

Studies of tone sequence perception: effects of uncertainty, familiarity, and selective attention

Charles S. Watson, Gary R. Kidd

Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Initial Experiments
4. Stimulus Uncertainty
5. Target-Tone Discrimination with Uncertainty about the Dimension of Change
6. Temporal Discrimination under Various Levels of Uncertainty and with Prolonged Training
7. Detection of Auditory Pattern Components and Informational Masking (IM)
8. What Do Listeners Learn Through Pattern Discrimination Training? Transposition Experiments
9. On the Relation between the Processing of Speech and of Nonspeech Complex Stimuli (Word-Length Tonal Patterns)
10. Individual Differences in Discrimination and Recognition of Complex Sounds
11. Summary
12. Acknowledgments
13. References

1. ABSTRACT

Listeners are able to resolve subtle spectral-temporal details of speech waveforms when attempting to recognize words spoken in the listeners' native language. This may provide support for the existence of some processing mechanisms that are specific to the sounds of speech. A more parsimonious interpretation is that all highly familiar spectral-temporal patterns (speech and nonspeech) are processed more efficiently than less familiar sounds. A series of studies of word-length tonal patterns, conducted over the past 30 years, has provided some insights into the roles of uncertainty and familiarity in the discrimination and identification of speech and complex nonspeech sounds. Major conclusions from these studies are presented. The findings are generally consistent with the hypothesis that the "special" nature of speech is in large part due to its familiarity and its importance in conveying useful information.

2. INTRODUCTION

The history of psychoacoustics, the study of the abilities of humans and other animals to detect, discriminate, and recognize sounds, began with the invention of electronic sound generation and amplification equipment. The first half century of that history was characterized by what may be loosely termed "cochlea-oriented" research and theory, the study of sensitivity and resolving power (spectral and temporal) for simple sounds, primarily isolated sinusoidal pulses. Following the arrival of technology to reliably create more complex auditory stimuli, especially through digital synthesis, it became apparent that the discrimination and recognition of complex sounds is limited by aspects of processing beyond the sensitivity and resolution of the early stages of the auditory system. Stimulus familiarity, learning, memory, and selective attention, and other aspects of processing

Studies of tone sequence perception

generally assumed to occur at higher levels of the nervous system, play more significant roles in determining listeners' abilities to hear complex sounds than for single tones, noise bursts, or clicks. Some of these more cognitive matters are the subject of this article.

A series of studies of human listeners' abilities to resolve the spectral-temporal details of sequences of pure tones was begun in 1972. As suggested above, one motive for this research program was that in the early 1970's auditory science appeared to be on the verge of exhausting the study of the processing of its traditional stimuli, single tones, noise bursts, and clicks. Many in the field of psychoacoustics began to move on to study the perception of complex sounds that shared more of the properties of sounds encountered in the everyday environment. Rather than tackle the full complexity of speech or familiar non-speech sounds (such as clapping hands or footsteps), it was decided to take a more modest step in that direction. Watson and his co-workers at the Central Institute for the Deaf chose sequences of tones because these stimuli could be designed to share some properties of naturally occurring sounds, while being composed of the previously well-studied sinusoidal pulses (1, 2). If there were important new aspects of human hearing to be discovered, that might be done by attempting to predict how pure tones would be heard when they were part of a sequence, as compared with their occurrence in isolation, or when adjacent to a second tone. The basic idea was to determine how well pure tone detection and discrimination in multi-tone contexts could be predicted from existing auditory theory and data, and to identify any new variables that influence performance for more complex stimuli.

It was recognized that there were a great many questions that might be asked about the hearing of tonal patterns and that a lengthy series of experiments would be required to investigate even a modest proportion of them. What follows is a summary of the more fruitful of those experiments conducted over the last 35 years. The move to sequences of tones from single tones was a very small step towards ecological validity, and it was recognized that these sequences might be such a nonrepresentative stimulus that even the most thorough knowledge of the manner in which they are processed would reveal little about the perception of other classes of complex sounds. However, as discussed below, differences in the principles that govern the perception of tone sequences, as opposed to more familiar complex sounds (such as speech sounds), may be as instructive as the similarities.

3. INITIAL EXPERIMENTS

The first question to be addressed was whether listeners could detect changes in the frequencies of individual components of "word-length" (0.5-sec) ten-tone patterns as well as might be expected from previous studies of discrimination using single tones, or from studies of temporal masking by a single adjacent tone. In later studies, the detection of changes in component duration and intensity were also studied, as was the influence of the surrounding context on component detection thresholds.

The tone patterns selected for initial study consisted of ten 45-msec tones (plus a 2.5-msec rise and decay for each tone). The duration of the individual components was intended to be roughly equal to that of the briefest phonemes in rapid speech, and the resulting 0.5-sec total duration was meant to be similar to the duration of many spoken words. The component frequencies of these ten-tone patterns were generally chosen from the frequency range of speech as heard over a telephone, about 300-3000 Hz. These constraints reflected a second motive for our studying complex nonspeech sounds, which was the assertion in the 1960's and 70's that the sounds of speech may be processed more accurately than other complex sounds, or that "speech is special." Since there had been very little study of the hearing of complex sounds other than speech, this claim seemed a challenge to study some unfamiliar complex sounds.

The basic tonal-pattern stimuli are illustrated in Figure 1, together with one of the psychophysical methods used to study them. In this method, a standard pattern is followed by two test patterns, one of which is the same as the standard while the other has been changed in some way. The most common changes to be detected in these experiments have been increments in the frequency, duration, or intensity of single components, but the detection of a component in a temporal position occupied by a silent gap in the standard has also been studied, as illustrated in the lower panel of Figure 1. In other experiments, a same-different method was used. Feedback about the correctness of the responses was always provided. The tonal patterns were generally presented at 75 dB SPL. Groups of four listeners were trained, 600-800 trials per day, generally with several sessions of training before experimental data were collected.

One of the early experiments (Experiment 3, from Watson *et al.*, 1975) will serve as a representative example. In this experiment, listeners were presented with 50 patterns, each a permutation of a ten-frequency catalog in which the frequencies were spaced at equal log (f) intervals between 256 and 892 Hz. On each trial, one of the 50 patterns was presented and one of its temporal components, designated as the "target", was subject to change in the form of a frequency increment (decrements were studied separately). Within some blocks of 100 trials, the same temporal component was always designated the target, while in other trial blocks, the target was drawn at random from 2, 4, or 8 possible temporal components. Thus, while the listeners had become very familiar with the task and with the catalog of patterns, the trial-to-trial stimulus uncertainty was quite high. The size of the frequency increments was varied in a quasi-adaptive manner, by reviewing the data after each daily session and adjusting the frequency increments to maintain performance between approximately 65 and 85 per cent correct for the average listener. The value of $\Delta f/f$ for 75 percent correct was interpolated for each condition from fitted psychometric functions. The four normal-hearing listeners had 20 hours of training on a similar pattern discrimination task prior to collecting the data in Figure 2; those data represent approximately 15,000 trials per listener.

