

AD-A286 563



MENTATION PAGE Dist: A

Form Approved OMB No. 0704-0188



estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson...

REPORT DATE		3. REPORT TYPE AND DATES COVERED ANNUAL 01 Sep 93 TO 31 Aug 94	
4. TITLE AND SUBTITLE DETERMINATION OF MULTIPLE SOUND SOURCES		5. FUNDING NUMBERS F49620-92-J-0489 61102F 2313/AS	
6. AUTHOR(S) Dr William A. Yost		8. PERFORMING ORGANIZATION REPORT NUMBER AFOSR-IR- 94 0713	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Parmlly Hearing Institute Loyola University Chicago 6525 North Sheridan Road Chicago IL 60626		9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR/NL 110 Duncan Ave Suite B115 Bolling AFB DC 20332-0001 Dr Tangney	
11. SUPPLEMENTARY NOTES		10. SPONSORING / MONITORING AGENCY REPORT NUMBER 94-35971 	
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.		A	
13. ABSTRACT (Maximum 200 words) The results studies showed that the SALT procedure produced weights that were much more consistent with subject reports than threshold measures. That is, the change in weights with increasing stimulus duration for the lateralization task indicated that the tones become more separable (analytic listening) as the overall duration increases. This does not occur with traditional measures of thresholds. In the amplitude modulation task, the modulated tones become more separable (analytic listening) as the modulation rates of the targets and distractor differ. Again traditional measures using thresholds do not show the same trend. Subjects also report that the modulated tones are different when the rates differ.			
14. SUBJECT TERMS		15. NUMBER OF PAGES	
		16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT (U)		18. SECURITY CLASSIFICATION OF THIS PAGE (U)	
		19. SECURITY CLASSIFICATION OF ABSTRACT (U)	
		20. LIMITATION OF ABSTRACT (U)	

DTIC ELECTE NOV 23 1994 S G D

94 11 33 3

DTIC QUALITY INSPECTED 8

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to **stay within the lines to meet optical scanning requirements.**

Block 1. Agency Use Only (Leave Blank)

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract	PR - Project
G - Grant	TA - Task
PE - Program Element	WU - Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Names(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. (If known)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of ..., To be published in When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement.

Denote public availability or limitation. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR)

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."

DOE - See authorities

NASA - See Handbook NHB 2200.2.

NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - DOD - Leave blank

DOE - DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports

NASA - NASA - Leave blank

NTIS - NTIS - Leave blank.

Block 13. Abstract. Include a brief (Maximum 200 words) factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code (NTIS only).

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

office partitions and lined with sound attenuating foam on all surfaces. Additional sound attenuating foam was placed in the room at various locations (e.g. directly behind the listener) in order to equalize the reflections measured at the location of the listener for each of the seven loudspeakers. The amplitude of the first reflection ranged from -17 and -21 dB re: the source across the seven loudspeakers, the loudspeaker outputs were matched in dBA output, and the seven loudspeakers were chosen to be within 2 dB of each other in the spectral region of 100 to 7000 Hz (the approximate bandwidth of the stimuli). No other attempts were made to equalize the loudspeakers, since we wanted to create a somewhat real-world listening environment.

The speech materials were 42 NU-6 words, the 26 letters of the alphabet, and the numbers 1 to 9 spoken by seven male talkers. Three judges listened to the words and letters to determine their intelligibility. Additional recording of the words, letters, or numbers were made; or replacement words were chosen until all three judges felt that the utterances from all seven talkers were equally intelligible.

During each test condition each loudspeaker presented a word of a unique male talker (e.g. loudspeaker 1 would always present the utterances of male talker 2, etc.). For each of the listening environments, there were three listening conditions: 1) one-at-a-time, in which a single word, letter, or number was presented from a loudspeaker chosen randomly from trial to trial; 2) two at a time, in which two words, letters, or numbers were presented simultaneously one from each of two loudspeakers that were chosen randomly, 3) three at a time in which three words, letters, or numbers were presented simultaneously one from each of three loudspeakers that were chosen randomly. In each of these conditions there were three lists: 1) 42 NU-6 words (6 words for each of the seven talkers) presented the first time (W1y), 2) the same NU-6 words were presented again but in a different random order (this repetition allowed an estimate of learning) (W2y), and 3) the 26 letters and 9 numbers were presented (let). The first time through each list the listener was asked to enter into the computer keyboard all of the words they heard. They could listen as many times to the words, letters, or numbers, as they wanted and the number of times they listen was recorded. After they finished indicating the words, letters, or numbers they heard; the utterances were repeated in exactly the same order and they were to indicate from which loudspeaker (1-7) they heard a utterance (the actual word, letter, or number they responded with during the word identification part of the experiment was shown). Five listeners participated in each base condition, in which the base condition was listening in one of the three environments and for the 1, 2, or 3 at time conditions (45 total listeners).

