence in the perception of speech. The Hague: Mouton.

Fenelon, B., & Blayden, J.A. (1968). Stability of auditory perception of words and pure tones under repetitive stimulation in neutral and suggestibility conditions. Psychonomic Science, 13, 285-286.

Garner, W.R. (1951). The accuracy of counting repeated short tones. Journal of Experimental Psychology, 41, 310-316.

Guilford, J.P., & Nelson, H.M. (1936). Changes in the pitch of tones when melodies are repeated. Journal of Experimental Psychology, 19, 193-202.

Goldstein, L.M., & Lackner, J.R. (1974). Alterations in the phonetic coding of speech sounds during repetition. Cognition, 2, 279-297.

Heise, G.A., & Miller, G.A. (1951). An experimental study of auditory patterns. American Journal of Psychology, 64, 68-77.

Helmholtz, H.L.F. (1954). On the sensations of tone as a physiological basis for the theory of music. New York: Dover. Reprint of 2nd English Edition of 1885 (A.J. Ellis, Trans), based upon the 3rd German Edition (1870) and rendered conformal with the 4th German Edition (1877).

Hirsh, I.J. (1959). Auditory perception of temporal order. Journal of the Acoustical Society of America, 31, 759-767.

Hirsh, I.J. & Sherrick, C.E. (1961). Perceived order in different sense modalities. Journal of Experimental Psychology, 62, 423-432.

Holloway, C.M. (1970). Passing the strongly voiced components of noisy speech. Nature, 226, 178-179.

Houtgast, T. (1972). Psychophysical evidence for lateral inhibition in hearing. Journal of the Acoustical Society of America, 51, 1885-1894.

Huey, E.B. (1968). The psychology and pedagogy of reading. Cambridge, Mass.: MIT press.

Kock, W.E. (1950). Binaural localization and masking. Journal of the Acoustical Society of America, 22, 801-804.

Lackner, J.R., Tuller B., & Goldstein, L.M. (1977). Some aspects of the psychological representation of speech sounds. Perceptual and Motor Skills, 45, 459-471.

Ladefoged, P. (1959). The perception of speech. In National Physical Laboratory Symposium No. 10., Mechanization of Thought Processes. Her Majesty's Stationery Office, London, 1, 309-471.

Ladefoged P. & Broadbent, D.E. (1960). Perception of sequence in auditory events. Quarterly Journal of Experimental Psychology, 12, 162-170.

Lashley, K.S. (1951). The problem of serial order in behavior. In Jeffress, L.A.(Ed.), Cerebral mechanisms in behavior: The Hixon Symposium. New York: Wiley, pp. 112-136.

Lass, N.J., & Gasperini, R.M. (1973). The verbal transformation effect: A comparative study of the verbal transformations of phonetically trained and non-phonetically trained listeners. British Journal of Psychology, 64, 183-192.

Lass, N.J., & Golden, S.S. (1971). The use of isolated words as auditory stimuli in eliciting the verbal transformation effect. Canadian Journal of Psychology, 25, 349-359.

Lass, N.J., West, L.K., & Taft, D.D. (1973). A non-verbal analogue to the verbal transformation effect. Canadian Journal of Psychology, 27, 272-279.

Layton, B. (1975). Differential effects of two nonspeech sounds on phonemic restoration. Bulletin of the Psychonomic Society, 6, 487-490.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. Language and Speech, 6, 172-187.

Locke, J. (1690). Concerning human understanding. London: Holt, Book 2, Chapter 14, Section 13, (Reprinted Oxford: Clarendon, 1894).

Massaro, D.W. (1979). Reading and listening (tutorial paper). In Kolers, P.A., Wrolstad, M.E., and Bouma H.(Eds.), Processing of visible language. New York: Plenum, pp. 331-354.

Miles, W.R. (1934). Age and the kinephantom. Journal of General Psychology, 10, 204-207.

Miller, G.A. (1962). Decision units in the perception of speech. IRE Transactions on Information Theory, IT-8, 81-83.

Miller, G.A. & Licklider, J.C.R. (1950). The intelligibility of interrupted speech. Journal of the Acoustical Society of America, 22, 167-173.

Moore, M.W., & Bliss, J.C. (1975). The Optacan reading system. Education of the Visually Handicapped, 7, 15-21.

Naeser, M.A., & Lilly, J.C. (1970). Preliminary evidence for a universal detector system--perception of the repeating word. Journal of the Acoustical Society of America, 48, 85 (Abstract).

Necker, L.A. (1832). Observations on some remarkable optical phenomena seen in Switzerland; and on an optical phenomenon which occurs on viewing a figure of a crystal or geometrical solid. London & Edinburgh Philosophical Magazine & Journal of Science, 1(3rd Series), 329-337.

Obusek, C.J. (1971). An experimental investigation of some hypotheses concerning the verbal transformation effect. Unpublished doctoral dissertation, University of Wisconsin, Milwaukee.

Obusek, C.J., & Warren, R.M. (1973). A comparison of speech perceptions in senile and well-preserved aged by means of the verbal transformation effect. Journal of Gerontology, 28, 184-188. (a)

Obusek, C.J., & Warren, R.M. (1973). Relation of the verbal transformation and the phonemic restoration effects. Cognitive Psychology, 5, 97-107. (b)

Ohde, R.N., & Sharf, D.J. (1979). Relationship between adaptation and the percept and transformations of stop consonant voicing: Effects of the number of repetitions and intensity of adaptors. Journal of the Acoustical Society of America, 66, 30-45.

Perl, N.T. (1970). The application of the verbal transformation effect to the study of cerebral dominance. Neuropsychologia, 8, 259-261.

Powers, G.L., & Wilcox, J.C. (1977). Intelligibility of temporally interrupted speech with and without intervening noise. Journal of the Acoustical Society of America, 61, 195-199.

Pritchard, R.M., Heron, W., & Hebb, D.O. (1960). Visual perception approached by the method of stabilized images. Canadian Journal of Psychology, 14, 67-77.

Riener, K.R., Bashford, J.A., Jr., & Warren, R.M. (1990). Increasing the intelligibility of speech through phonemic restorations. Journal of the Acoustical Society of America, 87, S71 (Abstract).

Riener, K.R., & Warren, R.M. (1990). Verbal organization of vowel sequences: Effects of repetion rate and stimulus complexity. Journal of the Acoustical Society of America (Abstract in press).

Riggs, L.A., Ratliff, F., Cornsweet, J.C., & Cornsweet, T.N. (1953). The disappearance of steadily fixated visual test objects. Journal of the Optical Society of America, 43, 495-501.

Roivainen, E. (1989). Verbal transformations in the aged. Perception, 18, 675-680.

Sadler, M.E. (1989). Effects of noise amplitude upon the rate and nature of transformations. Unpublished doctoral dissertation, University of Wisconsin-Milwaukee.

Samuel, A.G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. Journal of Experimental Psychology: Human Perception and Performance, 7, 1124-3113. (a)

Samuel, A.G. (1981). Phonemic restoration: Insights from a new methodology. Journal of Experimental Psychology: General, 110, 474-494. (b)

Samuel, A.G. (1987). Lexical uniqueness effects on phonemic restoration. Journal of Memory and Language, 26, 36-56.

Samuel, A.G., & Ressler, W.H. (1986). Attention within auditory word perception: Insights from the phonemic restoration illusion. Journal of Experimental Psychology: Human Perception and Performance, 12, 70-79.

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

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

Skinner, B.F. (1936). The verbal summator and a method for the study of latent speech. Journal of Psychology, 2, 71-107.

Sperling, G., & Reeves, G. (1980). Measuring the reaction time of a shift of visual attention. In Nickerson, R.S.(Ed.), Attention and Performance VIII. Hillsdale, N.J.: Erlbaum, pp. 347-360.

Taylor, M.M., & Henning, G.B. (1972). Verbal transformations and an effect of instructional bias on perception. Canadian Journal of General Psychology, 86, 231-245.

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

Thomas, I.B., Cetti, R.P., & Chase, P.W. (1971). Effect of silent intervals on the perception of temporal order for vowels. Journal of the Acoustical Society of America, 49, 84 (Abstract).

Thurlow, W.R. (1957). An auditory figure-ground effect. American Journal of Psychology, 70, 653-654.

Thurlow, W.R., & Elfner, L.F. (1959). Continuity effects with alternately sounding tones. Journal of the Acoustical Society of America, 31, 1337-1339.

Titchener, E.B. (1915). A beginner's psychology, New York: Macmillan.

Verschuure, H., Brocaar, M. (1983). Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. Perception & Psychophysics, 33, 232-240.

Vicario, G. (1960). L'effecto tunnel acoustico. Rivista di Psicologia, 54, 41-52.

Warren, R.M. (1961). Illusory changes of distinct speech upon repetition--The verbal transformation effect. British Journal of Psychology, 52, 249-258. (a)

Warren, R.M. (1961). Illusory changes in repeated words: Differences between young adults and the aged. American Journal of Psychology, 74, 506-516. (b)

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. (a)

Warren, R.M. (1968). Verbal transformation effect and auditory perceptual mechanisms. Psychological Bulletin, 70, 261-270. (b)

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

Warren, R.M. (1971). Identification times for phonemic components of graded complexity and for spelling of speech. Perception and Psychophysics, 9, 358-363.

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. (1976). Auditory illusions and perceptual processes. In Lass, N.J.(Ed.), Contemporary issues in experimental phonetics. New York: Academic Press, pp. 389-417.

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

Warren, R.M. (1984). Perceptual Restoration of Obliterated Sounds. Psychological Bulletin, 96 (2), 371-383.

Warren, R.M. (1985). Criterion Shift Rule and Perceptual Homeostasis. Psychological Review, 2 (4), 574-584.

Warren, R.M., & Ackroff, J.M. (1976). Two types of auditory sequence perception. Perception & Psychophysics, 20, 387-394.

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

Warren, R.M., Bashford, J.A., Jr., & Gardner, D.A. (1990). Tweaking the lexicon: Organization of vowel sequences into words. Perception & Psychophysics, 47(5), 423-432.

Warren, R.M., Gardner, D.A., Brubaker, B.S. & Bashford, J.A. Jr., (1990). Melodic and non-melodic sequences of tones: Effects of duration on perception. Music Perception, (in press).

Warren, R.M., & Gregory, R.L. (1958). An auditory analogue of the visual reversible figure. American Journal of Psychology, 71, 612-613.

Warren, R.M. & Meyers, M.D. (1987). Effects of listening to repeated syllables: category boundary shifts versus verbal transformation. Journal of Phonetics, 15, 169-181.

Warren, R.M., & Obusek, C.J. (1971). Speech perception and phonemic restorations. Perception & Psychophysics, 9, 358-362.

Warren, R.M., & Obusek, C.J. (1972). Identification of temporal order within auditory sequences. Perception & Psychophysics, 12, 86-90.

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

Warren, R.M., Obusek, C.J., Farmer R.M., & Warren, R.P. (1969). Auditory sequence: Confusion of patterns other than speech or music. Science, 164, 586-587.

Warren, R.M., & Sherman, G.L. (1974). Phonemic restorations based on subsequent context. Perception & Psychophysics, 16, 150-156.

Warren, R.M., & Warren, R.P. (1966). A comparison of speech perception in childhood, maturity and old age by means of the verbal transformation effect. Journal of Verbal learning and Verbal Behavior, 5, 142-146.

Warren, R.M., & Warren, R.P. (1968). Helmholtz on perception: Its physiology and development. New York: Wiley.