Studies of tone sequence perception

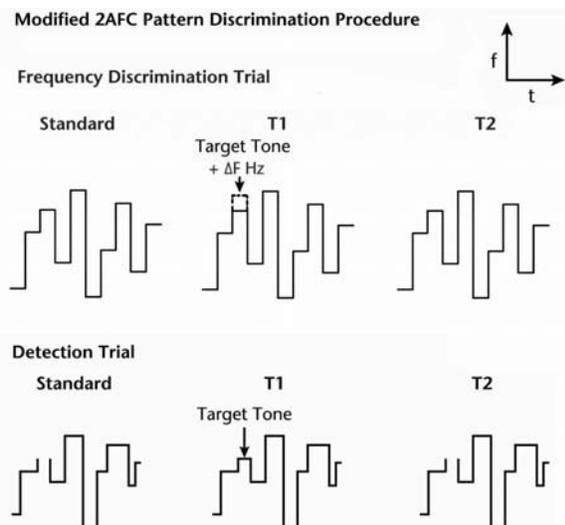


Figure 1. Schematic illustration of the modified two-alternative forced choice procedure used in many of the experiments described here. The upper panel shows a frequency discrimination trial using an isochronous pattern with a change in frequency applied to the third tone in the first test pattern. The lower panel shows a detection trial using an anisochronous pattern in which the third tone is present only in the first test pattern.

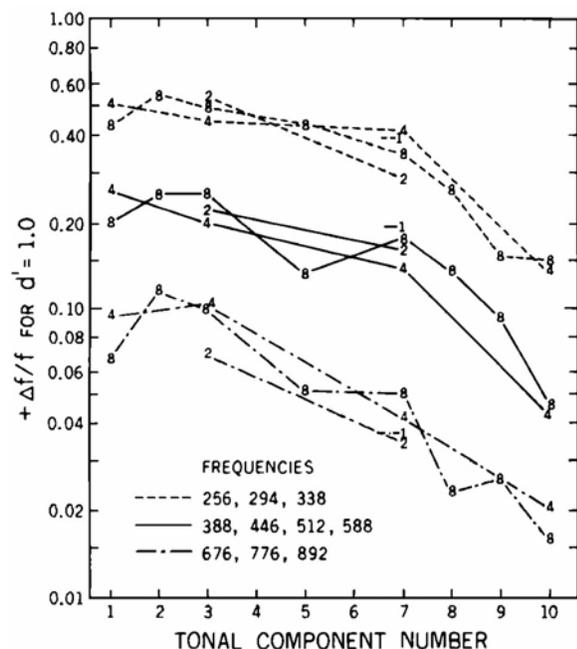


Figure 2. Illustration of strong effects on frequency discrimination of both component frequency and temporal position within the patterns, for four highly trained listeners. Thresholds are plotted separately for the three lowest, four middle, and three highest-frequency tones. The numbers plotted as individual data points represent the number of components subject to change within an experimental session. Data from Watson *et al.* (1), Experiment 3. Reproduced with permission from (1). Copyright 1975, Acoustical Society of America.

Figure 2 illustrates the influences of two variables on frequency discrimination thresholds for ten-tone patterns: component frequency and temporal position. Average thresholds were separately calculated for the three highest frequency components, the four mid-frequency components, and the three lowest frequency components, as indicated in the figure legend. High-frequency components were resolved more accurately than medium, and medium more accurately than low. The second strong effect was that of the temporal position of the target tones, with the least accurate frequency resolution for the first component and the most accurate for the last (tenth). This was not a simple onset or offset effect, but rather one that was gradually distributed over the time course of the pattern. The worst frequency resolution, represented by threshold values of $\Delta f/f$ for the three lowest frequencies in the first temporal position, was approximately 30 times larger than the threshold for high-frequency components in the tenth temporal position.

In a subsequent experiment, the degraded frequency resolution for lower frequency components within the patterns was found to be largely based on the relative rather than on the absolute component frequency. This is illustrated in Figure 3, from an experiment in which two different frequency ranges were employed, 294-892 Hz and 565-1500 Hz (Experiment 4, from Watson *et al.*, 1975). Data points in this figure show the frequency range of the target tones tested in each pattern (only five frequencies of the ten included in each pattern were subject to change).¹

The other variable illustrated in Figure 2 is the number of components subject to change in each block of trials, which is indicated by the numbers utilized as data symbols. No significant change in thresholds resulted from the increased stimulus uncertainty associated with a shift from a single component subject to change on each trial, to as many as eight components subject to change. This apparent lack of influence of a change in stimulus uncertainty was later shown to be a somewhat misleading result. When, as in this experiment, a new sample from a 50-pattern catalog was presented on each trial, stimulus uncertainty was already so high that further variation in it had no effect. Subsequent experiments showed that stimulus uncertainty (or familiarity) is the strongest single factor influencing the discrimination of the tonal sequences.

4. STIMULUS UNCERTAINTY

The generally poor frequency resolution shown by very highly trained listeners, as illustrated in Figures 2 and 3, was observed when the pattern presented on each trial was one of 50 sequential permutations of a ten-frequency catalog. It had been reasoned that to use only a single tonal sequence and a single one of its components as the target, would merely encourage the listeners to learn to ignore everything but the target stimulus. The goal was to determine how accurately listeners can hear *all of the components of a word-length pattern*, not how well they can ignore all but a single critical component. The results, however, led to a reconsideration of this rationale, because

Studies of tone sequence perception

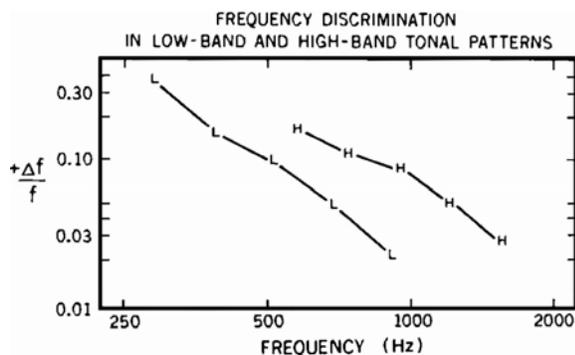


Figure 3. Relative frequency within a pattern is a determinant of frequency discrimination thresholds, in addition to absolute frequency. Average values of $\Delta f/f$ for $d' = 1.0$ are shown for the high- (H) and low-frequency (L) ranges used in Watson *et al.* (1), Experiment 4. Reproduced with permission from (1). Copyright 1975, Acoustical Society of America.

frequency resolution appeared to be so seriously degraded for all but the late-occurring, high-frequency components. Listeners possessed of such poor frequency resolution could hardly be expected to understand speech or enjoy music. Also, theories of the temporal and spectral spread of masking (e.g., 3, 4) do not predict such large degrading effects of the local context.

The control that was clearly needed was the experiment we had previously thought inappropriate: to simply present the same sequential pattern on every trial, within which the same temporal component would be subject to change. This was designated a “minimal uncertainty psychophysical procedure” or Min(U) for short (see definition of a Min(U) procedure in Watson *et al.* (2)). Since by that time a number of tonal-pattern experiments had been conducted, their results were ranked roughly by the level of uncertainty in each experiment. It was found that frequency resolution improved from the severely degraded levels illustrated in Figures 2 and 3 in this article (threshold values of $\Delta f/f > 0.5$) to nearly the accuracy expected for isolated pure tones of the same duration ($\Delta f/f < 0.02$), as stimulus uncertainty was reduced. The parameters of these experiments are shown in Table 1, from Watson *et al.* (2). Experiment 11 in that table was an example of a Min(U) experiment, in which the same pattern was presented on every trial and the second tonal component was always subject to change. The average threshold for $\Delta f/f$ was 0.014, only slightly larger than the values of 0.008 for isolated tones, obtained by the authors in Experiment 12, and by Liang and Chistovich (5), for tones of similar durations. It is important to emphasize that the listeners in all conditions had been trained for thousands of trials. Neither general familiarity with tonal patterns nor procedural confusion differed across these experiments; the only difference was the listeners’ total experience with specific patterns.

