(2)

AD-A238 788

ᴧENTATION PAGE

Form Approved
OMB No. 0704-0188

REPORT DATE

3. REPORT TYPE AND DATES COVERED
FINAL 15 Apr 87 to 14 Apr 90

4. TITLE AND SUBTITLE
TIME-FREQUENCY FACTORS IN AUDITORY PERCEPTION

5. FUNDING NUMBERS
AFOSR-87-0196
PE - 61102F
PR - 2313
TA - A6

6. AUTHOR(S)
Gregory Wakefield

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Electrical Engineering
and Computer Science
University of Michigan
1301 Beal Ave. - Rm. 4118 EECS Bldg.
Ann Arbor, MI 48109-2122

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR- 91 0638

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
Department of Life Sciences
Building 410
Bolling AFB DC 20332-6448

NL

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The final report of this three year research program summarizes our experimental and theoretical results concerning what we have termed cross-spectral temporal resolution. The basic stimulus is constructed of two simultaneously modulated sinusoidal carriers, f_1 and f_2. The modulation is either of the form of amplitude modulation $s(t)=a_1(t)\cos(2\pi f_1 t)+a_2(t)\cos(2\pi f_2 t)$ of frequency modulation $s(t)=\cos(2\pi(f_1 t+b_1(t)))+\cos(2\pi(f_2 t+b_2(t)))$ where, in both cases, the modulators are sinusoidal $a_i(t)=A_i(1+m_i\cos(2\pi f_{mi} t+0_i))$ and $b_i(t)=B_i\sin(2\pi f_{mi} t+0_i)$. The experiments measured the measured the sensitivity to variations in several of the stimulus parameters. From these experimental results, three theoretical results have been obtained. The first of these concerns the form of the ideal receiver for cross-spectral temporal resolution of amplitude modulated carriers. This model incorporates an auditory-nerve model and is based entirely on point-process statistics. The second of these concerns the form of the ideal receiver for cross-spectral temporal resolution of frequency modulated carriers. This model fails to account for the experimental results, which are, in general, much worse than would be predicted from the model and from a simple extension of the amplitude-modulation results. On the basis of this failure, a third general theoretical results has addressed the problem of uncertainty in auditory processing.

14. SUBJECT TERMS

15. NUMBER OF PAGES
12

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
UNLIMITED

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

Time-Frequency Factors in Auditory Perception
AFOSR-87-0193

- A Final Report -

Submitted to the

Air Force Office of Scientific Research

by

Dr. Gregory H. Wakefield

Department of Electrical Engineering
and Computer Science
University of Michigan
Ann Arbor, Michigan 48109

February 11, 1991

1 Overview

Coherent variations in the temporal structure of acoustic signals across different spectral regions have been shown to be significant in the auditory processing of complex sounds. Comodulation Masking Release, for example, occurs when a cueing band and masking band of noise share the same temporal envelope. The ability to segregate an acoustic stimulus into several acoustic sources may also depend on coherence in such temporal structure for harmonic complexes. The overall aim of the research on time-frequency factors has been to measure directly auditory sensitivity to variations in such coherent structure and to model this sensitivity using statistical decision theory.

The final report of this three year research program summarizes our experimental and theoretical results concerning what we have termed cross-spectral temporal resolution. The basic stimulus is constructed of two simultaneously modulated sinusoidal carriers, f_1 and f_2. The modulation is either of the form of amplitude modulation

$$s(t) = a_1(t)\cos(2\pi f_1 t) + a_2(t)\cos(2\pi f_2 t)$$

or frequency modulation

$$s(t) = \cos(2\pi(f_1 t + b_1(t))) + \cos(2\pi(f_2 t + b_2(t)))$$

where, in both cases, the modulators are sinusoidal

$$a_i(t) = A_i(1 + m_i\cos(2\pi f_{mi} t + \phi_i))$$

and

$$b_i(t) = \beta_i \sin(2\pi f_{mi} t + \phi_i)\,.$$

The experiments measured the sensitivity to variations in several of the stimulus parameters. From these experimental results, three theoretical results have been obtained. The first of these concerns the form of the ideal receiver for cross-spectral temporal resolution of amplitude modulated carriers. This model incorporates an auditory-nerve model and is based entirely on point-process statistics. The second of these concerns the form of the ideal receiver for cross-spectral temporal resolution of frequency modulated carriers. This model fails to account for the experimental results, which are, in general, much worse than would be predicted from the model and from a simple extension of the amplitude-modulation results. On the basis of this failure, a third general theoretical result has addressed the problem of uncertainty in auditory processing.