Warren, R.M., & Warren, R.P. (1970). Auditory illusions and confusions. Scientific American, 223, 30-36.

Welford, A.T. (1958). Ageing and human skill. London: Oxford University Press.

Wheatstone, C. (1835). Remarks on Purkinje's experiments. Report of the British Association, 551-553.

Wickelgren, W.A. (1965). Distinctive features and errors in short-term memory for English vowels. Journal of the Acoustical Society of America, 38, 583-588.

Wickelgren, W.A. (1966). Distinctive features and errors in short-term memory for English consonants. Journal of the Acoustical Society of America, 39, 388-398.

Wiley, R.L. (1968). Speech communication using the strongly voice components only. Doctoral dissertation, Imperial College, University of London.

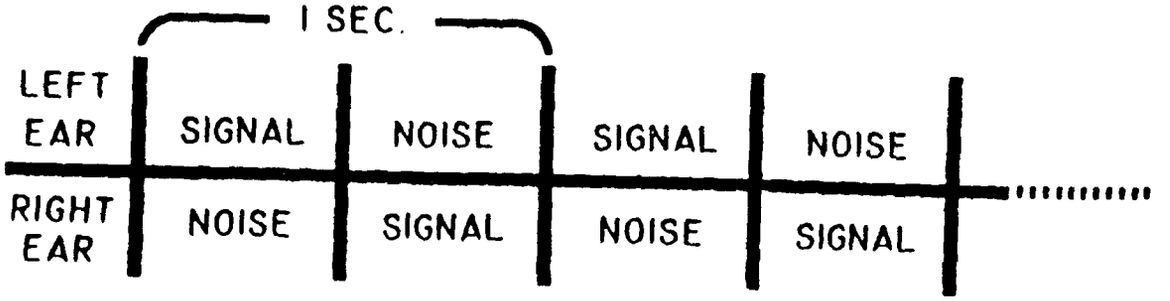


Figure 1. Alternating pattern of stimulation at each ear used to produce delateralization of the signal through contralateral induction. [Source: Warren, R.M., & Bashford, J.A., Jr. (1976). Auditory contralateral induction: An early stage in binaural processing. Perception & Psychophysics, 20, 380-386.]

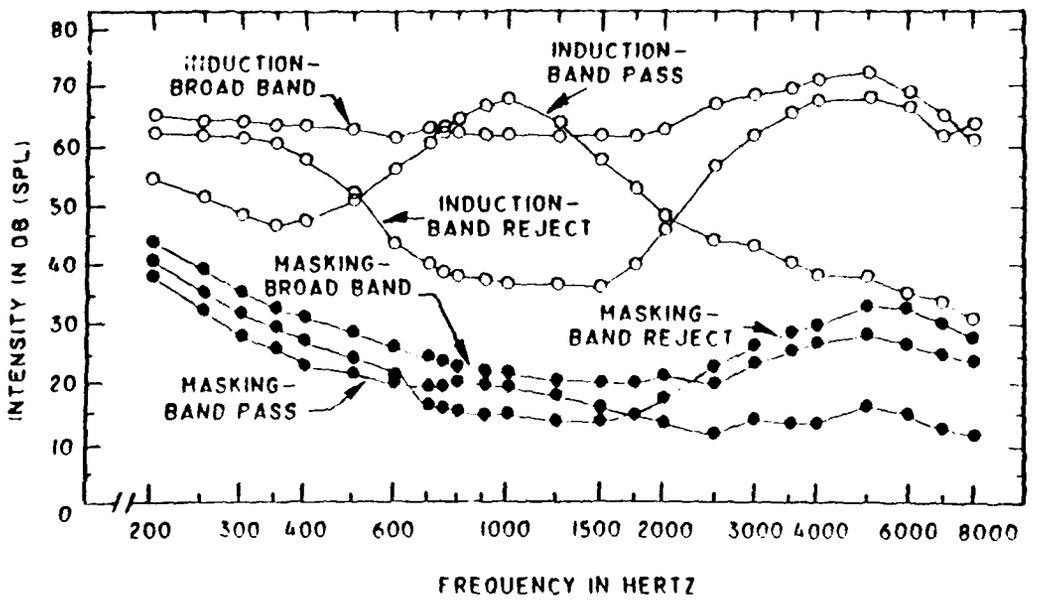


Figure 2. Upper limit for delateralization (contralateral induction) of pure tones (24 different frequencies) delivered to one ear and presented along with three contralateral 80dB SPL noises: broad-band (white) noise; narrow-band (1/3-octave) noise centered at 1000 Hz; and band-reject noise (spectral gap one octave wide) centered at 1000 Hz. The detection thresholds of the tones are shown as well, and the difference between the upper limit for delateralization and the masked detection thresholds represent the range over which contralateral induction occurred. [Source: Warren, R.M., & Bashford, J.A., Jr. (1976). Auditory contralateral induction: An early stage in binaural processing. Perception & Psychophysics, 20, 380-386.]

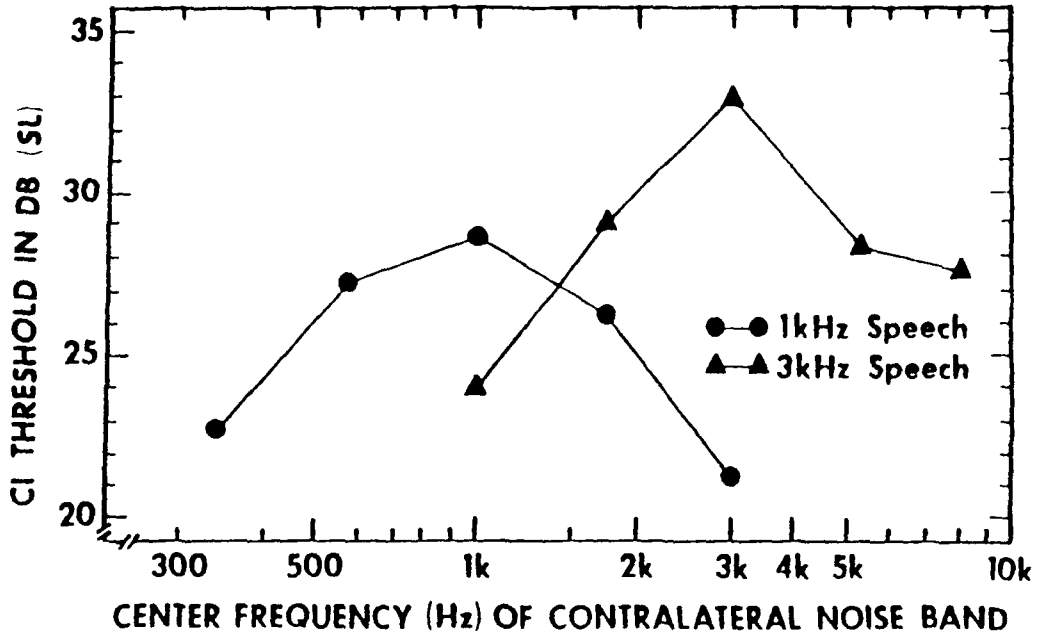


Figure 3. Upper intensity limit for delateralization (contralateral induction) of filtered speech bands (1/3-octave) centered at 1000 Hz and 3000 Hz which were delivered to one ear while contralateral 80 dB 1/3-octave noise bands of various centered frequencies were presented to the opposite ear. Intensity of speech is given as sensation level (SL) or dB above threshold. [Source: Warren, R.M., & Bashford, J.A., Jr. (1976). Auditory contralateral induction: An early stage in binaural processing. Perception & Psychophysics, 20, 380-386.]

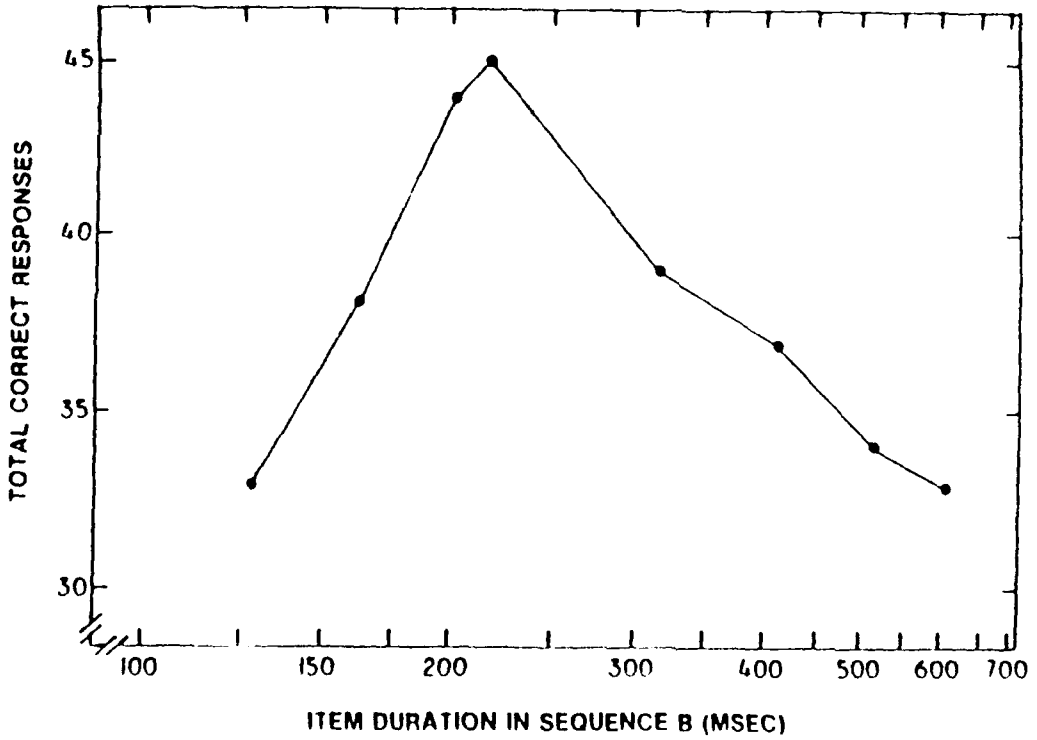


Figure 4. The extent of temporal mismatch permitting pattern recognition. Scores for correct same- or- different judgments are shown for pairs of recycled sequences consisting of the same four sounds arranged in either identical or permuted orders. Sequence A of the pair always had component sounds lasting 200 msec, and the duration of items in Sequence B is given by the abscissa. The maximum score for correct responses is 60. Each data point is based on responses of separate groups of 30 subjects. [Source: Warren, R.M. (1974). Auditory pattern discrimination by untrained listeners. Perception & Psychophysics, 15, 495-500.]