Large effects of uncertainty were also shown for the discrimination of changes in duration of pattern components, as well as for the detection of pattern components (as discussed below). Given the magnitude

and the generality of the uncertainty effects, they are probably the most important findings to emerge from this research program. However, the explanation for these effects remains unclear. Although the manipulation of the number of different patterns presented in an experiment has been described in terms of uncertainty, it does not necessarily follow that the effect is due to an informational limit. One alternative explanation is that the ability to focus attention on an individual target component (and, therefore, resolve it with greater accuracy) improves as listeners become more familiar with the spectral-temporal structure of the individual patterns. According to this view, the number of different patterns used in an experiment would be of little consequence if, in advance, a listener had been made very familiar with some or all of the patterns. Although this hypothesis has not been tested with very large numbers of patterns, Spiegel and Watson (6) found that after extensive practice with each of eight patterns under Min(U) testing conditions, subjects performed at approximately the same level when patterns were randomly selected from the eight-pattern set on each trial. Thus, stimulus uncertainty effects may be understood in terms of familiarity and attentional focusing, rather than in terms of an inherent informational limitation on processing. This conclusion is similar to that of Shiffrin and Schneider (7), who were studying similar phenomena in visual search tasks at the same time these auditory studies were conducted. Those authors explained the excellent performance achieved under “consistent mapping” conditions (similar to the Min(U) described here) compared to that under varied mapping conditions (similar to High(U)) in terms of “automaticity”, meaning that visual search had become automatic and was no longer dependent on the number of items in the search set (i.e., targets) or the number of items in the search display. Thus, according to the automaticity theorists, once a stimulus becomes very familiar, discrimination and identification tasks can be performed extremely efficiently, with very little demand on attentional resources. Meaningless visual and auditory patterns appear to be similar in the efficiency with which they are processed, once they have become familiar.

Another test of the concept of an informational limit governing performance was carried out in a series of experiments in which frequency discrimination for a single component was measured in tone sequences with a range of total durations and numbers of components (8, 9). Results from the first set of experiments, in which all patterns were isochronous, suggested that the critical variable was the number of components in the patterns, as shown in Figure 4. The major surprise was the broad range of total durations over which this relationship appeared to hold, from patterns as brief as 125 ms (a medium-duration phoneme) to more than 1500 seconds (the duration of a short phrase or sentence). However, with isochronous patterns, the number of tones is confounded with the proportion of the total pattern duration taken up by the target tone (and by each of the context tones). When proportional duration was manipulated independently of the number of components, requiring the use of anisochronous patterns, it was found that proportional duration of the target tone, relative to total pattern duration, was the critical

Table 1. Frequency discrimination experiments with various stimulus catalogs

Experiment	No. of listeners	No. of experimental sessions 1-h session/day ²	Temporal position of test component	delta-f/f for d' = 1.0	Characteristics of stimulus catalog		
					No. of signal frequencies	No. of temporal components subject to change	No. of patterns in catalog
1	4	23(10)	4	0.5	540	4	2.3x10 ¹⁹
2	4	9(10)	7	0.35	460	7	4.2x10 ²⁶
3	4	17(12)	2	0.25	10	8	50
4	3	10(36)	3	0.14	1	1	50
5	4	14(22)	4	0.10	5	4	50
6	4	5(10)	7	0.09	5	1	50
7	3	27(12)	4	0.07	5	4	5
8	4	1(10)	7	0.06	1	1	50
9	4	6(42)	2	0.06	5	4	20
10	4	8(34)	2	0.019	4	4	1
11	4	8(20)	2	0.014	1	1	1
12	4	10(10)		0.0080	Single 40-msec tones, 1 kHz		
Liang and Chistovich	-- ¹	-- ¹		0.0082	Single 40-msec tones, 1 kHz		
Henning	2	-- ¹		0.0028 ³	Single 40-msec tones, 1 kHz		
Moore	3	10(2h)		0.0024	Single 40-msec tones, 1 kHz		

¹ Data not available ² Quantity in parentheses is number of training sessions prior to data collection. ³ Interpolated value. The stimuli in Experiments 1-11 were sequential patterns of ten 40-msec tones. In all cases except Experiments 1 and 3 the frequency range of the patterns was 500 to 1500 Hz; in those two experiments it was 300 to 3000 and 250-900 Hz, respectively. The frequencies of the ten tonal components in Experiments 3 through 11 were spaced at equi-log intervals over the total frequency range of the patterns. In Experiment 1 the total frequency range was divided into five equi-log intervals, two random samples from each interval comprised a ten-tone pattern. In Experiment 2 the frequencies of the components were randomly sampled from the entire frequency range of the patterns. In all experiments the stimulus on each trial was a random sample from the catalog described in the rightmost three columns of the table *with* the constraint that each combination of characteristics occurred equally often in the cases of Experiments 3 through 11. The test-tone frequencies for the data reported in the table were: Experiment 1, 753 to 1195 Hz; Experiment 2, 800 to 1100 Hz; Experiment 3, 388, 446, 512 or 588 Hz; and Experiments 4 through 12, 920 Hz. Reproduced with permission from (2). Copyright 1975, Acoustical Society of America.

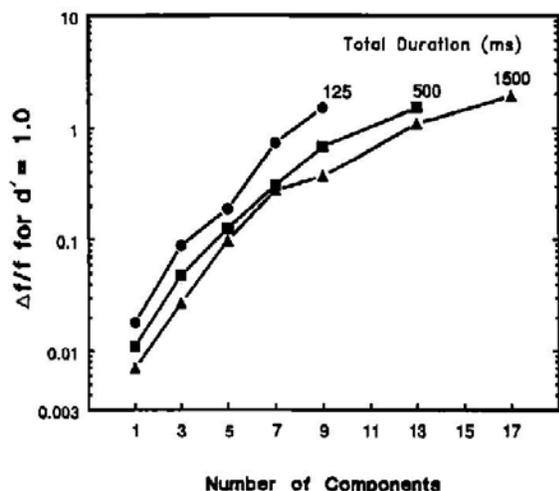


Figure 4. Discrimination thresholds for each of the three total durations, shown as a function of the number of components in the pattern. From Watson *et al.* (8, Experiment 3). Reproduced with permission from (8). Copyright 1975, Acoustical Society of America.

variable (9). As shown in Figure 5, changes in proportional duration had a substantial effect on performance, while large changes in the number of components, in the target tone duration, and in the total pattern duration had relatively small influences.