The psychophysical methods used to measure cross-spectral temporal resolution were based on Levitt's "two-down one-up" adaptive procedure. During the initial phases of our research, we became concerned with the lack of sound theoretical results on the distribution of the threshold statistics. This motivated our development of a mathematical model for the procedure, which, in turn, led directly to an analytical form for the distribution of threshold. Such an approach later was used to develop a more efficient method, which will be summarized in the last section of the report.

2 Amplitude-Modulation Features (Years One and Two)

2.1 Effects of Modulation Frequency

Phase disparity thresholds were measured for a two-carrier sinusoidally amplitude modulated signal as a function of modulation frequency. The standard was of the form

$$s_s(t) = (1 + \cos(2\pi f_m t))\cos(2\pi f_1 t) + (1 + \cos(2\pi f_m t))\cos(2\pi f_2 t)$$

in which the envelopes of the two carriers were in-phase. On the comparison trials, a phase offset was introduced in the envelope of the upper-frequency carrier

$$s_c(t) = (1 + \cos(2\pi f_m t))\cos(2\pi f_1 t) + (1 + \cos(2\pi f_m t + \Delta\phi))\cos(2\pi f_2 t)$$

and the subject's task was to determine which interval, in a 2AFC task, contained the phase-shifted envelopes. Typical results were shown in a previous report for modulation frequencies of 8, 16, 32, and 64 Hz. The carriers were 500 Hz and 2000 Hz. In general, there was little effect of modulation frequency on performance at the lowest three frequencies when threshold is referenced in units of phase, rather than absolute units of delay. For 64 Hz, several subjects showed slight, but significant, increases in threshold. Measurements above 64 Hz showed that performance degraded substantially in approximately 50% of the subjects. Other subjects showed better performance (modulation of 128 Hz) but also reported that the strength of a residue pitch served as a cue. In light of these results, we believe that performance in this task is limited to frequencies less than 100 Hz, and that above this upper limit, performance is determined by a qualitatively different cue. This upper limit is similar to that reported by Green for temporal effects in profile analysis.

2.2 Effects of Initial Phase Offset

A phase disparity was introduced between the modulators of the standard to determine the effects that phase offset had on performance. Thus,

$$s_s(t) = (1 + \cos(2\pi f_m t))\cos(2\pi f_1 t) + (1 + \cos(2\pi f_m t + \phi_{off}))\cos(2\pi f_2 t)$$

and

$$s_c(t) = (1 + \cos(2\pi f_m t))\cos(2\pi f_1 t) + (1 + \cos(2\pi f_m t + \phi_{off}\Delta\phi))\cos(2\pi f_2 t)$$

Phase disparity thresholds were measured for five different phase offsets, ranging from 0 degrees to 180 degrees in steps of 45 degrees. All subjects showed threshold functions that were non-monotonic with phase offset and had a single minimum at 90 degrees. Subjects differed slightly with respect to sensitivity as a function of phase offset. These effects were observed for all frequency modulators.

2.3 Effects of Amplitude Modulation

Modulation depth was varied from 100% to 12.5% in factors of two to determine the extent to which cross-spectral temporal resolution depends on highly-modulated signals. In general, performance degrades rapidly for modulation depths less than 50% and essentially approaches chance at 180 degrees for 12.5% modulation. This finding is interesting in light of the fact that subjects are still able to detect modulation of the carriers even in conditions in which they are performing at chance with respect to the relative timing of such modulators. These results were confirmed by Yost and Sheft.

2.4 Effects of Carrier Separation

Carriers were separated from 1 to 8 octaves to determine whether frequency separation affected performance. There was a small degradation in performance with frequency separation for all subjects, but these effects were small (5-10% in performance along the psychometric function) when compared to the effects of amplitude modulation or modulation frequency. Under no condition was it possible to drive performance to the theoretical limits (50% correct at 180 degrees phase disparity) as it was for these other two variables. These results were confirmed, partially, by Yost and Sheft, who also showed that more complicated results occur when the carriers are separated by less than an octave. This is to be expected since under these conditions leakage of the envelopes into common channels will lead to strong phase effects in the driven activity from such channels.