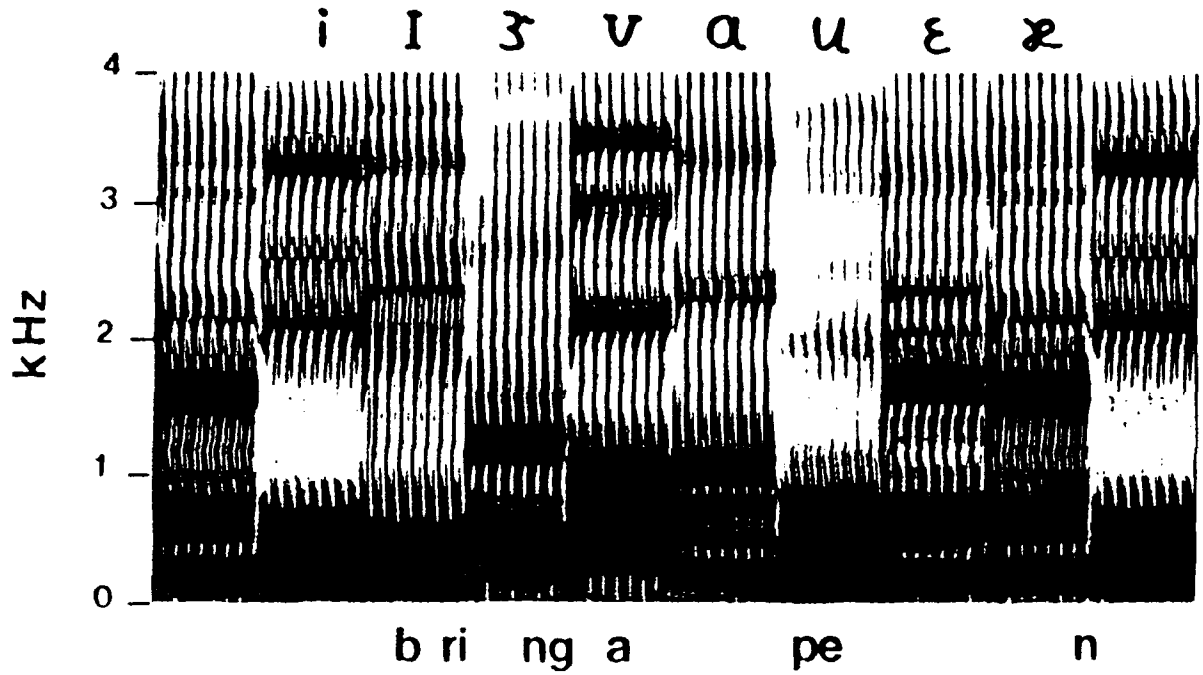


Figure 5. Mapping of an illusory phrase heard with a repeated sequence of eight 80-msec vowels.

APPENDIX 4
Poster presented at May 1990
Acoustical Society of America
Meeting, Abstract published
J. Acoust. Soc Am., 1990, 87, S24

Auditory Memory for Long-Period Random Waveforms

Brad S. Brubaker and Richard M. Warren
Department of Psychology
University of Wisconsin-Milwaukee
Milwaukee, WI 53201

Overview

The present study demonstrates that repeated segments of white noise having repetition periods ranging from 55 to 500 ms could be remembered and identified subsequently (beyond the limit of echoic storage). Recognition of the noise segment is not limited to the same recycled format, since identification can be accomplished when the familiar noise segment is subsequently presented as a portion of a longer noise statement. [A related paper by Bashford & Warren (Poster # I 7 in this session) examined the analogous ability to recognize band limited patterns embedded within broadband signals.]

Introduction

Auditory perception would be of limited use if we were unable to recognize extended temporal patterns. Studies dealing with the recognition of long acoustic patterns have usually employed sequences of sounds such as tones or vowels. However, the use of discrete familiar elements may introduce special mechanisms associated with the perceptual organization of these stimuli, and hence obscure the basic rules governing perception of temporal patterns in general. Only relatively few studies have investigated pattern recognition using the more general case of unfamiliar randomly derived waveforms.

Guttman & Julesz (1963) reported that the repetition of noise segments having durations up to 1 s can be detected effortlessly, with repetition still detectable (with some difficulty) up to 2 s. They described two perceptual types of infrapitch repetition detection. The "whooshing" range extends from 1 through 4 Hz, and is accompanied by a complex pattern of discrete events which grows in detail and

richness with continued listening. The "motorboating" range extends from 4 Hz to the lower limit of pitch (20 Hz), and is characterized by a homogeneous timbre or quality characteristic of the particular waveform. Among the studies dealing with factors influencing the detection of iterance for random waveforms repeated at infrapitch frequencies are those of Schubert & West (1969), Pollack (1969, 1990) and Warren & Bashford, (1981). A second line of studies dealing with the recognition of long temporal patterns has employed non-periodic random waveforms (e.g., Pfafflin, 1968; Pfafflin & Mathews, 1966; Pollack, 1975).

The present investigation has combined these two lines. It was considered that recycling a segment excised from noise would reinforce its memory trace, and hence recognition would be facilitated when the pattern was presented at a later time. Pilot studies indicated that subsequent recognition was indeed facilitated by recycling. In addition, it was found that the repeated waveform need not be presented in the same recycled fashion as that used initially: The noise segment could be recognized subsequently when it was part of a longer pattern. An interesting distinction between these two presentation methods is that when the recycled pattern is embedded in a longer noise segment, the loop is cut, and a beginning and end of the pattern is formed. The ability of listeners to recognize the noise segment when surrounded by additional noise indicates that recognition is not restricted to the original cyclical format but can be extended to linear derivatives. It would seem that the recognition of complex long period patterns can be facilitated greatly when prior presentation is in a recycled format. These observations served as the basis for the present study.

General Method

"Frozen noise" segments were prepared by bandpass filtering gaussian noise (100 to 4000 Hz) and storing the segments in a digital form. Stimuli were derived from the stored segments and presented as "parent/derived" pairs. The derived segment was excised from the parent frozen noise, and was always one-third the parent's duration (see Figure 1). Both the parent and derived waveforms were recycled (repeated without pauses) when presented to listeners. The listeners' task was to either tap when the derived segment occurred within the subsequently presented parent (Experiment 1), or to tell whether a longer segment was or was not the parent (Experiment 2). Stimuli were presented at 80 dBA through diotically

wired headphones. Three psychoacoustically trained listeners were used as subjects. All had previous experience in listening to repeating noise segments and distinguishing between two recycling noises of the same period.

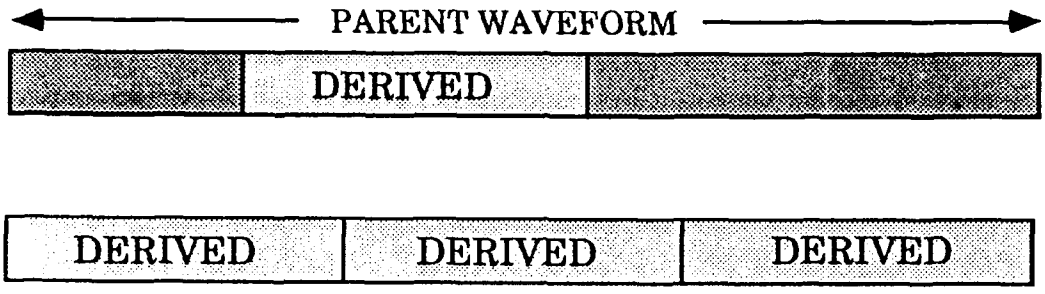


Figure 1. Representation of Parent and Derived Recycled Noise

Experiment 1

Thirty different parent/derived pairs of noise segments were used. The derived recycled noise always had a 500 ms period (2 Hz repetition frequency). The subject listened for as long as desired (usually about 20 s). The stimulus was then turned off for an interstimulus interval of 30 seconds, which was followed by presentation of the parent recycled noise having a period of 1.5 s. The subject's task was to tap a key in synchrony with the occurrance of the derived segment which had been heard earlier. The median value for ten consecutive taps was taken to be the estimate of the perceived location of the derived segment within that parent for each of the 30 parent/derived stimulus pairs. Since each parent is three times as long as the derived segment the chance probability of a correct response was .333.

Table 1. Accuracy of locating a 500 ms derived pattern within a 1.5 s parent

Listener	Percent Correct	z	p
BB	83.3	3.9	<.0001
JB	90.0	4.5	<.0001
KR	76.6	3.4	<.001
Overall	**83.3**	**6.6**	**<.0001**

As shown in Table 1, all subjects could locate the portion of the parent waveform that matched the derived waveform with an accuracy significantly above the values expected by chance.

Experiment 2

In the previous experiment the derived recycled noises were heard as "whooshing", with a fine structure containing repetitive components such as "clanks" and "thumps". Can this recognition task be accomplished when the derived recycled noise is in the motorboating range of infrapitch periodicity and hence heard as a homogeneous repetitive sound?

As in experiment 1, listeners heard a recycled noise followed by a second recycled noise having a repetition period three times that of the first. On half the trials, the longer recycled noise was the parent of the derived segment (that is, it contained the shorter noise segment heard earlier), and on the remaining trials the longer recycled noise was an independent pattern. The repetition period of the derived waveform ranged from 500 (2 Hz) to 55 ms (18.2 Hz) with the values given in Figure 2.

The subject's task was to indicate whether the second stimulus did or did not contain the first. Each subject made thirty judgements at a particular duration before advancing to the next set. For half the stimulus pairs of each set, the longer stimulus contained the waveform of the shorter. All subjects started with the 500 ms duration for the shorter member of the stimulus pair and proceeded through the six sets of stimuli in order of decreasing duration. A five second interstimulus interval was used.

The results are shown in Figure 2. As can be seen, performance was above the chance level of 50% correct at all repetition periods. The pooled identification accuracy for derived recycled noises in the "whooshing" range (81.7%) was significantly higher than pooled identification performance in the motorboating range (70.5 %, $z=2.47$, $p<.01$)

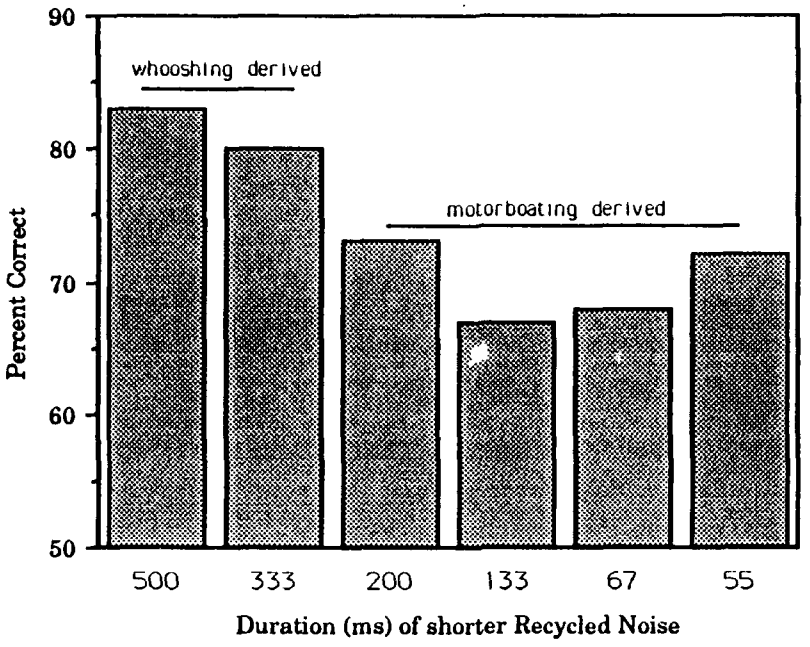


Figure 2. Accuracy of identifying the presence of the shorter pattern within the longer waveform in Experiment 2. [62.2% is significantly above chance ($z=1.65$, $p<.05$).]

Summary and Discussion

Detection of repetition of long-period noise segments requires that information concerning the stimulus be stored in a form permitting recognition when it comes around again. There have been a number of studies dealing with the ability to detect repetition of random waveforms repeated at infratonal frequencies (e.g. Guttman and Julesz, 1963; Warren & Bashford, 1981; Schubert & West, 1969; Pollack, 1990). The ability to recognize non-repeating random waveforms has also been investigated by delivering two patterns separated by a short silent gap (e.g. Pfafflin, 1968; Pfafflin & Mathews, 1966; Hanna, 1984).