Later, Lutfi (10) devised an elegant model (designated “CoRE” for Component Relative Entropy), which successfully predicted virtually all of our results with tonal patterns (including those summarized in Table 1) by

formalizing the association between uncertainty and resolving power in a model that also incorporates the proportional duration principle through duration-based weighting of pattern components. Lutfi interpreted the success of the CoRE model as indicating that the effect of proportional duration reflects an informational limitation, expressed in terms of proportional variance. However, Kidd (11) has suggested that it is the duration-weighting component of the model that accounts for the proportional duration effect and that uncertainty or variance-based factors are not responsible for the effect. This point of view is illustrated by a single-trial demonstration in which a relatively small frequency change in a target tone within a sequence is easily detected by most listeners when the proportional duration is large (e.g., 40%), but is very rarely heard by listeners (without training/familiarization) when the proportional duration is 10% or less.

Although the effects of familiarity and proportional duration show that there are factors other than informational limitations that can have a major influence on auditory pattern discrimination, an information-based limitation has certainly not been ruled out. More experiments of the sort conducted by Spiegel and Watson (6), described earlier, are needed to resolve this question. Recapping, in such experiments listeners are trained under Min(U) with each of several patterns, and are then tested with those patterns presented in random order. An elimination of uncertainty effects resulting from this sort of familiarization would be consistent with both an explanation in terms of automaticity (7) and with therecognized abilities to resolve *familiar* subtle phonemic

Studies of tone sequence perception

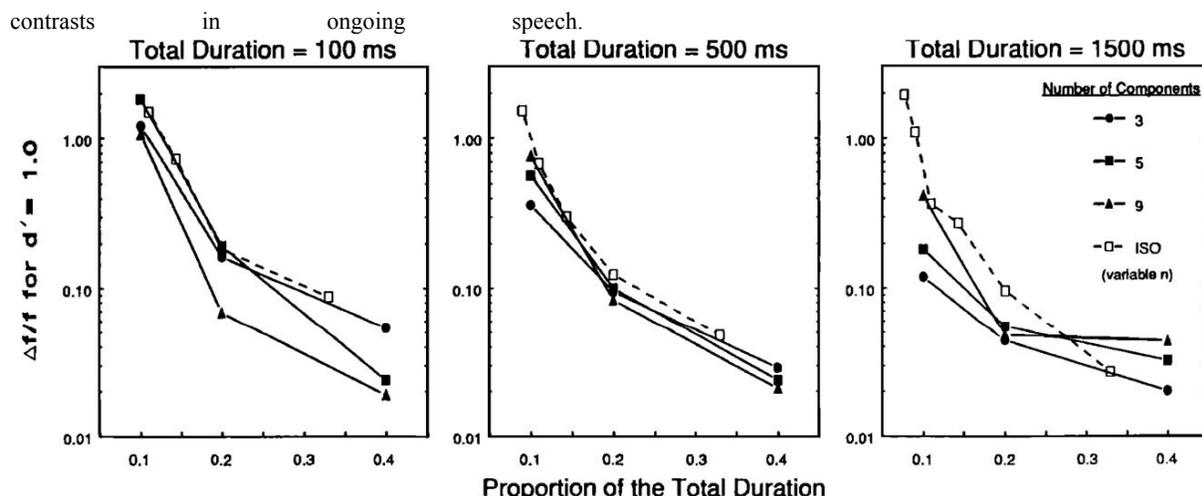


Figure 5. Data from Kidd and Watson (9), Experiment 1, plotted as a function of the target tone's proportion of the total pattern duration (PTD). Although total pattern duration has a substantial effect on performance for patterns with PTD = 0.1 (especially with smaller numbers of components), PTD is clearly the major determinant of performance. Reproduced with permission from (9). Copyright 1975, Acoustical Society of America.

5. TARGET-TONE DISCRIMINATION WITH UNCERTAINTY ABOUT THE DIMENSION OF CHANGE

The initial experiments had demonstrated large effects of the pattern context on the ability to detect changes in the frequency of individual tonal components, under conditions of medium or high trial-to-trial stimulus uncertainty. In an effort to determine whether stimulus uncertainty had similar effects on the detectability of changes in component duration and level, Watson and Kelly (12) conducted what was referred to in the confines of the laboratory as a "super-uncertainty experiment." Listeners were trained for twelve hours to discriminate ten-tone patterns similar to those described earlier, except that their frequency range was increased to 300-3000 Hz. That range was divided into ten equi-log intervals. Eight frequencies were drawn for each pattern, one from each of the ten intervals, to be used as the frequencies of the context (non-target) tones. By this method of selecting the context tones, no pattern was ever repeated in the experiment. The target tone on each trial was either the third or the eighth temporal component and had a frequency of either 534 or 1688 Hz (both target-tone frequency and temporal position were randomly assigned). On each trial the dimension subject to change was randomly selected as either an increment in frequency, intensity, or duration. The data from this very-high-uncertainty condition are shown by the rightmost psychometric functions (labeled "1") in each panel of Figure 6 (from 12). These data are for the 534-Hz target tones and were collected following 12 hours of adaptive-tracking training (approximately 8400 trials) with three normal-hearing listeners. Following this condition, in which listeners heard a new pattern on every trial and had to be alert for changes in component frequency, intensity, and duration, three additional conditions were run with reduced levels of uncertainty.

The psychometric functions labeled "4" show discrimination under Min(U) conditions. In Condition 2, the target position and signal dimension were randomized as under Condition 1, but the frequencies of the context tones were fixed. In Condition 3, the context tones were fixed, the target tone was 534 Hz, and it was always presented as the third tone in the pattern. Discrimination performance thus varied from extremely high threshold increments under "super uncertainty" of over 1000 Hz, 9-10 dB, and 60-80 ms, to values only slightly elevated from those expected for isolated tones, under Min(U) conditions.

6. TEMPORAL DISCRIMINATION UNDER VARIOUS LEVELS OF UNCERTAINTY AND WITH PROLONGED TRAINING

During this same era, several investigators began to conduct studies of listeners' temporal resolving power for speech and speech-like sounds, in some cases hypothesizing that speech and language disorders might be a consequence of "temporal processing disorders" (e.g., Tallal and Piercy, 1973) (13). Returning to the study of a single dimension of the tonal patterns, Espinoza-Varas and Watson (14) trained five listeners with normal hearing to detect temporal increments in the duration of target tones, under three conditions of stimulus uncertainty. A duration increment was added to a single tone in a ten-tone pattern, while all other tone durations remained constant (resulting in an increase in the total pattern duration as well as in the target tone). A ten-frequency catalog, ranging from 500-1500 Hz, was used to generate the patterns. In Condition 1, the target tone was randomly selected to have one of the five different frequencies (565, 721, 920, 1175 or 1500 Hz) and could occur at sequential position 1, 4, 7 or 10. Frequencies of the remaining nontarget tones were randomized, yielding 9! possible temporal sequences for each of the 20 possible targets. In uncertainty condition 2, the target-tone positions were selected as in Condition 1,