2.5 Model of Cross-spectral temporal resolution

An auditory-nerve model (Siebert) that was used originally to evaluate performance in binaural perception (Colburn) was studied with respect to cross-spectral temporal resolution. The model consists of two channels, each of which is driven by the envelope of one of the two carriers. The channel outputs are Poisson point processes whose instantaneous rates are the input envelope signals. When formulated as an ideal receiver, the decision statistic depends on the number of coincident events across the two channels. In his PhD thesis, Colburn considered such a model for binaural perception and evaluated its performance under a number of simplifying assumptions. These assumptions allowed him to derive analytic approximations of the ideal receiver's performance. We studied the exact model and developed computer simulations to evaluate the performance of the receiver. Excellent agreement was obtained between psychometric functions from the experimental data and those of the ideal receiver when the coincidence window (the only free parameter of the model) was approximately 5 ms (averaged across observers). The model was simplified to show that ideal performance is approximately monotonic with the *difference in correlation* between the two envelopes for a given modulation depth. In general, the model can account for the effects of phase offset and modulation frequency on performance, but does not account precisely for the effects of modulation depth on performance. Furthermore, the model predicts that performance should not depend on differences in envelope structure, but only on the correlation between envelopes. This prediction was confirmed by experiments in which multicomponent envelopes with various phase offsets, but identical correlation differences, were used to modulate the sinusoidal carriers.

2.6 The effects of envelope energy

It was hypothesized that signals with equal envelope energy and equal correlation differences would be equally discriminable on the basis of (i) the success of the auditory nerve model with respect to correlation differences and (ii) the fact that the *trend* of the model with modulation depth was consistent with the results and appear as "energy" terms in the form of the receiver. Psychophysical testing supported the hypothesis for a wide range of deterministic and stochastic envelopes.

2.7 Information transmission in point-process systems

The success of the model with respect to most of the major stimulus parameters motivated further research into refining the model. This led us to focus on the statistics of point-process systems as a preliminary stage to further modeling. The results of our first set of results has been published recently in Biological Cybernetics. We are continuing our research in this area, but immediate results with respect to the problem of characterizing cross-spectral temporal resolution are not expected for two-three more years.

2.8 Conclusions

Cross-spectral temporal resolution for two-carrier signals is determined by the difference in envelope correlation and envelope energy over a broad range of modulation frequencies and envelope energy. This suggests that an observer can be completely characterized by a single set of measures. Using this set of measures, it may be possible to study impaired auditory function, either due to sensorineural disease or reverberant/noisy environments through changes in the correlation/energy functions. An auditory-nerve model is a powerful tool for studying auditory performance. Its main drawbacks are the intractability of the mathematics with respect to analytical, closed-form results. The negative impact of these drawbacks on research can be reduced by suitable simulation techniques. An example of these techniques is provided by the completed research and several novel approaches to information transmission have been developed.

3 Frequency-Modulation Features (Years Two and Three)

3.1 Parametric effects

Cross-spectral temporal resolution appears to have similar characteristics regardless of whether the temporal information is presented in the form of amplitude modulation or frequency modulation. In general, the effects noted above for amplitude modulation hold as well for sinusoidal frequency modulation of tonal carriers: (i) when normalized to phase offset, performance is relatively constant as a function of modulation frequency up to approximately 64 Hz, (ii) phase thresholds are convex-cup shaped as a function of initial phase offset with a minimum in the neighborhood of 90 degrees, (iii) performance degrades with decreasing modulation "depth" - index in the case of FM, and (iv) carrier separation has little effect on performance. There is, however, a major difference in *sensitivity* to these manipulations between AM and FM. In general, best-case FM thresholds are 3-10 times larger than best-case AM thresholds. This is qualitatively in agreement with work by Carlyon who observed that subjects could not discriminate FM phase disparities for modulation indices in the "narrowband" region of FM in which sinusoidal FM has approximately the same spectral characteristics as sinusoidal AM. From this result, Carlyon incorrectly however concluded that the auditory system is insensitive to FM temporal excursions; our findings show that the FM must be introduced at a much larger modulation index in order to be processed across the frequency spectrum.

3.2 Modeling FM cross-spectral temporal resolution

The degraded performance noted in the experiments above suggests that similar mechanisms are involved in processing cross-spectral temporal patterns induced by amplitude modulation and by frequency modulation. This hypothesis can be addressed by considering the effective amplitude modulation present in an auditory channel tuned to the carrier that is introduced by frequency modulation. That such amplitude modulation exists is well documented in the literature and reflects the fact that each channel is narrowband - as the instantaneous frequency of the sinusoidal FM sweeps through the passband and into the stopband of the channel the output of the channel will be modulated from high amplitude to low amplitude.