In the present experiments, recycling of the noise segments made it possible for the pattern to be recognized subsequently in a format different from that used initially. The results obtained demonstrate that the memory trace of a recycled noise was stable enough to survive a time delay exceeding the limit for echoic storage, and also flexible enough to permit matching to a linear statement of the pattern

embedded in a longer pattern of frozen noise.

Experiment 1 utilized parent/derived recycled noise pairs. The shorter derived recycled noise had a period of 500 ms (repetition frequency of 2 Hz) and was within the "whooshing" range of detectable periodicity characterized by patterns containing discrete components such as "clanks" and "thumps". The longer parent recycled noise had a period of 1.5 s (repetition frequency of .67 Hz) and was below the limit of the "whooshing" range as described by Guttman and Julesz (1963). It was found that after hearing the derived stimulus, listeners could accurately locate its position within the parent waveform following a 30 s interval.

Unlike the whooshing range of recycled noises (1 to 4 Hz) , the motorboating range (4-20 Hz) is characterized by a homogoneous quality or timbre. Could waveform recognition be accomplished within this range as well? Experiment 2 also used parent/derived stimulus pairs, with durational ratios of 3:1. Six durations of the derived stimuilus were used, ranging from 2 to 18 Hz as shown in Figure 2. It was found that performance was above chance for all repetition frequencies. However, identification accuracy was best when the derived pattern was in the whooshing range ($z=2.47$, $p<.01$).

The present study has shown that recycling random patterns provide a convienient method for producing a stable and flexible memory for complex long period waveforms (frozen noise segments repeated at infrapitch frequencies). The memory extends beyond the limit of echoic storage and can be generalized to patterns presented in linear format. [Work supported by AFOSR and NIH.]

References

Guttman, N. & Julesz, B. (1963) Lower limits of auditory periodicity. *Journal of the Acoustical Society of America*, **35**, 610.

Hanna, T.E. (1984) Discrimination of reproducible noise as a function of bandwidth and duration. *Perception and Psychophysics*, **36**, 409-416.

Pfafflin, S. M. (1968) Detection of auditory signal in restricted sets of reproducible noise. *Journal of the Acoustical Society of America*, **43**, 487-490.

Pfafflin, S. M. & Mathews, M. V. (1966) Detection of auditory signals in reproducible noise. *Journal of the Acoustical Society of America*, **39**, 340-345.

Pollack, I. (1969) Depth of sequential auditory information processing. *Journal of the Acoustical Society of America*, **46**, 952-964.

Pollack, I. (1975) Identification of random auditory waveforms. *Journal of the Acoustical Society of America*, **58**, 1262-1271.

Pollack, I. (1990) Detection and discrimination thresholds for auditory periodicity. *Perception & Psychophysics*, **47**, 105-111.

Schubert , E. D. & West, R. A. (1969) Recognition of repeated patterns: A study of short-term auditory storage. *Journal of the Acoustical Society of America*, **46**, 1493-1501.

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

APPENDIX 5
Poster presented at May 1990
Acoustical Society of America
Meeting, Abstract published
J. Acoust. Soc Am., 1990, 87, S71

Increasing the Intelligibility of Speech through Multiple Phonemic Restorations. Keri R. Riener, James A. Bashford, Jr., and Richard M. Warren (Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI 53201).

Overview

Earlier studies have shown that the extent of illusory continuity of interrupted speech increases with both the masking potential of an interrupting noise and with the amount of linguistic context provided by the signal. The present study examines objective effects of filling regularly spaced gaps in speech with noise, and provides evidence that restoration of intelligibility has spectral and contextual requirements similar to those governing illusory continuity. These results are consistent with the theory (Warren et al., 1972; Warren, 1984) that considers phonemic restoration to be one example of a general compensatory process called 'auditory induction,' which can produce illusory continuity of both verbal and nonverbal signals interrupted by potential maskers, and which can restore intelligibility of interrupted speech given minimal sentential context.

Introduction

A wide variety of acoustic signals may appear to be continuous when portions are replaced or obliterated by an appropriate extraneous sound. Such illusory continuity has been demonstrated under conditions of periodic interruption for both verbal and nonverbal signals (Miller & Licklider, 1950), and has also been studied under conditions involving single deletions of complete phonemes or syllables from otherwise intact words or sentences

(Warren, 1970; Warren & Obusek, 1971; Obusek & Warren 1973).

Early studies using nonverbal signals such as pure tones demonstrated that illusory continuity through periodic interruptions required that the interrupting sound be a potential masker of the signal (Houtgast, 1972; Warren et al., 1972). This finding, coupled with the discovery of phonemic restorations occurring with single deletions of speech, led Warren et al. (1972) to propose that illusory continuity of interrupted sounds (including speech) is produced by a process which routinely compensates for the effects of intermittent masking. This process, called 'auditory induction,' is considered to use the neural input afforded by an extraneous noise as substrate for synthesis of contextually likely signals.

Confirmation of the predicted spectral requirements in the case of verbal induction has come in part from studies employing single interruptions of sentences and words. Restoration of single speech fragments has been found to be more robust when the extraneous sound is acoustically similar to the missing speech (e.g., Layton, 1977; Samuel, 1981b). In more recent work employing multiple interruptions, Bashford & Warren (1987a) have obtained evidence of a detailed correspondence between masking and illusory continuity of speech, as well as evidence of a large effect of linguistic context upon the durational limit of verbal induction. When listeners in that study were presented with narrowband discourse periodically interrupted by different narrowbands of noise, the duration of interruption required for the detection of gaps in speech increased (relative to a silent gap condition) by as much as 225 ms when gaps were filled with noise having the same center frequency, but increased by only 50 ms with noise bands two octaves removed in center frequency. Similar variation in the durational limits of continuity were

observed as a function of context: Addition of broadband noise to gaps within broadband speech produced a 250 ms increase in threshold gap durations for normal discourse but only a 100 ms increase both for isolated words and for the same discourse passage read backward at the same word rate.

If phonemic restoration represents a linguistic adaptation of the more general phenomenon of auditory induction, it is of interest to know under what conditions this specialized form can actually aid in the comprehension of spoken language. As will be discussed in greater detail below, several investigators have observed noise-induced enhancements of intelligibility for speech subjected to multiple interruptions (Cherry & Wiley, 1967; Holloway, 1970; Powers & Wilcox, 1977; Verschuure & Brocaar, 1983; Bashford & Warren, 1987b). However, spectral effects have not been examined. Also, as pointed out by Verschuure & Brocaar (1983), previous observations of differential enhancement of intelligibility that were possibly due to linguistic context have been confounded with other differences in experimental conditions. The present study was designed to examine spectral and contextual requirements for restoration of intelligibility under uniform conditions. Experiment 1 examined the intelligibility of narrowband filtered sentences periodically interrupted by silence or by different narrowbands of noise. Experiment 2 examined the intelligibility of broadband, contextually graded speech signals interrupted either by silence or by broadband noise.

General Methods

Stimuli

All speech stimuli were produced by the same male speaker and initially recorded (bandpass filtered from 100 Hz to 10 kHz) as described by Bashford and

Warren (1987b). The stimuli were then rerecorded, with the levels of individual words/sentences adjusted so that the slow-peak amplitudes of individual items in all stimulus sets varied by less than 2 dB. Speech and interrupting pink noise (when present) were passed to the listener by separate electronic switches triggered alternately (10 ms rise/fall) to produce a 50% duty cycle. On/off times for speech were 175 ms in Experiment 1 and 200 ms in Experiment 2.

Listeners in both experiments were tested individually in an audiometric room with stimuli delivered through diotically wired headphones. Speech signals were always presented at a slow-peak level of 70 dBA and interrupting noise was delivered at 80 dBA. Listeners were instructed to repeat back each stimulus as best they could during a 3 s interstimulus interval, and they were encouraged to guess when unsure.

Experiment 1

Speech stimuli were bandpass filtered (1.5 kHz center frequency with slopes of 48 dB/octave) CID (Central Institute of the Deaf) sentences used previously in broadband form by Powers and Wilcox (1977) and Bashford & Warren (1987b). These sentences were designed to serve as samples of "everyday American speech" and are arranged in 10 lists of 10 sentences with each list containing a total of 50 phonetically balanced key words that vary in syllabic composition and positioning within sentences.

Separate groups of 30 Introductory Psychology students were randomly assigned to 6 interruption conditions. For listeners in one experimental group, gaps in the filtered sentences were filled with silence. For listeners in the remaining conditions, the sentences were interrupted by one of five 1/3 octave bands of pink noise (filter slopes of 45 dB/octave) having center frequencies of 375, 750, 1500, 3000, and 6000 Hz, respectively. Within experimental conditions,

the order of stimulus presentation was blocked and pseudorandomized so that each ten-sentence list occurred 3 times in each block.

Results

Repetition accuracy scores were subjected to an analysis of variance, which yielded significant main effects for interrupter [$F(5/174) = 7.72$, $p<.0001$] and blocks [$F(9/1566) = 55.32$, $p < .0001$], and a nonsignificant interaction [$F(45/1566) = 0.609$, $p > .98$]. Subsequent Tukey HSD analysis of the block effect indicated that repetition accuracy increased monotonically up to the ninth block of sentences, indicating that listeners gradually improved in their ability to extract information from the fragmented sentences. However, the nonsignificant interaction also indicates that the listeners' adaptation to interruption did not alter the effects of the different interrupters upon intelligibility.

Figure 1 presents percent repetition accuracy collapsed across blocks and plotted as a function of the center frequency of interrupting noise. Intelligibility was higher for sentences interrupted by noise having the same center frequency than for sentences interrupted by silence or by the remaining noise bands ($p < .05$ or better by Tukey HSD). Intelligibility did not differ across the nonmatching noise band conditions and was statistically equivalent to that obtained in the silent gap condition (about 32%).

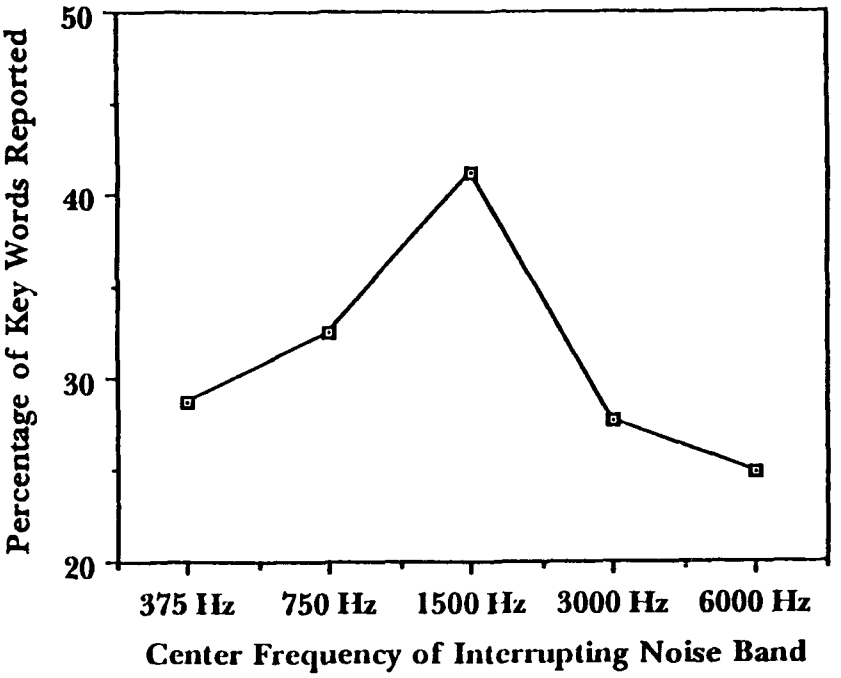


Figure 1: The effect of center frequency of interpolated noise bands upon the intelligibility of narrowband speech (1500 Hz center frequency) CID sentences ("everyday speech") were used.