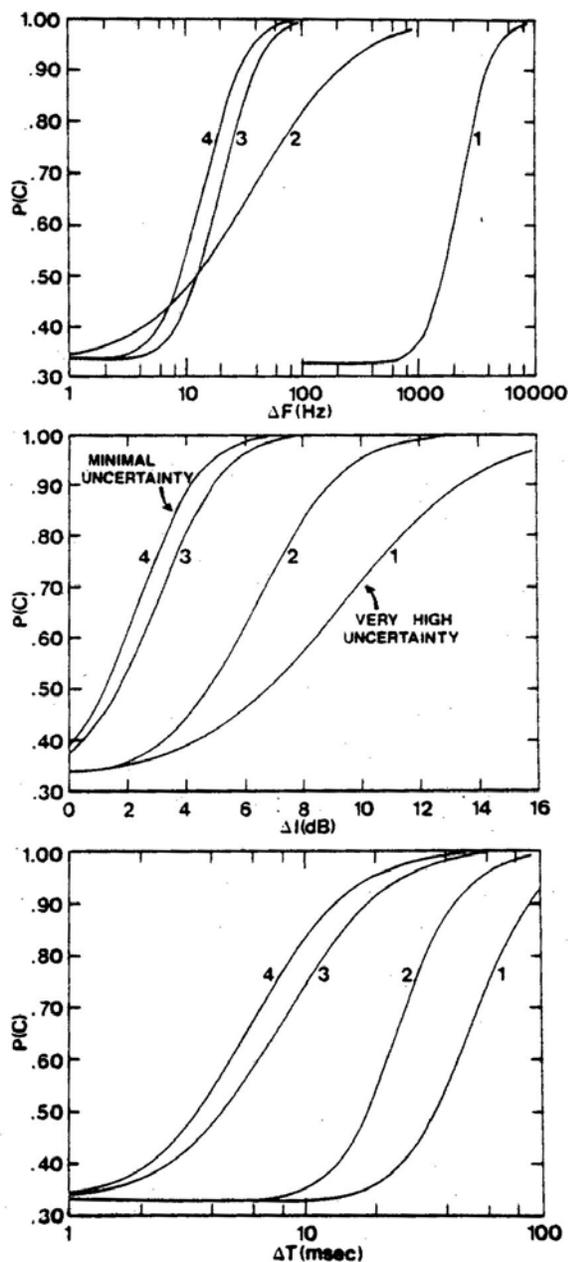


Figure 6. Mean-slope, mean-intercept psychometric functions fitted to pattern-discrimination data collected from three observers after 12 hours of training. Functions labeled 1 are for a very high level of stimulus uncertainty in which the “signals” to be detected were changes in the frequency, intensity, or duration of single components of a new pattern presented on each trial. Those labeled 4 are for a minimal-uncertainty replication of that experiment. Functions 2 and 3 are for intermediate levels of stimulus uncertainty, as described in the text. Reproduced with permission from Charles S. Watson and Lawrence Erlbaum Associates (12).

but only ten different sequences of context frequencies were used. In uncertainty condition 3 Min(U) training was conducted for four different patterns (a-d in Figure 7), each

of which also occurred under each of the other uncertainty conditions.

Figure 7 shows learning data for these three conditions, for four patterns, labeled a-d, obtained over the course of about 75,000 trials. Even after such prolonged training there was little indication that performance on these duration-discrimination tasks was approaching asymptotic levels, for any single level of uncertainty. It was clear, however, that changes in trial-to-trial uncertainty were associated with significant reductions in the discrimination thresholds. A simple summary of these results is that the thresholds were roughly predicted by the Weber ratio ($\Delta t/t$) of about 0.10, obtained in numerous earlier studies of temporal discrimination (e.g., 15). However, that ratio under the high uncertainty condition appeared to be applied to the total pattern duration, while under Min(U) it was applied to the duration of individual tonal components. In a control experiment to test this hypothesis, temporal increments were added uniformly to all of the components in a ten-tone pattern, rather than to a single target component. This distributed-increment procedure yielded roughly the same threshold, in terms of total Δt as did the original condition, in which the increment was added to only a single component. That is, under the high level of stimulus uncertainty, the listeners were basing their judgments on changes in total pattern duration, rather than on the duration of the target component. The observation, under higher levels of stimulus uncertainty, that at least 15,000 to 20,000 trials are required to approach asymptotic performance is worth keeping in mind when considering the results of many temporal discrimination experiments that have used far less training.

7. DETECTION OF AUDITORY PATTERN COMPONENTS AND INFORMATIONAL MASKING (IM)

The effects of stimulus uncertainty described in previous sections were sometimes termed instances of “informational masking” (IM) in the reports of those experiments (e.g., 16), although those effects were on discrimination performance, rather than detection. Some observations did suggest, however, that the *effective levels* of the target tones were reduced under high uncertainty conditions.

In the early experiments with tonal patterns, it was noticed that the degraded frequency resolution under high uncertainty conditions was approximately what would be expected if the levels of the target components were reduced by 40-50 dB relative to the contextual components (normally presented at 75 dB, SPL). Spiegel and Watson (6) investigated this apparent reduction in effective level by actually decreasing or increasing the target-tone levels by various amounts, relative to the context tones, under minimal- and high-uncertainty conditions. Significantly degraded frequency resolution, relative to that expected with isolated tones, was observed under High(U), until the target tones were increased to about +10 dB, while accurate

Studies of Tone Sequence Perception

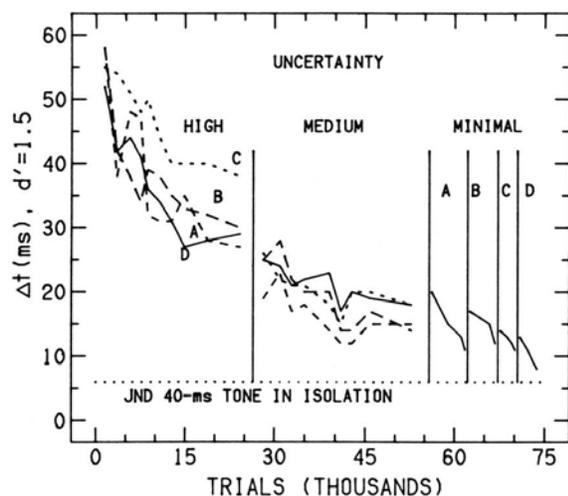


Figure 7. Improvement in threshold values of delta-t as a result of lengthy training, obtained under high, medium, and minimal stimulus uncertainty, from Espinoza-Varas and Watson (14), Experiment 1. The temporal position and frequency of the test tones were: 4/1175 Hz (A); 7/565 Hz (B); 4/565 Hz (C); and 7/1175 Hz (D). The dotted horizontal line shows the expected threshold for 40-ms tones presented in isolation (cf. Creelman, 1962) (15). Because of the large number of test-tone conditions under high and medium uncertainty, the data of three to five sessions were collapsed to obtain the data points. The data points are plotted at the midpoint of the time interval over which data were collected. The total number of sessions from the beginning to the end of this experimental series is 119, distributed over a period of approximately 8 months. Reproduced with permission from (14). Copyright 1975, Acoustical Society of America.

discrimination was maintained under Min(U) until the target tones were reduced by almost 50 dB, relative to the context tones. This was interpreted as representing an IM effect since the same interfering stimuli (the context tones) resulted in severe degradation in performance under one level of uncertainty and had little or no effect under another. (Watson (17) discusses the difference between IM resulting from stimulus uncertainty and that associated with target-masker similarity.)