All utterances were lowpassed filtered at 7,000 Hz (they were played out at 16,000 Hz rate), were normalized to the same rms level, and when more than one word at a time was presented the utterances were temporally align so that the temporal middle of all utterances were the same. Since the difference in duration of the utterances was maximally 128 ms, the maximum onset (offset) separation was 64 ms. All utterances were presented at 70 dBA and mixed with a continuous broadband noise presented at 65 dBA.

The data for the letters and numbers were scored once for total correct as responded by the listeners. For the words there were three levels of analysis: 1) Level 1 (Wx1)-scored directly as the listener recorded their answers; 2) Level 2 (Wx2)-corrections were made for spelling and homonyms (e.g. dear for deer), 3) Level 3 (Wx3)-additional corrections were made for words or near words that might have been the combination of the words presented (e.g. mop and fall yielding mall), for close homonyms (e.g. drain for rain), and suffixes added by the listener that were not in the spoken word (e.g. homes for home).

RESULTS-DISCUSSION

Performance decreases from the 1-at-a-time to the 2-at-a-time to the 3-at-a-time listening conditions. Performance is best in the normal listening environment, worse in the one-mic and one-phone environment, and intermediate for the KEMAR environment. These differences are greater for the word identification tasks (Wxy) than for the letter and number identification tasks (let). Localization performance is very good in the normal environment, fairly accurate with KEMAR, and at or barely above chance (chance is 1/7 or 14.3%) in the one-mic and one-phone environment. Some listeners did perform above chance in that they recognized some of the voices and assigned them a consistent speaker number and some of the times this assignment was correct. Such a strategy also meant that some listeners also scored below chance.

In the normal listening condition localization performance is best for the loudspeakers directly in front of the listener (4,5, and 6), and performance is lower for the side loudspeakers (1,2,6, and 7). Because the computer terminal and keyboard were directly in front of the listener, they were most likely facing toward loudspeaker 4 during the experiment (KEMAR also faced loudspeaker 4). The brief duration of each sound would have made it unlikely that significant head movements would have occurred. Thus, the data are consistent with poor localization acuity for sounds off to one side.

For the normal and KEMAR listening conditions performance improves the further apart the sounds from the two or three loudspeakers become. There is a small change in performance as a function of loudspeaker separation in the one mic-one phone conditions.

In the KEMAR conditions listener's reported frustrations in not being able to turn toward the perceived location of an utterance. Many listeners believed this inability hindered their performance.

CONCLUSIONS

Binaural hearing does seem to play a role in divided attention tasks that characterize the cocktail party problem. This appears to be especially true when the listener does not have a great deal of familiarity with the messages and when three words were presented simultaneously. Coupling a listener's head movements to the position of KEMAR relative to the sound sources might improve performance when listening in the KEMAR environment. Thus in answer to Cherry's question, "On what logical basis could one design a machine ('filter') for carrying out such an operation (solving the cocktail party problem)?"; spatial hearing does provide one such logical basis.

Analytic and Synthetic Listening

R.H. Dye, William A. Yost, and Stanley Sheft

I. INTRODUCTION

A major role for hearing is to determine the sources of sounds. Since the sounds from many sources are combined into one complex sound field as the input to the auditory system, the auditory system must determine what aspects of this sound field are unique to each sound source. Analytic listening indicates that a listener is able to segregate the information in a sound field according to sources, while synthetic listening represents a failure to segregate. A new procedure,

the Synthetic /Analytic Listening Task (SALT), is used to measure analytic and synthetic listening in two tasks: a lateralization task in which interaural time differences are the basis for sound source segregation, and a modulation discrimination task in which amplitude modulation is the basis for sound source segregation. The SALT procedure is more useful in describing performance related to sound source determination than are the more traditional procedures based on measures of threshold.