These results are consistent with earlier findings for illusory continuity of speech obtained by Bashford and Warren (1987a) under the same filtering conditions, and indicate that restoration of intelligibility follows the same acoustic rule governing auditory induction in general. Experiment 2 was designed to examine possible linguistic requirements of verbal induction.

Experiment 2

In an early study dealing with the intelligibility of periodically interrupted speech, Miller & Licklider (1950) observed illusory continuity of isolated monosyllables through brief, noise-filled gaps but found no accompanying increase in repetition accuracy. In contrast, subsequent studies employing

sentences or prose as stimuli have (as in experiment 1 above) demonstrated substantial increases in intelligibility due to interpolated noise (Bashford & Warren, 1987b; Verschuure & Brocaar, 1983; Powers & Wilcox, 1977; Holloway, 1970; Cherry & Wiley, 1967). Although this suggests that restoration of intelligibility has certain contextual requirements, another interpretation has been offered by Verschuure and Brocaar (1983) who presented listeners with interrupted versions of syntactically normal sentences which were either semantically normal or anomalous (e.g., "The church bell drank a sailor."). They found increased intelligibility with noise interpolation for both types of stimuli. Considering their semantically anomalous sentences equivalent to word lists, they suggested that the earlier failure of Miller & Licklider (1950) to obtain an objective effect upon intelligibility was due to the use of trained listeners, who presumably had sufficient skill to obviate the need for interpolated noise. However, semantically anomalous sentences may not be equivalent to word lists. The present experiment was designed to permit a direct comparison of restoration effects for naive listeners using isolated words and syntactically normal sentences which varied in the extent of semantic context. In addition, the results obtained permitted a comparison of restoration of continuity of speech (reported previously) with restoration of intelligibility. As we shall see, continuity and intelligibility effects produced by interpolated noise are experimentally separable.

Method

Each of the eight separate groups of 31 listeners (Introductory Psychology students who did not serve in Experiment 1) served in one of the eight listening conditions (four different types of broadband speech, which were periodically interrupted either by silence or by broadband pink noise. In addition to the CID

("everyday speech") sentences employed in Experiment 1, there were three conditions involving monosyllabic words, including: 1) isolated monosyllables taken from the same phonetically balanced lists used by Miller and Licklider (1950) and produced in isolation at a rate of one word each three seconds, 2) monosyllables appearing as the final word within nonpredictive carrier sentences (example "Ruth hopes she called about the junk."), and 3) the same monosyllables appearing as the final word in highly predictive carrier sentences ("Throw out all this useless junk."). The latter two conditions were derived from the SPIN (Speech Perception in Noise) test of Kalikow et al. (1977).

All speech stimuli were interrupted periodically with a 200 ms on/off cycle, with speech presented at a slow-peak level of 70 dBA. The noise, when present, was delivered at 80 dBA.

Results

Figure 2 presents the scores for percent repetition accuracy, collapsed across blocks and plotted as a function of context and interrupter. Analysis of variance yielded a significant context X interrupter interaction [$F(3,240) = 21.59$, $p < .0001$] that was found through simple effects tests to be due in part to the absence of a noise-induced restoration of intelligibility for words presented in isolation [$F(1,240) = 0.0$, $p > .9$]. For the remaining three conditions presenting key words in sentential context, addition of noise to gaps did significantly improve intelligibility [$F(1,240) \geq 43.4$, $p < .001$]. Additional simple effects tests indicated that restoration of intelligibility was greater for key words in highly predictive carrier sentences than in neutral carriers or CID sentences [$F(2,182) = 4.03$, $p < .02$]. Restoration of intelligibility did not differ significantly between the latter two conditions [$F(1,120) = 1.32$, $p > .1$].

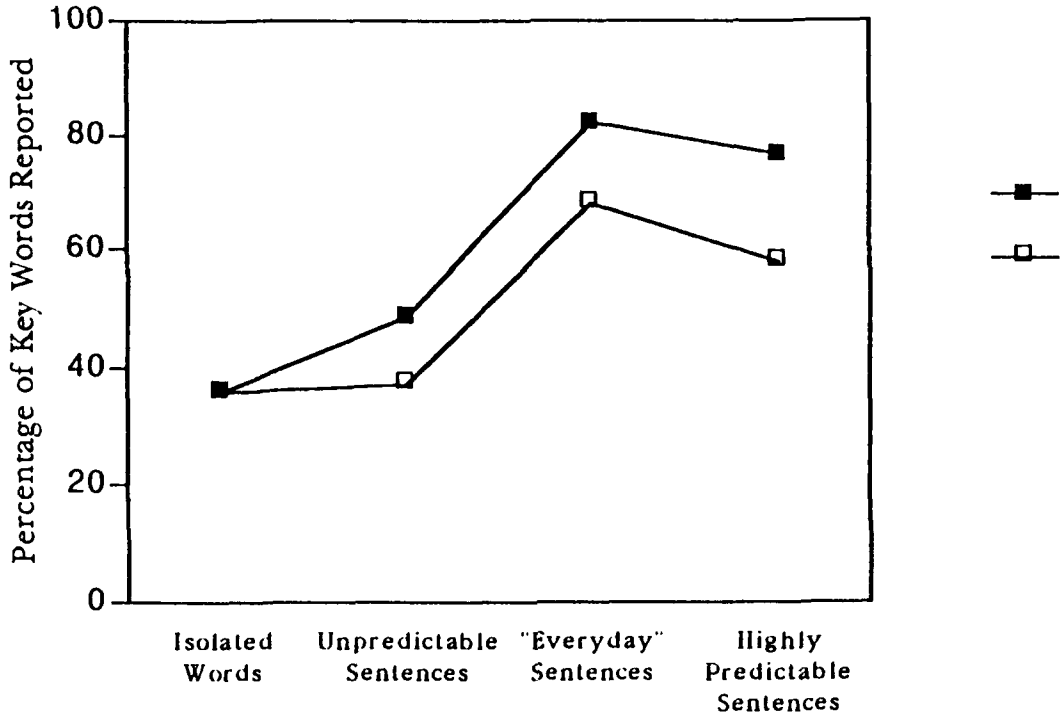


Figure 2: The effect of context on restoration of intelligibility by noise. The stimuli were PB word lists, low predictable SPIN test sentences, CID ("everyday speech") sentences, and high predictable SPIN test sentences

Conclusions

This study provides additional support for the hypothesis that phonemic restoration is a specialized, linguistic form of auditory induction (Warren, et al., 1972, Warren, 1984). Auditory induction requires that an interpolated sound be a potential masker of the signal for restoration of continuity, and Experiment 1 involving narrowband speech and various narrowband noises demonstrated that this requirement also applies to the restoration of intelligibility. Experiment 2 measured the effects of contextual information upon the restoration of intelligibility and provides evidence that top-down

factors play an important role. When word lists were used, no increase in intelligibility was produced by the addition of noise to gaps in the speech signal. But when a sentential framework was present lacking semantic pointers (low predictability key words), noise did enhance intelligibility. Still greater enhancement was found when the sentential framework contained several semantic pointers for the key words.

The inability of interpolated noise to enhance the intelligibility of interrupted word lists is interesting in view of reports in the literature that the addition of noise to gaps in isolated words does produce illusory continuity through brief gaps (Miller & Licklider, 1950; Bashford & Warren, 1987a). It appears that even though the information necessary to identify the missing phonetic components is lacking with word lists, illusory continuity occurs nevertheless as with nonverbal acoustic signals. [Work supported by AFOSR and NIH]

References

Bashford, J. A., Jr., & Warren, R. M., (1987a). Multiple phonemic restorations follow the rules for auditory induction. Perception & Psychophysics, 42(2), 114-121.

Bashford, J. A., Jr., & Warren, R. M., (1987b). Effects of spectral alternation on the intelligibility of words and sentences. Perception & Psychophysics, 42(5), 431-438.

Cherry, E., & Wiley, R. (1967). Speech communications in very noisy environments. Nature, 214, 1164.

Dannenbring, G. L. (1976). Perceived auditory continuity with alternately rising and falling frequency transitions. Canadian Journal of Psychology, 30, 99-114.

Holloway, C. M., (1970). Passing the strongly voiced components of noisy speech. Nature, 226, 178-179.

Houtgast, T. (1972). Psychophysical evidence for lateral inhibition in hearing. Journal of the Acoustical society of America, 51, 1885-1894.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L., (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. Journal of the Acoustical Society of America, 61(5), 1337-1351.

Layton, B. (1975). Differential effects of two nonspeech sounds on phonemic restoration. Bulletin of the Psychonomic Society, 6, 487-490.

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. Journal of the Acoustical Society of America, 22, 167-173.

Powers, G. L., & Wilcox, J. C. (1977). Intelligibility of temporally interrupted speech with and without intervening noise. Journal of the Acoustical Society of America, 61, 195-199.

Obusek, C., & Warren, R. M. (1973). Relation of the verbal transformation and phonemic restoration effects. Cognitive Psychology, 5, 97-107.

Samuel, A. G., (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. Journal of Experimental Psychology: Human Perception and Performance, 7, 1124-1131.

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

Verschuure, J., & Brocaar, M. P. (1983). Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. Perception & Psychophysics, 33(3), 232-240.

Warren, R. M., Wrightson, J. M., & Puretz, J., (1988). Illusory continuity of tonal and infratonal periodic sounds. Journal of the Acoustical Society of America,

84(4), 1338-1342.

Warren, R. M., & Obusek, C. J., (1971). speech perception and phonemic restorations. Perception & Psychophysics, 9, 358-362.

Warren, R. M., (1984). Perceptual restoration of obliterated sounds. Psychological Bulletin, 96, 371-383.

Warren, R. M., Obusek, J. M., & Ackroff, J. M., (1971) Auditory induction: Perceptual synthesis of absent sounds. Science, 176, 1149-1151.

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

APPENDIX 6
Poster presented at May 1990
Acoustical Society of America
Meeting, Abstract published
J. Acoust. Soc Am., 1990, 87, S24

Pattern recognition within spectrally isolated regions of broadband complex sounds. James A. Bashford, Jr., and Richard M. Warren (Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI 53201)

Overview

Earlier studies dealing with memory for long-period complex sounds have demonstrated that listeners can effortlessly detect the repetition of frozen noise when segments as long as 1 s are recycled without pause (Guttman & Julesz, 1963; Pollack, 1969; Warren & Bashford, 1981). In the present study, recycled noise segments (RNs) were used to examine the ability to recognize individual patterns occupying limited spectral regions of broadband infratonal sounds. [A companion paper presented by Brubaker & Warren (poster I8 in this session) examines the analogous ability to recognize particular broadband noise segments embedded within longer broadband RNs.] Experiment 1 of the present study examined listeners' ability to recognize individual RNs (repetition rates from 16 Hz to 1 Hz) which were presented first under 2-octave bandpass filtering and then presented broadband (50 Hz to 10 kHz). Experiment 2 reversed the order of presentation, and required listeners to identify bandpass derivatives of previously heard broadband RNs without knowing in advance which derivative band (low, mid or high) would be presented for recognition. Our trained listeners were able to perform both of these tasks with high accuracy (approximately 90 %) under some conditions. Recognition of patterns heard first bandpass and then within broadband "parent" RNs was accurate for low (150-600 Hz) and midrange (600-2400 Hz) frequency bands, but poor for a high frequency band spanning the range from 2.4 kHz to 9.6

kHz. When the order of presentation was reversed (broadband RNs first), the main difference observed was that performance involving the lowest spectral region was generally poorer. It is suggested that attentional factors associated with the perceptual dominance of the spectral midrange of broadband signals was responsible for the poor performance obtained when the broadband signal was presented first. Frequencies above 2.4 kHz provided a poor substrate for complex pattern recognition regardless of presentation order.