Leek and Watson (18) shifted the basic experimental paradigm from discrimination to the measurement of detection thresholds for target components, presenting the same pattern on every trial, but with uncertainty about which component would be designated the detection target (as illustrated in the lower panel of Figure 1). The results were consistent with the previous frequency discrimination studies, in that the degree of threshold shift depended on the frequency and temporal position of the target components, with more masking for earlier-occurring and lower-frequency components. With prolonged training under these conditions (8,000-10,000 trials), it was found that the detection thresholds for all components dropped from an initial 65-75 dB to 25-30 dB. The average time course of learning showed this change to occur quite gradually over several thousand trials, or about 10-15 hours of testing. Two additional observations were

made in this experiment. One was that some subjects showed patterns of learning to detect individual components in which thresholds might remain at 70+ dB for as many as 4000 trials and then drop down to 20-30 dB almost as rapidly as the adaptive tracking procedure would permit (i.e., in 15-20 trials). More often, learning progressed in a gradual fashion. It was instructive that the precipitous changes would not have been noticed had the data been averaged over subjects and conditions. The other surprising result was that when the same listeners were later tested with six additional patterns (each of which was studied for 3000 trials, under minimal uncertainty), very little informational masking was observed after the initial 100 or so trials with each of the new patterns, contrasting with the thousands of trials required for the first pattern. During their experience with the first pattern they had apparently (based on the data and on their verbal reports) learned to "listen for faint components in the background of strong ones."

8. WHAT DO LISTENERS LEARN THROUGH PATTERN DISCRIMINATION TRAINING? TRANSPOSITION EXPERIMENTS

As described in the previous sections, listeners' abilities to detect small changes in the frequency, intensity, or duration of pattern components (or to detect those components at reduced levels) improve following prolonged training. But exactly what they are learning about the patterns, over the course of such training, is not clear. Undoubtedly listeners learn about the context as well as the target tones and about properties of the patterns other than the specific frequencies and durations of the tones. At least some incidental learning about the patterns seems very likely, but not all of this learning may be incidental, in the sense that it is irrelevant to the detection of changes in the target tones. That is, knowledge of the pattern structure may be useful, either in helping to focus attention on target components, or in identifying relational information associated with target-tone changes (e.g., changes in relative frequencies or in relative durations among the target and nontarget components).

One approach to this question would be to determine the detectability of a change in a previously untested component, following prolonged minimal uncertainty training to detect changes in a different component. We have yet to conduct such a straightforward test of incidental learning. A different approach was taken by Kidd and Watson (19), who asked whether listeners learn to detect absolute changes in target tones or relative changes between target and context tones. These experiments used the modified 2AFC procedure Figure 1, but, in this case, the two comparison patterns were transposed (i.e., all frequencies increased by the same percentage) such that the relative frequencies of the standard pattern were preserved while the absolute frequencies were increased. On each trial, the two comparison patterns were transposed by the same randomly selected amount, ranging from 2 to 12 semitones. A frequency increment

Studies of Tone Sequence Perception

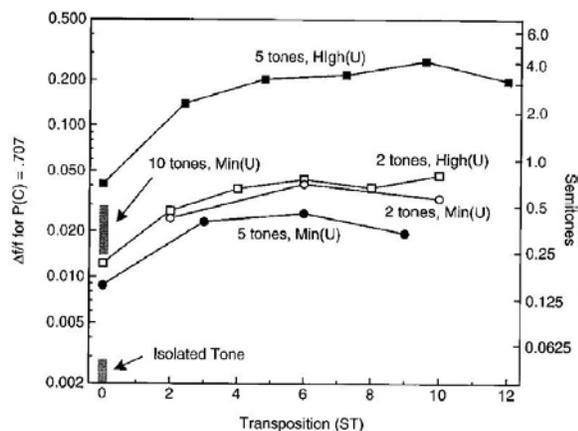


Figure 8. The major effects of frequency transposition on pattern discrimination, from Kidd and Watson (19). Overall performance levels are consistent with earlier frequency-discrimination experiments (dark bar labeled “10 tones”), as are relative levels of performance under different levels of uncertainty and stimulus complexity. A large uncertainty effect is apparent for five-tone patterns, but not for two-tone patterns, and the difference between performance with five-tone and two-tone patterns is minimized when subjects become familiar with the five-tone patterns. Reproduced with permission from (19). Copyright 1975, Acoustical Society of America.

or decrement was applied to the target tone in one of the transposed comparison patterns. Thus, subjects could not perform the task based on absolute frequency changes between standard and comparison patterns or between the comparison patterns (e.g., by choosing the comparison pattern with a higher average frequency).

Subjects were tested under Min(U) and High(U) conditions to determine the effect of transposition for familiar and unfamiliar patterns. Under Min(U), the same pattern was presented on every trial, while it was changed on every trial under High(U). Patterns consisted of either five tones or only two tones. Five-tone (rather than ten-tone) patterns were used to lower the level of difficulty slightly under High(U) conditions. Two-tone patterns were used to assess the processing of relative frequency information with the minimum number of components required to do so.

Figure 8 illustrates the major results of the transposition experiments. When required to detect pattern changes solely on the basis of relative frequency, listeners can still detect changes in component frequency, although not as well as in the non-transformed baseline condition when both absolute- and relative-frequency cues are available. Although there is a slight tendency for thresholds to increase with the amount of transposition (especially with longer patterns and higher uncertainty) these functions are relatively flat for levels of transposition from 2 to 12 semitones. As would be expected, frequency discrimination for five-tone patterns was more severely degraded by transposition under high uncertainty than under minimal uncertainty conditions. Because the

difference between High(U) and Min(U) performance increases with the amount of transposition, it appears that Min(U) training facilitates the learning of relative-frequency relationships even more than it facilitates the detection of changes in absolute frequency.

Performance with two-tone patterns (open symbols in Figure 8) demonstrates the expected minimal effects of changes in uncertainty when the target's proportion of the total pattern duration is as large as 0.5. This result may also reflect the simplicity of having but a single frequency interval on which to base a decision. That performance under minimal uncertainty is as good or better with five as with two tones suggests that the surrounding context can be helpful when the task demands reliance on relative-frequency information.

9. ON THE RELATION BETWEEN THE PROCESSING OF SPEECH AND OF NONSPEECH COMPLEX STIMULI (WORD-LENGTH TONAL PATTERNS)

As noted in the introduction to this essay, when these experiments were begun, an unanswered question was whether small perturbations in the waveforms of an utterance could be detected more accurately than like-size changes in comparably complex laboratory-generated nonspeech sounds. The answer is now clear, and it is “yes and no.” Under high-uncertainty testing conditions, changes in the individual tonal components of 0.5-sec ten-tone patterns are very poorly resolved. So badly, in fact, that a listener equipped with the width of filters and apparently coarse temporal processor that is implied by the High(U) data would probably find speech quite unintelligible. On the other hand, the spectral-temporal acuity with which these same nonspeech stimuli can be processed under Min(U) conditions exceeds that demonstrated in all but a very few speech experiments. It is important to emphasize that the poor performance under High(U) was observed after many days and even weeks of training with those stimuli. These are not effects that are mitigated by becoming generally familiar with this class of stimuli or with the task.