A direct test of the Equivalence hypothesis is to present FM-equivalent amplitude modulation at the two carrier frequencies to determine whether the degraded performance is consistent with a mechanism that computes a correlation between the envelopes of the channel outputs at the two carrier frequencies. When phase disparity thresholds are obtained for such signals, performance is an order of magnitude better than that observed for FM waveforms. It is necessary to de-tune the auditory filter (using a roex model) by a factor of 5 in order to achieve similar performance in a limited set of conditions.

There are a number of possible reasons for the failure of this simple hypothesis to explain the original FM results. One possibility is that the envelopes are radically different from those assumed by the model. This can be dismissed partially by noting that the de-tuning must be substantial before AM performance begins to approximate that in the FM case. Thus, any "reasonable" envelope transformation is likely to fail to fit the original data.

A second possibility is that the auditory system re-integrates the channel data into some quasi-wideband signal prior to cross-spectral temporal resolution. This hypothesis was studied informally by the investigator and a Masters student under the form of mathematical reconstruction theory. We believe that there are some interesting and potentially very powerful results that can be obtained under this approach. When coupled with the statistical bounds obtained by Edwards and Wakefield for auditory nerve firings, a more complete model of information transmission by the VIII-th nerve and what effects it has on subsequent performance may be possible. We would be interested in pursuing this research in a more formal manner, if the AFOSR is interested.

A third possibility is that the frequency modulation introduces enough "time-dispersion" in the processing of envelopes from different channels that performance is degraded. This possibility and the need to model quantitatively the effects of signal uncertainty on performance led to the final year of study in the research project.

4 Uncertainty in auditory signal processing (Year Three)

Rather than beginning with a new experimental domain of uncertainty, we selected that of Neff and of Green as the domain of study. Green has shown that the detection of a sinusoid in wideband noise can be degraded by randomizing the frequency of the sinusoid on a trial-by-trial basis. Somewhat surprising is the result that this degradation is relatively small; thresholds are increased by 3 dB when the sinusoid can range over a 3 kHz band as opposed to a sinusoid of known frequency. Neff has shown that the detection of a sinusoid in multicomponent sinusoidal maskers can be degraded by randomizing the frequency composition of the maskers on a trial-by-trial basis. In light of Green's data for signal uncertainty, Neff's data for masker uncertainty are all the more surprising; thresholds can be elevated by as much as 15-20 dB for certain types of multicomponent maskers and this elevation occurs prior to achieving a "dense" spectral masker as k, the number of components in the masker, ranges from 1 to 300.

To account for his results, Green used the theory of signal detection to show that the signal uncertainty problem is a composite hypothesis problem for which some results have been obtained through statistical simulation. That is, the standard signal detection problem is modeled as a decision problem in which the observation is drawn either from noise alone (H_0, the null hypothesis) or from a signal + noise condition (H_1, the alternative hypothesis). In this case, the likelihood ratio defines the test statistic for the statistically optimum receiver:

$$\Lambda(o) = \frac{p(o \mid H_1)}{p(o \mid H_0)}.$$

When the signal is uncertain, the alternative hypothesis is a composite hypothesis. That is, the observation may have come from the addition of the first possible signal to noise, or the second possible signal to noise, etc. In this case, the likelihood ratio takes the form

$$\Lambda(o) = \frac{\sum_{i=1}^{N} p(o \mid H_{1i}) p_i}{p(o \mid H_0)}$$

which can be factored into the sum of the likelihood ratios for each possible signal, weighted by the probability p_i that the i-th signal occurs on any given trial. That is,

$$\Lambda(o) = \sum_{i=1}^{N} \frac{p(o \mid H_{1i})}{p(o \mid H_0)}.$$

In Neff's experiments, however, uncertainty is introduced in the null and alternative hypothesis. Therefore, the *structure* of the ideal receiver is qualitatively different from that of the single-composite hypothesis and the simple hypothesis receivers. This receiver, known as the double-composite hypothesis receiver, has the form

$$LAMBA(o) = \frac{\sum_{i=1}^{M} p(o \mid H_{1i})p_i}{\sum_{i=1}^{M} p(o \mid H_{0i})p_i}.$$

The primary advantage of SDT in psychophysics is that it provides an optimal mathematical form of the observer by which human performance can be compared. Information can be degraded and this effect can be modeled from the standpoint of the ideal observer to determine whether this better matches human performance, as in the case of the energy detector. However, the primary disadvantage is that performance is difficult, if not intractable, to summarize analytically using SDT. Green relied on simulations and an asymptotic argument developed by Peterson to infer the effects of signal uncertainty on the performance of the ideal receiver in the single composite hypothesis case. According to this argument, a larger effect of uncertainty on performance is expected if the observer is initially certain about the signal class. The smaller effects on human performance can be attributed to an initial degree of uncertainty in the human observer concerning the signal class. Such an asymptotic argument was shown to be false by Birdsall and his colleagues in the 1960's. A corrected form of a suboptimal receiver appears to yield similar results and uphold Green's basic conclusions.