Introduction

It has been reported previously (Guttman & Julesz, 1963; Pollack, 1969; Warren & Bashford, 1981) that listeners can detect the repetition of noise segments at rates extending several octaves below the 20 Hz limit for perception of pitch. Infratonal repetition at frequencies from about 19 Hz down to 4 Hz (noise segment durations of from 53 ms to 250 ms) is heard as a periodic pattern (called "motorboating" by Guttman and Julesz) which has a homogeneous quality that differs for individual noise segments but little or no temporal fine structure. Repetition of longer noise segments, in the so-called "whooshing" range extending from about 4 Hz down to 1 Hz, results in perceptual organizations that are temporally more complex: Listeners may report that a single whooshing pattern contains features such as mechanical "bumps" and "clanks" or may report patterns resembling sounds from the natural environment such as the chirping of a cricket. Typically, these patterns appear to occupy most if not all of the waveform period.

The subjective reports described above suggest that repetition detection for these long-period waveforms involves the recognition of complex

integrated patterns rather than the detection of singularities. Objective evidence indicating complex pattern recognition has been reported by Warren and his coworkers (Warren & Bashford, 1981; Brubaker & Warren, 1987; Bashford & Warren 1988) who used RNs consisting of three or more concatenated, equal duration segments and found that listeners could readily detect the permutation of order within the periodic sounds in both the motorboating and whooshing ranges of repetition.

The present study deals with the recognition of patterns within limited spectral regions, using RNs that were matched in repetition frequency (pattern duration) and derived either from identical or independent noise segments. Only informal observations have been reported concerning the perception of infratonal repetition in different regions of the spectrum. Guttman and Julesz (1963) stated that neither highpass nor lowpass filtering prevented the detection of infratonal periodicity. In addition, Warren and Bashford (1981) swept a 1/3-octave filter through motorboating and whooshing RNs and reported that repetition could be heard at all center frequencies. However, listeners in our laboratory have observed the clarity of repetition to be much greater at low and midrange frequencies than at high center frequencies (above about 3 kHz).

In order to examine further the salience of pattern information provided by different regions of the RN spectrum, we conducted pilot experiments in which listeners were presented with narrowband filtered RNs followed either by the original broadband versions of the same RNs or by different broadband RNs having the same repetition frequency. Listeners were required to judge whether the patterns perceived with the narrowband stimuli were also present in the broadband signals. With 1/3 octave

filtering, listeners reported that patterns heard with narrowbands never closely matched the patterns heard broadband. (One factor interfering with identification of a 1/3 octave pattern when presented as a part of a broadband pattern becomes apparent if we consider that a critical band is approximately 1/3 octave wide. This means that excitation from the higher and lower frequency regions adjacent to the 1/3 octave RN, when present as part of a broadband signal, would mix with and change the pattern from what it was when presented alone.) However, when bandpass RNs had a width of one octave or more, some of our laboratory staff were able to perform the task with little training. For RNs having repetition frequencies greater than about 4 Hz, narrowband and broadband samples of the same iterated segment could be matched on the basis of a similarity in quality or timbre. For RNs repeating at rates lower than about 4 Hz, listeners identified matching stimuli on the basis of discrete features (e.g., "thumps" and "clanks") heard first within the narrowband and then within the broadband stimuli, either as perceptually discrete patterns or as subcomponents of new patterns. Consistent with our earlier observations concerning the apparent salience of pattern information in different frequency bands, our pilot listeners appeared to have greater difficulty recognizing patterns derived from high frequency regions of the RN spectrum.

Experiment 1 of the present study was designed to verify these initial observations under formal conditions. Experiment 2 involved a more difficult task: Listeners were presented first with a broadband RN and then a bandpass RN of the same repetition frequency. Their task was to judge whether the bandpass RN was derived from the previous broadband

stimulus. When listening to the broadband RN, listeners did not know which frequency range would be used for the subsequent test stimulus.

General Methods

Stimuli

Segments of white noise were digitized (50 kHz sampling rate, 16-bit quantization) and recycled without pause using an Eventide delay line (Model SP 2016). The recirculating analog output of the delay line was bandpass filtered from 50 Hz to 10 kHz (48 dB/octave slopes) to produce the "broadband" signals which were recorded at 15 ips on a 16-track recorder. For each of 4 repetition rates, sixteen RN samples were recorded in parallel on the same segment of tape so that they were simultaneously available on playback during the experiment. The 4 RN frequencies consisted of two motorboating rates of approximately 16 Hz and 8 Hz (segment durations of 64 ms and 124 ms, respectively) and two whooshing rates of 2 Hz and 1 Hz (segment durations of 500 ms and 1000 ms, respectively). Three sets of 64 broadband stimuli were prepared in this fashion, with each of the 192 stimuli based on an independent noise capture. One stimulus set was used for preliminary training of listeners and the remaining two sets were used as formal stimuli in Experiments 1 and 2, respectively.

Narrowband stimuli, each two octaves wide, were prepared by bandpassing the broadband RNs (48 dB/octave slopes) with nonoverlapping cutoffs: the low band had cutoffs of 150-600 Hz, the midrange band had cutoffs of 600-2400 Hz, and the high band had cutoffs of 2400-9600 Hz.

Procedure

Listeners made two-interval forced choice judgments either for narrowband RNs followed by broadband RNs (Experiment 1) or for broadband RNs followed by narrowband RNs (Experiment 2). On each

trial, a listener was given as much time as needed to listen for a salient quality or pattern associated with the RN. The listener then terminated the first interval by pressing another key and, following a 2 s interstimulus interval, received an 8 s presentation of the comparison RN. The listeners responded "yes" or "same" if they believed that the RNs contained a common component, and they received trial by trial feedback.

Experiment 1

The listeners in this experiment were three musicians, two professional and one amateur, who were not experienced in psychoacoustic experimentation prior to this study. Before participating in the formal experiment, they received approximately 10 hours of training with equal numbers of judgments under the 12 experimental conditions (three frequency bands at each of 4 repetition rates). All testing was conducted in a sound attenuating chamber with stimuli delivered diotically through Sennheisser headphones (Model HD 230) at an average level of 83 dBA.

In each experimental session, listeners were presented with 16 "same" and 16 "different" trials involving a single RN frequency and a single narrowband filtering condition. Each combination of conditions was presented twice in counterbalanced order, yielding a total of 24 sessions and a total of 64 judgments (32 "same" trials and 32 "different" trials) in each condition for each listener.

Results

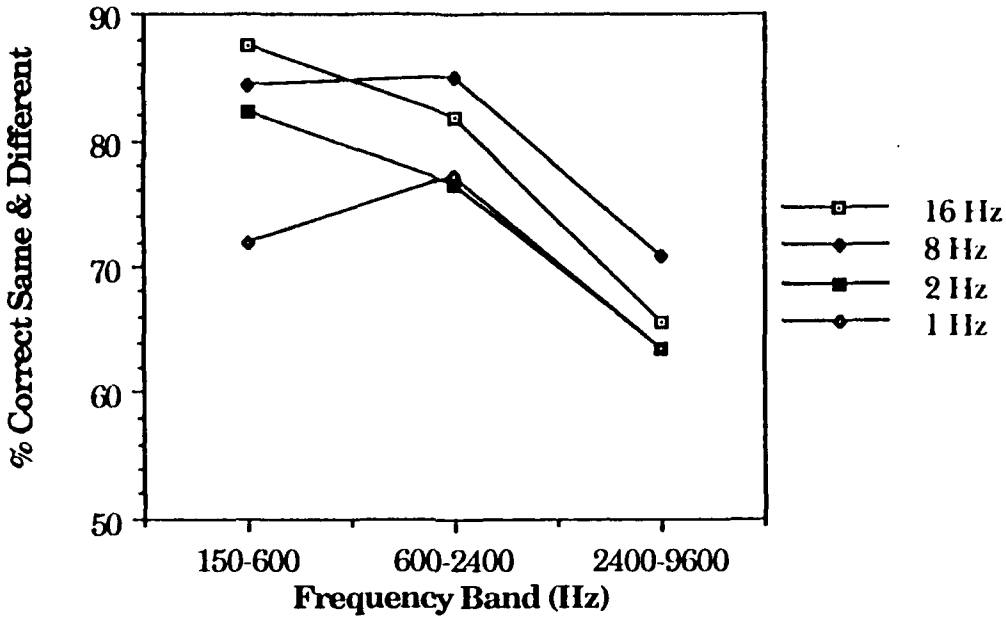


Figure 1: Overall recognition accuracy for broadband "parents" of previously heard narrowband recycled noise segments in Experiment 1. [57.3% is above chance ($z = 2.02$, $p < .05$)]

Figure 1 presents the average percentage of correct recognition responses for each combination of RN frequency and narrowband filtering (192 judgments per data point). Performance was well above chance in all conditions ($z \geq 3.75$, $p < .0002$ or better), ranging from 63.5% to 87.5% accuracy. In keeping with earlier pilot observations, recognition accuracy for the high frequency band was lower than that obtained for the middle band at all repetition rates ($z \geq 2.79$, $p < .01$ or better), and was also lower than that obtained for the low frequency band at all repetition rates but 1 Hz ($z \geq 3.18$, $p < .002$ or better). Within filtering conditions, there were only two significant differences attributable to RN frequency: 1 Hz RNs were recognized less accurately in the lowest frequency band ($z \geq 2.43$, $p < .02$ or better), and 2 Hz RNs were recognized less accurately than 8 Hz RNs ($z = 2.07$, $p < .05$) in the middle frequency band, suggesting that recognition was

generally more difficult for whooshing range RNs.

The results of this experiment demonstrate clearly that the perceptual organization of complex, long-period waveforms is sufficiently distinct and stable within the low and midrange regions of the spectrum to permit listeners to recognize spectrally isolated patterns when they recur as part of a broadband signal. Experiment 2 examined the salience of pattern information within different spectral regions of broadband RNs in the absence of direct priming: Listeners were presented with broadband RNs followed by filtered RNs, and on a given trial did not know which of the three frequency bands would be presented for recognition.

Experiment 2

Participants in this experiment included one of the musicians from the first experiment and two members of the laboratory staff. Listeners received 5 to 10 hrs of training at the experimental task prior to formal testing. Aside from the order of filtering conditions within trials, the procedure in this experiment was similar to that employed in the first experiment. In each experimental session, listeners were randomly presented with 5 "same" and 5 "different" trials for each of the three narrowband filtering conditions at a single RN frequency. This procedure was repeated four times for each listener at each repetition rate in counterbalanced order, yielding a total of 40 judgments (20 "same" trials and 20 "different" trials) for each spectral band per listener.