An obvious objection to this attempt at a rapprochement between our understanding of the differences between the processing of speech and nonspeech stimuli is that it might appear to require that words be presented under Min(U) conditions if they are to be resolved with the spectral-temporal precision demonstrated for tonal patterns under those same conditions. That is, we would resolve “baseball” very well if it were said over and over again by the same talker. There are two answers to this concern. One is that, as demonstrated in vocoder experiments, as well as in High(U) experiments with speech (20), speech is not routinely resolved with anything close to the accuracy demonstrated for tonal patterns under Min(U). The other is the demonstration by Spiegel & Watson (6) that after listeners have been trained to discriminate small changes in eight different tonal patterns, each trained under Min(U) conditions, the patterns can be presented in random order and the precise Min(U) discrimination skills are

Studies of Tone Sequence Perception

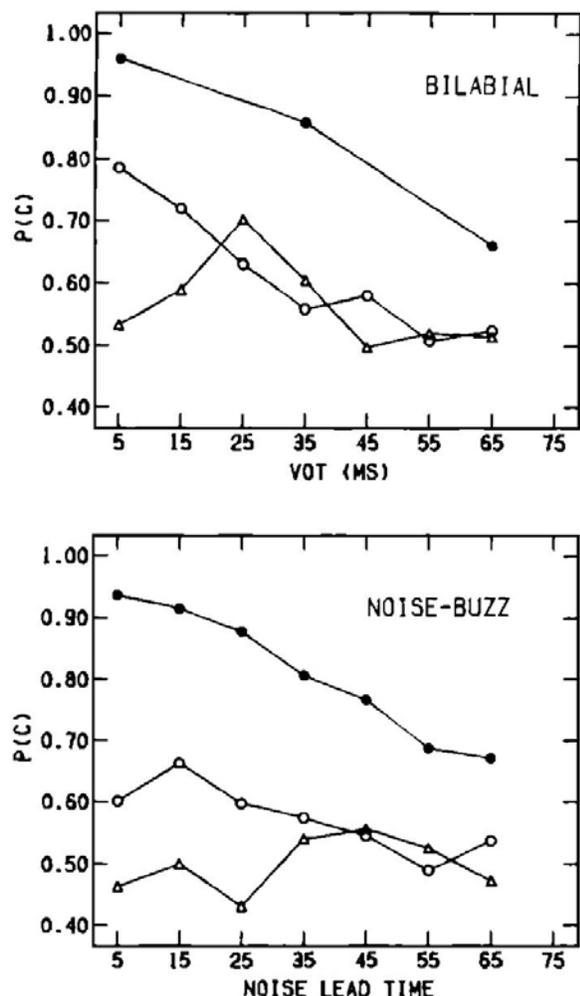


Figure 9. Effects of stimulus uncertainty on discrimination results from bilabial VOT experiments (upper panel), and from noise-lead-time experiments (lower panel). Data from the same-different tasks are shown with filled circles for minimal uncertainty (Δ -VOT=5 ms) and with open circles for high uncertainty (Δ -VOT=10 ms). Data from an ABX task are shown with open triangles (Δ -VOT=10 ms). From Kewley-Port, Watson, and Foyle (20). Reproduced with permission from (20). Copyright 1975, Acoustical Society of America.

maintained. They are now familiar stimuli, just as are the words of one's native language

Training and testing under minimal uncertainty conditions has been shown to have effects on speech sounds similar to those described above for tonal patterns. *Speech sounds* that are not discriminable under High(U) testing can be made so by resorting to Min(U) testing, even if those sounds are within the same speech category. Kewley-Port *et al.* (20) demonstrated this, as shown in Figure 9. Not only could the within-category sounds of synthesized tokens of /ba/ with voice-onset times (VOTs) less than those at the category boundary between /ba/ and

/pa/ (about 20-25 ms) be distinguished accurately, a ten ms increment in VOT was most accurately detected at VOT=5 ms. Discrimination performance decreased monotonically with increasing values of VOT, rather than showing the nonmonotonic relation to VOT obtained in many earlier experiments conducted using higher levels of uncertainty. Discrimination of VOT differences under Min(U) conditions is thus consistent with Weber's Law, which describes discrimination performance for many other sensory stimuli. This appears to be a lesson worth recalling when attributing psychophysical performance for a specific class of sounds to some inherent peculiarity of the processor. Particularly accurate resolution of some spectral-temporal detail within a complex waveform may simply be a consequence of having learned that this detail normally bears more information than some other parts of the waveform (e.g., the categorical boundary between /pa/ and /ba/). Given an opportunity and an incentive to learn to attend to other portions of the waveform, listeners may be perfectly able to do so.

10. INDIVIDUAL DIFFERENCES IN DISCRIMINATION AND RECOGNITION OF COMPLEX SOUNDS

The preceding sections have described effects on pattern discrimination that are common to most listeners. It was observed early on, however, that differences among listeners in complex pattern discrimination experiments tend to be considerably larger than those found in studies with simpler stimuli, especially with isolated tones. Large individual differences are also observed in the case of patterns composed of multiple simultaneous tones, termed "profiles" by Green and his colleagues (see 21). This has been illustrated in studies by Neff and Dethlefs (22) and by Drennan and Watson (23), each of which concluded that the distribution of performance in these tasks is essentially unimodal, rather than bimodal, indicating that there are not two distinct types of processing, or processors. Prolonged training greatly reduces the range of performance, with the initially poor performers tending to catch up to those who have the best initial thresholds, but existing evidence suggests that the rank ordering of performance tends to be maintained despite lengthy training.

Studies of individual differences using auditory test batteries consisting of various single-tone and multitone spectral-temporal discrimination tasks have shown there to be little commonality between the abilities to resolve details of simple or complex unfamiliar laboratory test sounds and the recognition of speech or other familiar sounds under difficult listening conditions (24, 25). In the presence of masking noises, the best listeners can identify familiar sounds at much lower signal-to-noise ratios than can the worst. The S/N for the worst listeners can be 7-10 dB higher than for the best, for the same level of performance. Apparently the best listeners can identify a familiar word or an environmental sound despite portions of the sound being inaudible, while much more of the complex waveform is

Studies of Tone Sequence Perception

required by the worst listeners. This may imply that the best listeners have either more complete and accurately stored representations, greater skill in recognizing stimuli based on incomplete information, or some combination of the two. Whatever the explanation, the individual-differences studies suggest that speech and other familiar everyday sounds may be *special* because of their familiarity rather than because of a unique mode of processing peculiar to sounds of a particular stimulus category.²

11. SUMMARY

The ability to detect or to discriminate changes in brief individual components of unfamiliar complex sounds under conditions of high trial-to-trial uncertainty is substantially inferior to that for familiar sounds, or sounds that have become familiar through minimal uncertainty training. The effective levels of pattern components are often reduced by as much as 30-50 dB, compared to thresholds for the same components within very familiar sounds. Lower-frequency components, those occurring earlier in patterns, and those that occupy a small proportion (less than about 20 per cent) of the total pattern duration are most severely affected. After extensive practice with unfamiliar sounds under minimal uncertainty psycho-physical procedures, the effects of trial-to-trial pattern uncertainty are greatly reduced or eliminated. It appears that pattern familiarity helps listeners focus attention on individual components more effectively, despite high levels of trial-to-trial uncertainty. Transposition experiments have begun to probe the question of what exactly is learned as a tonal pattern becomes familiar, making it clear that listeners learn more than just the specific frequencies and durations of the components. This is suggestive of humans' abilities to recognize syllable constituents as produced by many different vocal tracts.