The fact that the structure of the double-composite hypothesis receiver is different from that of the single-composite case raises the possibility that the differences observed experimentally may reflect fundamental differences in the performance of these ideal receivers. We have evaluated this hypothesis by generalizing a technique first introduced by Birdsall and Spooner, known as the ESP receiver, to provide a non-parametric distribution-free set of theorems for bounding performance. These results allow us to infer directionality of effects from small-scale simulations of the ideal receiver's performance to the type of large-scale conditions studied by Neff. From this combination of distribution-free bounding theorems and small-scale simulations we have (i) confirmed the results of Green and (ii) demonstrated the effects observed by Neff. A report of these results is under preparation for publication and will be available shortly.

5 Adaptive procedures

The final material to be summarized is research that occurred during the first two years of the program and recently emerged into an improved psychophysical adaptive procedure. A more detailed account was provided in earlier reports along with copies of manuscripts that were submitted for publication. The major result is that Levitt's two-down, one-up adaptive procedure is conjectured to estimate the 70.7% correct point of the psychometric function, but the tools to substantiate this conjecture are in fact not adequate. We provided a more detailed proof of this conjecture for a truncated procedure and argued that the result holds as the truncation points (the bounding point near 50% and 100% at which the procedure is guaranteed to reverse) are extended to infinity. The approach utilized a Markov model of the adaptive procedure. Also based on this model, we were able to derive exact expressions for the small sample distribution of the threshold estimator. We showed that the estimator is biased for the cases typically measured in the psychoacoustics literature. More recently, we used these Markov tools to develop a modified small-sample procedure which reduces the variance of the estimator by a factor of 3-5. This provides a substantial reduction in the bias of the estimator for the "standard" stopping criterion of 12-14 reversals. The basic procedure substitutes a one-down, one-up adaptive rule for the first k reversals and then switches to the two-down, one-up adaptive rule. The Markov theory is used to determine an optimum k at which the switching occurs.

6 Presentations of research results

The following are abstracts of presented material that was sponsored by the AFOSR. In addition, several manuscripts are in preparation or in revision. Earlier manuscripts on the AM cross-spectral phase disparity and on the Levitt tracking procedure have been withdrawn due to (i) results on a better neural model that have strong implications concerning CMR as well as CS phase disparity and (ii) the discovery of the new "switched" adaptive procedure and a better approach for handling the truncation problem in the original analysis. Copies of these manuscripts as they are submitted will be forwarded to the AFOSR.

Publications

Edwards, B. W. and Wakefield, G. H. (1990). "On the statistics of binned neural point processes: The Bernoulli approximation and AR representation of the PST histogram," *Biological Cybernetics*, to be published, December, 1990.

Abstracts

Wakefield, G. H. (1987). "Detection of Envelope Phase Disparity," *J. Acoust. Soc. Am.*,, 81, S1(34).

Wakefield, G. H. and Edwards, B. W. (1987). "Discrimination of Envelope Phase Disparity," *J. Acoust. Soc. Am.*, 82, S1(41).

Edwards, B. W. and Wakefield, G. H. (1988). "Small sample statistical analysis of Levitt's adaptive psychophysical procedure," *J. Acoust. Soc. Am.*, 83, S1(17).

Edwards, B. W. and Wakefield, G. H. (1989). "The effect of envelope energy on the discrimination of cross-spectral envelope phase disparity," *J. Acoust. Soc. Am.*, 85, S1(121-2).

Wakefield, G. H. and Edwards, B. W. (1989). "Cross-spectral envelope phase discrimination for FM signals," *J. Acoust. Soc. Am.*, 85, S1(122).

Feinman, G. R. and Wakefield, G. H. (1990). "Uncertainty in signals and maskers: A signal detection theoretic analysis," *J. Acoust. Soc. Am.*, 88, S146.

Invited Papers

Wakefield, G. H. and Edwards, B. W. (1989). "Cross-spectral envelope processing in the auditory system," Special Session on the Perception of Complex Sounds, 1989 meeting of the Midwest Psychological Association, May, 1989, Chicago, Illinois.