II. LATERALIZATION

A. THE SYNTHETIC/ANALYTIC LISTENING TASK.

For the lateralization work each trial consisted of two intervals, with the first providing a diotic presentation of a 753-Hz cue tone that served to indicate the intracranial midline and the target frequency. The second interval presented the test signal, which was a two-tone complex comprised of the 753-Hz target and a 553-Hz distractor. Data were gathered in blocks of 100 trials, with the target and distractor each presented at ten different interaural delays that ranged from left-leading to right-leading and were symmetrically placed about zero. Each possible pairing of target and distractor delay was presented once, in a random order, during each block of trials. Subjects were instructed to indicate whether the *target component* appeared to the left or right of the intracranial midline as marked by the cue tone presented during the first interval. Feedback concerning the position of the target was provided to listeners on a trial-by-trial basis.

The durations of the signals were 25, 50, 100, 200 or 400 ms, with the target and distractor gated simultaneously with 10-ms rise-decay times. Nine listeners were run under conditions in which the interaural delays of the target and distractor ranged from $-90 \mu\text{s}$ to $+90 \mu\text{s}$ in $20\text{-}\mu\text{s}$ steps, with negative delays indicating left-leading signals and positive delays indicating right-leading signals. The level of each component was 53 dB SPL. Within a block of 100 trials, the duration was fixed. Matrices of left-right judgments were generated for each block of trials, with target delay plotted on the abscissa and distractor delay plotted on the ordinate.

B. ANALYSIS OF LEFT-RIGHT RESPONSES.

Extracted from each composite matrix of responses is the slope and x-intercept of the best-fitting linear boundaries between left and right responses. Assume that the interaural delay of the target and the interaural delay of the distractor are related by separate functions to the percepts arising from each (perceived laterality, in this case),

$$X_i = f(\text{IDT}_{T_i}), Y_j = f(\text{IDT}_{D_j}) \quad (1)$$

where X_i is the percept associated with the i th interaural difference of time of the target and Y_j is the percept associated with the j th interaural difference of time of the distractor. Assume that the decision variable used by listeners is a weighted combination of the percepts arising from the target and the distractor dimensions, with w_T and w_D representing the weights given to the target and the distractor perceptual dimensions, respectively. Left-leading signals produce negative values of the percept and right-leading signals produce positive values. Listeners respond "Right" if:

$$(w_T X_i + w_D Y_j) > C \text{ and "Left" if } (w_T X_i + w_D Y_j) < C, \quad (2)$$

where C is the decision criterion used for making left and right responses on the basis of the

decision variable.

$$Y_j = (-w_T/w_D)X_j + C/w_D \quad (3)$$

and the slope of the linear boundary between left and right responses is the ratio of the weights given to the two perceptual dimensions ($m = -w_T/w_D$). Often it will be convenient to normalize the weights so that $w_T + w_D = 1.0$, so:

$$w_T = m/(m-1) \text{ and } w_D = 1/(1-m), \quad (4)$$

the y-intercept, multiplied by w_D , provides an estimate of the decision criterion, C . Analytic performance will be associated with w_T 's near 1.0. Synthetic performance will be associated with w_T 's near 0.5, reflecting equal weighting of the target and distractor.

Without making assumptions about the shapes of the perceptual distributions that result from presentations of left-leading and right-leading targets (distributed uniformly and discretely in stimulus space), little can be said about the shape of the boundary between left and right responses. However, multivariate normal distributions provide good models for many naturally occurring categories. As long as the covariances associated with the "left" and "right" perceptual distributions are the same, the optimal boundary between them will be linear. As such, the assumption that the decision bound is linear appears to be a reasonable one.

Best linear boundaries were found with an algorithm that minimizes the sum of the Euclidian distances between the boundary and misclassified responses ("left" responses to the right of boundary, "right" responses to the left of the boundary). Boundaries associated with target weights ranging from 0.0 to 1.2 in steps of 0.01 were assessed at values of C ranging from -90 to +90 μs . Although this particular algorithm does not seek to maximize classification accuracy, it generally yielded boundaries that fell within the range of those that produced the highest percentage of correct classifications.