Results

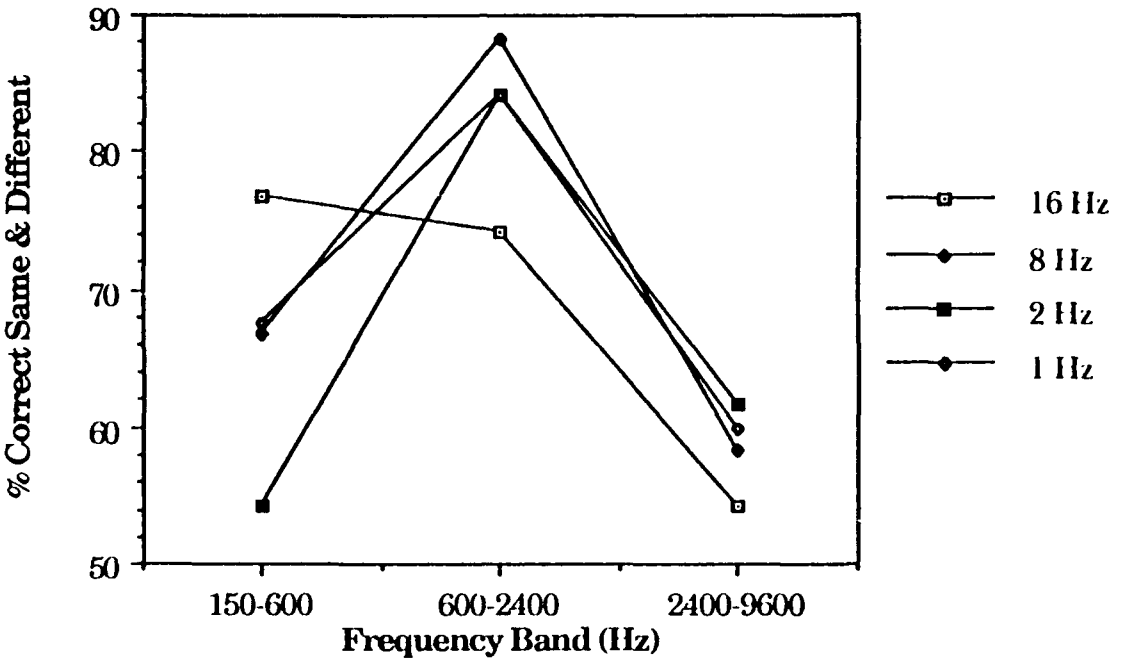


Figure 2. Overall recognition accuracy for bandpass derivatives of previously heard broadband recycling noise segments in Experiment 2. [59.2% is above chance ($z = 2.01$, $p < .05$)]

Figure 2 presents the average percentage of correct recognition responses in the 12 experimental conditions (120 judgments per data point). As found in experiment 1, recognition accuracy for the highest frequency band was poor at all repetition rates, and in the present experiment fell below chance for the 8 Hz and 16 Hz RNs. For the middle band, accuracy was not only well above chance at all RN rates ($z \geq 4.13$, $p < .0001$) but generally equivalent to performance obtained with that band in Experiment 1, ranging from 74 % to 88 % correct. Interestingly, recognition accuracy for the low frequency band in this experiment fell below that obtained with the midrange band at all repetition rates except 16 Hz, suggesting that midrange frequencies tend to dominate in the perception of longer period

infratonal patterns unless listeners are primed by listening to a particular range as in Experiment 1.

Discussion

Previous work has indicated that the detection of infratonal repetition of noise segments is based upon the recognition of complex patterns. The present study indicates that the recognition of such infratonal patterns can proceed somewhat independently within different regions of the RN spectrum, and that memory for these patterns is sufficiently distinct to permit later recognition when they appear in an altered spectral context (see Brubaker and Warren, poster I8, for analogous recognition performance for broadband noise segments embedded within longer broadband RNs).

The results of both experiments 1 and 2 indicate that high frequency regions of the RN spectrum contribute but little to complex pattern recognition. Somewhat similar results have been reported by Hanna (1984) for discrimination of filtered noise segments presented as one-shot patterns.

The results of experiment 2 indicate that midrange frequencies of a broadband signal may suppress processing of infratonal patterns in another frequency region unless listeners have been primed with a spectrally isolated pattern from that region. The perceptual dominance of midrange frequencies may be due in part to the relative sensation level of the three bands. It may also be related to regional differences in resolution of the complex, running spectral profiles afforded by infratonal RNs. Somewhat similar frequency effects have been reported for the discrimination of stationary profiles (Green & Mason, 1985; Green, 1988).

[Work supported by AFOSR and NIH]

References

Bashford, J. A., Jr., & Warren, R. M., (1988). Discrimination of recycled word-length sequences. Journal of the Acoustical Society of America, Suppl. 1, Vol 84, S141.

Brubaker, B. S., & Warren, R. M., (1987). Detection of infratonal repetition of frozen noise: Singularity recognition or pattern recognition? Journal of the Acoustical Society of America, Suppl. 1, Vol. 82, S93.

Green, D. M., and Mason, C. R., (1985). Auditory profile analysis: Frequency, phase, and Weber's law, Journal of the Acoustical Society of America, 77(3), 1155-1161.

Green, D. M., (1988). Profile Analysis: Auditory Intensity Discrimination. New York: Oxford University Press.

Guttman, N. & Julesz, B. (1963). Lower limits of auditory periodicity analysis. Journal of the Acoustical Society of America, 35(4), 610.

Hanna, T. E., (1984). Discrimination of reproducible noise as a function of bandwidth and duration. Perception & Psychophysics, 36(5), 409-416.

Pollack, I. (1969). Depth of sequential auditory processing. Journal of the Acoustical Society of America, 46, 952-964.

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

APPENDIX 7
Poster presented at May 1990
Acoustical Society of America
Meeting, Abstract published
J. Acoust. Soc Am.,1990,87,S160-161

Mapping the Organization of Vowel Sequences into Words.

Magdalene H. Chalikia and Richard M. Warren (Deparatment of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI 53201)

Previous research [R. M. Warren, J. A. Bashford, and D. A. Gardner, (1990)] has shown that when listeners hear recycled sequences of steady-state vowels they do not perceive them as a succession of vowels. Illusory phonemes are introduced and real or nonsense words are heard. Often the sequence of vowels is split perceptually in two simultaneous words differing in both quality and phonemic content. The present study employed sequences of eight 80-ms vowels, and mapped the perceptual phonemes to acoustic phonemes by terminating the repeated sequence at various positions and determining the last sound heard in the perceived word for each position. When two simultaneous words were heard, they both were mapped. Relations between the acoustic phonemes and the perceived phonemes will be described and implications concerning the perceptual organization of speech will be discussed. [Work supported by NIH and AFOSR.]

Mapping the Organization of Vowel Sequences into Words

Studies employing recycled vowel sequences (usually with 3 or 4 vowels) have shown that the naming or identification of order is easily accomplished at 200 ms/item, but is not possible at item durations below 100 ms. (Cole & Scott, 1973; Cullinan, Erdos, Schaefer, & Tekieli, 1977; Dorman, Cutting, & Raphael, 1975; Thomas, Cetti, & Chase, 1971; Thomas, Hill, Carroll, & Garcia, 1970; Warren, 1968; Warren, Obusek, Farmer, & Warren, 1969; Warren & Warren 1970). None of these studies reported observations involving vowel durations below the threshold for identification of order.

In a recent study, Warren, Bashford, and Gardner (1990) have shown that when listeners hear recycled sequences of three steady-state vowels at durations of 30 ms - 100 ms/item, they do not perceive them as a succession of vowels. Instead illusory phonemes are introduced and there is a tendency to hear real or nonsense words, which are different for each arrangements of the vowels.

In addition, when other sequences were used, each consisting of a different random arrangement of a recycled set of ten steady-state 40 ms vowels, listeners heard words and could identify a particular sequence among several on the basis of its verbal correlate. It was also possible to distinguish between pairs of sequences having minimal differences in structure (e.g. reversing the order of a single contiguous pair of vowels), through verbal mediation.

Usually the sequence of vowels was split perceptually into two parts. One organization was a verbal from. The other was either a noise

or a second verbal form. When two simultaneous words were heard, they seemed to differ in both the voice of the speaker and the phonemic content.

Let me play a vowel sequence, so that you will get an idea of what it sounds like. One of the forms heard for this sequence by some listeners, including myself, is 'bluber'. I also hear ' bad baby', as a second form. See what you can hear.

Clearly, that study with brief vowels demonstrated a dissociation between temporal order judgments and discrimination performance. Also, that *acoustic* sequences need not correspond to *perceptual* sequences. Patterns formed by particular arrangements of speech sounds may be recognized as "temporal compounds" , i.e. groupings having special holistic characteristics, without the need for identification of constituents (Warren, 1982, 1988).

The present study attempted to determine the correspondence of the individual speech sounds forming illusory words, to the vowels actually present at that time. It is possible to map the perceptual phonemes to acoustic phonemes, but not through methods that might appear to be the most obvious. Placing an acoustic marker such as a click in one of the vowels would not work, since clicks (and other extraneous sounds) are mislocalized in speech (Ladefoged, 1959; Warren & Obusek, 1971). Increasing the intensity of a vowel appreciably and then listening for a corresponding increase in the level of speech sounds in the illusory word doesn't work, since we found that listeners continue to hear the illusory word, and the increased intensity results in hearing the vowel veridically, but as an extraneous sound which cannot be localized in the word. Deleting a vowel and listening for the disappearance of a portion of the illusory word doesn't work, because the illusory word can change to another form.

However, there is a method that does appear to work quite well (eg. Warren, 1971; Warren & Sherman, 1974). When the repeated vowel sequence is abruptly terminated, the illusory word or words also stop suddenly, and it is easy to designate the last speech sound heard. By presenting the sequence several times, and systematically changing the point of termination on successive presentations, it is possible to map the perceptual phonemes to the acoustic phonemes. We used this method in the present study, interrupting the sequence at the beginning, middle, and end of each vowel.

The stimuli were sequences of eight steady-state vowels, each vowel lasting 80-ms. The sequences differed only in the arrangement of the vowels. The vowels were the ones that you see in the Figure, and correspond to the words *herd, head, hod, hoot, heed, hid, had, hood* .

(Figure 1)

Before the actual experiments started all listeners mapped an English word (academic) and a nonsense word (blandit). We did this to establish that the subjects could follow the instructions and that they could match the perceptual phonemes to the acoustic phonemes.

Subjects next mapped the illusory words heard with a recycled vowel sequence. Initially, we mapped one word per subject. Each listener first heard a sequence, for as long as it was necessary to establish a verbal form, and then went through the mapping procedure for that form. If two forms were heard, he or she was instructed to report only the one that was more salient. After completing the mapping, the process was repeated with a different arrangement of the vowels.

(Figure 2)

Figure 2 shows the spectogram of a vowel sequence, for which a listener heard 'happy boy'. The results of the mapping, at the bottom of the Figure, indicate where the different phonemes of 'happy boy' were heard.

The following results were obtained:

(1) All listeners reported hearing some kind of polysyllabic verbal form (mostly nonsense). However, the syllables reported always followed the phonological rules of English. In addition to the form, there was always some type of residue.

(2) In general, most listeners heard idiosyncratic forms, with some common syllables appearing in the forms reported by subjects listening to the same sequence. When that was the case, the common illusory syllables corresponded roughly to the same stimulus vowels.

We then used other listeners that were required to map both forms if two forms were heard. The results from that study indicated the following:

(1) When two forms were heard, the signal was split into a primary and a secondary form. Usually the primary form was of greater phonetic complexity than the secondary form. More often than not the two forms did not have phonemes in common and their starting points were different (as shown by the mapping). (2) The two voices heard typically had different timbres, suggesting that perhaps different spectral regions could account for these differences.