There are large individual differences in the ability to recognize very familiar patterns, including both speech and non-speech sounds, under listening conditions that render portions of the sounds inaudible. It is clear that individual differences in spectral and temporal resolving power measured with single tones and unfamiliar complex sounds is surprisingly independent of the speech-to-noise ratio required to recognize speech, for normal-hearing listeners. Some listeners need to hear much more of the total patterns than do others in order to recognize familiar sounds. The extent to which the range of individual abilities to recognize speech and other familiar sounds is due to differences in memory capacity, selective attention, or perhaps some general ability to make effective use of incomplete information about familiar patterns remains to be determined, but these are the candidates.

12. ACKNOWLEDGMENTS

This research has been supported by grants from several agencies over the past three decades. Those agencies include the National Institutes of Health, the Office of Naval Research, and the Air Force Office of Scientific Research.

13. REFERENCES

1. Watson, C. S., Wroton, H. W., Kelly, W. J., & Benbassat, C. A.: Factors in the discrimination of tonal patterns. I. Component frequency, temporal position and silent intervals. *J. Acoust. Soc. Am.*, 57, 1175-1185 (1975)
2. Watson, C. S., Kelly, W. J., & Wroton, H. W.: Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. *J. Acoust. Soc. Am.*, 60, 1176-1186 (1976)
3. Elliott, L. L.: Backward and forward masking of probe tones of different frequencies. *J. Acoust. Soc. Am.*, 34, 1116-1118 (1962)
4. Widin, G. P., & Viemeister, N. F.: Masker interaction in pure-tone forward masking. *J. Acoust. Soc. Am.*, 68, 475-479 (1980)
5. Liang, C. A., & Chistovich, L. A.: Frequency difference limens as a function of tonal duration. *Sov. Phys. Acoust.*, 6, 75-80 (1961)
6. Spiegel, M. F., & Watson, C. S.: Factors in the discrimination of tonal patterns. III. Selective attention and the level of target tones. *J. Acoust. Soc. Am.*, 69, 223-230 (1981)
7. Shiffrin, R. M., & Schneider, W.: Controlled and automatic human information processing: 2. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190 (1977)
8. Watson, C. S., Foyle, D. C., & Kidd, G. R.: Limits of auditory pattern discrimination for patterns with various durations and numbers of components. *J. Acoust. Soc. Am.*, 88, 2631-2638 (1990)
9. Kidd, G. R., & Watson, C. S.: The "proportion-of-the-total-duration rule" for the discrimination of auditory patterns. *J. Acoust. Soc. Am.*, 92, 3109-3118 (1992)
10. Lutfi, R. A.: A model of auditory pattern analysis based on component-relative-entropy. *J. Acoust. Soc. Am.*, 94, 748-758 (1993)
11. Kidd, G. R.: The influence of pattern-based frequency expectations on resolving power for components of multi-tone sequences. *J. Acoust. Soc. Am.*, 97, Pt. 2, 3271 (1995)
12. Watson, C. S., & Kelly, W. J.: The role of stimulus uncertainty in the discrimination of auditory patterns. In D.J. Getty & J. H. Howard (Eds.), *Auditory and Visual Pattern Recognition* (pp. 37-59). Hillsdale, NJ: Lawrence Erlbaum (1981)
13. Tallal, P. & Piercy, M.: Developmental aphasia: Impaired rate of nonverbal processing as a function of sensory modality. *Neuropsychologia* 11: 389-398 (1973)
14. Espinoza-Varas, B., & Watson, C.S.: Temporal discrimination for single components of nonspeech patterns. *J. Acoust. Soc. Am.*, 80, 1685-1694 (1986)
15. Creelman, C. D.: Human discrimination of auditory duration. *J. Acoust. Soc. Am.*, 34, 582-593 (1962)
16. Watson, C. S.: Uncertainty, informational masking, and the capacity of immediate auditory memory. In W. A. Yost & C. S. Watson (Eds.), *Auditory Processing of Complex Sounds* (pp. 267-277). Hillsdale, NJ: Lawrence Erlbaum (1987)
17. Watson, C. S.: Some comments on informational masking. *Acustica*, 91, 502-512 (2005)
18. Leek, M. R., & Watson, C. S.: Learning to detect auditory pattern components. *J. Acoust. Soc. Am.*, 76, 1037-1044 (1984)

Studies of Tone Sequence Perception

19. Kidd, G. R., & Watson, C. S.: Detection of frequency changes in transposed sequences of tones. *J. Acoust. Soc. Am.*, 99, 553-566 (1996)
20. Kewley-Port, D., Watson, C. S., & Foyle, D. C.: Auditory temporal acuity in relation to category boundaries; speech and nonspeech stimuli. *J. Acoust. Soc. Am.*, 83, 1133-1145 (1988)
21. Green, D. M.: *Profile analysis: Auditory intensity discrimination*. Oxford Psychology Series, No. 13. New York: Oxford University Press (1988)
22. Neff, D. L., & Dethlefs, T. M.: Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *J. Acoust. Soc. Am.* 98, 125-134 (1995)
23. Drennan, W., & Watson, C. S.: Sources of variation in profile analysis. I. Individual differences and extended training. *J. Acoust. Soc. Am.* 110, 2491-2497 (2001)
24. Surprenant, A. M., & Watson, C. S.: Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *J. Acoust. Soc. Am.*, 110, 2085-2095 (2001)
25. Watson, C. S., & Kidd, G. R.: On the lack of association between basic auditory abilities, speech processing, and other cognitive skills. *Seminars in Hearing*, 23, 83-93 (2002)
26. Luce, P. A., & Pisoni, D. B.: Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, 19, 1-36 (1998)

Footnotes: ¹ The results from conditions in which a negative delta-f was used have been omitted from Figures 2 and 3. In general, the findings are similar to those using a positive delta-f, but the pattern of results is more complex. This is because negative frequency changes applied to the lower-frequency tones (which are generally more difficult to resolve than higher-frequency tones) become more salient as the frequency of the altered target tone approaches or exceeds the lower bound of the pattern. This leads to inverted-U-shaped functions of threshold versus frequency for lower tones. The same boundary phenomenon also occurs with positive delta-f values, but this does not complicate the threshold-by-frequency functions because the boundary effect and the within-pattern frequency effect both lead to decreased thresholds with increasing frequency. ² No mention has been made of the influence of similarities among the sounds that have become familiar, but recent research has clearly demonstrated the influence of this variable on word recognition, termed “lexical-density” effects (e.g., 26). These effects could easily be modeled with non-speech stimuli such as tonal patterns.

Key Words: Psychoacoustics, Auditory Pattern Perception, Review

Send correspondence to: Dr. Charles Watson, Department of Speech and Hearing Sciences, Indiana University, 200 S. Jordan Ave., Bloomington, IN 47405 Tel: 812-855-4621, Fax: 812-855-5531, E-mail: watson@indiana.edu

<http://www.bioscience.org/current/vol12.htm>