C. SUBJECTS AND TRAINING.

All nine listeners had extensive prior experience in lateralization experiments. Prior to data collection, all listeners received at least 15-18 hours of training during which they lateralized 753-Hz tones in isolation as well as in the presence of a 553-Hz distractor.

III. AMPLITUDE MODULATION

A. STIMULI

For the modulation work, two stimuli were presented per trial, the cue and the test stimulus. The cue consists of the amplitude-modulated target and is defined as:

$$S_c(t) = (1+m_t \sin(2\pi f_m t))(\cos(2\pi f_c t)), \quad (5)$$

with f_m =target modulation frequency, f_c =target carrier frequency, and m_t =target depth of modulation.

The test stimulus consists of the amplitude-modulated target and an amplitude modulated distractor and is defined as:

$$S_T(t) = (1+m_t \sin(2\pi f_m t))(\cos(2\pi f_c t)) + (1+m_d \sin(2\pi f_m t))(\cos(2\pi f_d t)), \quad (6)$$

with f_{md} =distractor modulation frequency, f_{cd} =distractor carrier frequency, and m_d =distractor depth of modulation.

The target depth of modulation (m_t) in the cue stimulus was always -15 dB in terms of $20 \log m_t$, but in the test stimulus it and the distractor depth of modulation ($20 \log m_d$) could take on one of ten values (-22.5, -21.0, -19.5, -18.0, -16.5, -13.5, -12.0, -10.5, -9.0, and -7.5 dB; five depths lower than -15 dB and five greater than -15 dB). The carrier frequencies (f_{ct} and f_{cd}) were either 1000 or 4000 Hz; and the modulation frequencies of the distractor (f_{md}) were 4, 8, 16, 32, and 64 Hz, while the modulation rate of the target (f_{mt}) was always 16 Hz. All stimuli were 500 ms in duration. The overall level of the cue and test stimuli was 75 dB SPL and there was a continuous 20-dB Spectrum Level broad band noise (filtered only by the headphones; a similar noise has been used in studies of MDI to make it difficult for listeners to process modulation in a single frequency band).

The listeners' task was to decide if the depth of modulation of the target carrier in the test stimulus was greater or lower than the target when it is in the cue stimulus, noting that the distractor carrier in the test stimulus was modulated with one of the ten depths of modulation. Feedback was provided consistent with the modulation depth of the target in the test relative to that in the cue. For instance, the target carrier in the test stimulus may have been modulated at a -22.5 dB depth of modulation and the distractor carrier modulated at a -7.5 dB depth of modulation. In this case the listener should respond "lower" since the target depth of modulation is lower in the test (-22.5 dB) than it was in the cue (-15 dB). To respond in this way, the listener would have to be able to ignore the fact that the distractor carrier in the test was modulated with a greater depth of modulation (-7.5 dB) than the target in the cue. That is, the listener had to perform analytically in attending to the target carrier and ignoring the distractor carrier. Figure 6 is a schematic description of the stimulus condition and procedure.

B) LISTENERS.

Four listeners with normal hearing were tested in a four-person, sound proof room. There were 100 trials per block, such that on each trial one of the 100 combinations of 10 depths of target modulation and 10 depths of distractor modulation was presented each once. Eight blocks of trials (800 trials) were used to estimate the target and distractor weights.

C) ESTIMATING THE WEIGHTS:

In applying the SALT procedure to this stimulus situation, we followed the method described above. First, a linear boundary, separating the responses, is described by equations (2) and (3) as:

Listeners respond: "Higher" if $(w_T X_i + w_D Y_j) > C$ and "Lower" if $(w_T X_i + w_D Y_j) < C$,

where C is the decision criterion used for making higher and lower responses on the basis of the decision variable (higher means that the target was modulated with a greater depth in the test than in the cue stimulus; and lower means the opposite). The slope of the linear boundary between lower and higher responses is the ratio of the weights given to the two perceptual dimensions ($m = -w_T/w_D$).