The results confirm those of the Warren, Bashford and Gardner study (1990) indicating that, when one hears a recycled vowel sequence, an obligatory perceptual transformation of the original sounds occurs, so that

it is not possible to hear the actual phonemes. Listeners are able to list the ordered sequence of speech sounds composing the words that they hear, but these phonemes are derived from the forms (i.e. they are inferred) and bear little correspondence to the phonemes that form the acoustic input.

When the vowel sequence is interrupted in a particular location the reports of the terminal phoneme depend on the verbal organization heard at that time. In the case when two forms are reported, an interruption at the same location (i.e. the same acoustic phoneme) gives rise to two perceived phonemes, one appropriate for each form.

These results suggest that the identification of phonemes and their orders requires the mediation of a prior syllabic organization. Also, it is suggested that a determination of the component sounds and their orders is not necessary for verbal organization to take place (Brubaker & Warren, 1988; Warren, Bashford, & Gardner, 1990).

REFERENCES

BRUBAKER, B. S., & WARREN, R. M. (1988). Learning to identify phonemic orders. *Journal of the Acoustical Society of America,* **84** (Suppl. 1), S154.

COLE, R. A., & SCOTT, B. (1973). Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology*, **27**, 441-449.

CULLINAN, W. L., ERDOS, E., SCHAEFER, R., & TEKIELI, M. E. (1977). Perception of temporal order of vowels and consonant-vowel syllables. *Journal of Speech and Hearing Research*, **20**, 742-751.

DORMAN, M. F., CUTTING, J. E., & RAPHAEL, L. J. (1975). Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance*, **104**, 121-129.

LADEFOGED, P. (1959). The perception of speech. In *National Physical Laboratory Symposium No. 10: Mechanisation of thought processes* (pp. 309-417). London: Her Majesty's Stationery Office.

THOMAS, I. B., CETTI, R. P., & CHASE, P. W. (1971). Effect of silent intervals on the perception of temporal order for vowels. *Journal of the Acoustical Society of America*, **49**, 84.

THOMAS, I. B., HILL, P. B., CARROLL, F. S., & GARCIA, B. (1970). Temporal order in the perception of vowels. *Journal of the Acoustical Society of America,* 48, 1010-1013.

WARREN, R. M. (1968). Relation of verbal transformations to other perceptual phenomena. *Conference Publication No. 42: Institution of Electrical Engineers* (Suppl. 1), 1-8.

WARREN, R. M. (1971). Identification time for phonemic components of graded complexity and for spelling of speech. *Perception and Psychophysics* , **9** (4), 345-349.

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

WARREN, R. M. (1988) Perceptual bases for the evolution of speech. In M. E. Landsberg (Ed.), *The genesis of language* (pp. 101-110), Berlin: Mouton de Gruyter.

WARREN, R. M., & OBUSEK, C. J. (1971) Speech perception and phonemic restorations, *Perception and Psychophysics*, **9 (3B)**, 358-362.

WARREN, R. M., OBUSEK, C. J., FARMER, R. M., & WARREN, R. P. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science*, **164**, 586-587.

WARREN, R. M. , & SHERMAN, G. L. (1974) Phonemic restorations based on subsequent context. *Perception and Psychophysics*, **16**, 150-156.

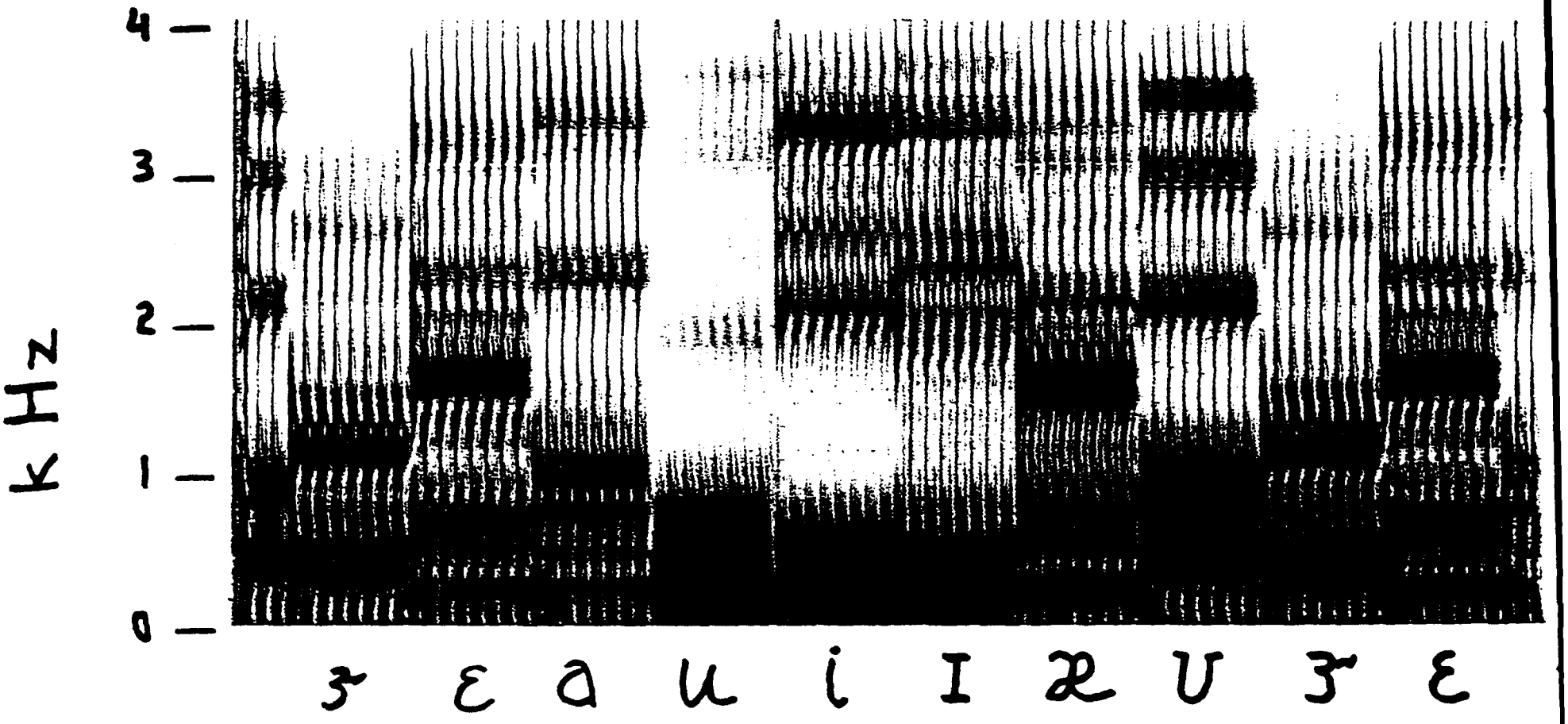
WARREN, R. M., & WARREN, R. P. (1970). Auditory illusions and confusions. *Scientific American*, **223** (December), 30-36.

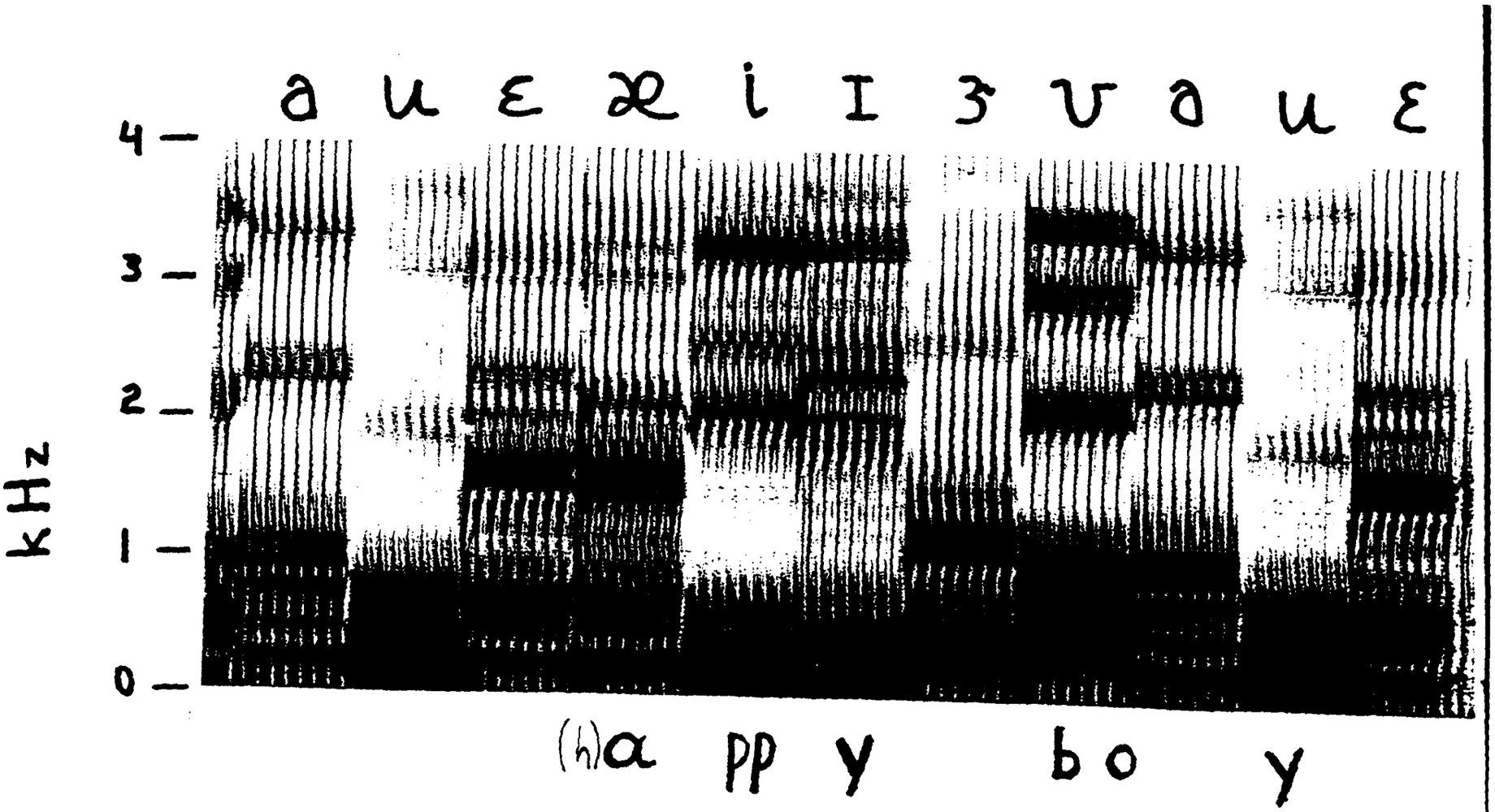
WARREN, R. M., BASHFORD, J. A. JR., & GARDNER, D. A. (1990). Tweaking the lexicon: Organization of vowel sequences into words. *Perception and Psychophysics*, **47 (5)**, 423-432.

Figure Captions

Figure 1. Spectograms of the eight vowels used, those in *herd, head, hod, hoot, heed, hid, had, and hood.*

Figure 2. Spectograms of the eight vowels in a different arrangement, for which a listener heard 'happy boy'.


4
3
2
1
0
kHz
ɝ ɛ ɑ u i ɪ ɚ ʊ ɝ ɛ


ɑ ʊ ɛ æ i I ɜ ʊ ɑ ʊ ɛ
4
3
2
1
0
kHz
(h)ɑ pp y b o y