For each condition and listener the 800 trials in the ten by ten response matrix were used to determine the linear boundary which best partitions the "high (H)" and "low (L)" responses. As explained above, the linear boundary is determined by an algorithm that minimizes the sum of the

Euclidian distances between the linear boundary and misclassified responses (i.e., an "L" response to the right of the boundary, "H" response to the left of the boundary).

IV. RESULTS AND CONCLUSIONS

The results for both studies showed that the SALT procedure produced weights that were much more consistent with subject reports than threshold measures. That is, the change in weights with increasing stimulus duration for the lateralization task indicated that the tones become more separable (analytic listening) as the overall duration increases. This does not occur with traditional measures of thresholds. In the amplitude modulation task, the modulated tones become more separable (analytic listening) as the modulation rates of the target and distractor differ. Again traditional measures using thresholds do not show the same trend. Subjects also report that the modulated tones are different when the rates differ.

In attempting to describe how listeners process the many sound sources that make up most of our everyday listening environments, we often need to determine if they are analytic or synthetic in processing the spectral and temporal acoustic information. The SALT procedure offers a new method for determining, on a listener by listener basis, the degree to which a set of acoustic parameters is being processed analytically or synthetically. By being able to describe listeners' performance in this way, we hope to be able to more completely describe the process of hearing.

List of Publications for AFOSR Grant F49620-92-j-0489-P.I. William A. Yost

1. Dye, R.H., Yost, W.A., Stellmack, M., and Sheft. A Stimulus Classification Procedure for Assessing the Extent to Which Binaural Processing is Spectrally Analytic or Synthetic, *Journal of the Acoustical Society of America*, in press
2. Shofner, W.P. and William A. Yost, Repetition Pitch: Auditory Processing of Rippled Noise in the Chinchilla, in *Hearing: Physiology and Psychophysics*, Springer-Verlag Press, in press.
3. Stellmack, M. A. and Dye, R. H. "The Combination of Interaural Information across Frequencies: The effects of number and spacing of components, onset asynchrony, and harmonicity" *Journal of the Acoustical Society of America* 93, 2933-2947, 1993.
4. Yost, W.A., Fay, R.R., Popper, A.(Co-Editors), *Psychoacoustics*, Springer-Verlag Press, 1993
5. Yost, William A. Overview of Psychoacoustics, in *Psychoacoustics*,(W.A. Yost, R.R. Fay, and A. Popper, eds), Springer-Verlag Press, 1993
6. Yost, William A and Stanley Sheft. Auditory Perception, in *Psychoacoustics*, (W.A. Yost, R.R. Fay, and A. Popper, eds), Springer- Verlag Press, 1993
7. Yost, William A. Perceptual Models of Localization, in *Proceedings of Perception of Reproduced Sound*, Audio Engineering Society, 12, 1993
8. Yost, William A. *Fundamentals of Hearing: An Introduction*, Academic Press, New York, 1994
9. Yost, William A., New Developments in the Study of Spatial Hearing, *Audiology Today* 6(4), 9-12, 1994.
10. Yost, William A. Modulation Detection Interference: Across-Spectral Processing and Sound Source Determination, *Hearing Research*, 79 (1/2), 48-59, 1994.
11. Yost, William A. The Cocktail Party Effect: 40 Years Later, in *Localization* (R. Gilkey and T. Anderson, eds) Erlbaum Press, New Jersey, in press
12. Yost, William A, Dye, R.H., and Sheft, S. Analytic and Synthetic Listening, in *Current*

Topics in Acoustical Research, (J.Menon, Ed.), Research Trends, in press

AFOSR Technology Transfer/Transition Information for 1994

Directorate: Chemistry and Life Sciences
Program Manager:

Grant # or Lab Task #: Not applicable

PI or Lab Task Manager :

PI telephone #:

Institution or AF Lab:

Address:

What is transitioned/transferred? (Briefly describe your basic research idea, result, theory, methodology, process, device, fact, knowledge, software etc., which significantly influences the direction of a customer's applied program in government, industry or academia):

Transitioned to whom (who is the customer?):

Name of applied program:

Organization (company, government agency, institution, etc):

Contact (person's name):

Telephone:

Date of transition (year, month or day):

Goal of applied program:

Purpose of transition (how is the direction of the relevant applied program changed by the transition):

8 Sept 94