NIH Public Access
Author Manuscript
Hear J. Author manuscript; available in PMC 2010 January 21.

Published in final edited form as:
Hear J. 2008 September 1; 61(9): 26.

Training listeners to identify the sounds of speech: I. A review of past studies

Charles S. Watson, PhD,
President of Communication Disorders Technology, Inc., (CDT) of Bloomington, IN, and Professor Emeritus in Speech and Hearing Sciences at Indiana University.

James D. Miller, PhD,
Principal Scientist at CDT, an Adjunct Professor of Speech and Hearing Sciences at Indiana University, and Director of Research Emeritus at Central Institute for the Deaf, St. Louis.

Diane Kewley-Port, PhD,
Vice-President of CDT and Professor of Speech and Hearing Sciences at Indiana University.

Larry E. Humes, PhD, and
Professor of Speech and Hearing Sciences at Indiana University.

Frederic L. Wightman, PhD
Research Director of the Heuser Hearing Institute and Professor of Psychology and Brain Sciences and of Surgery at the University of Louisville.

In this first section of a two-part article we offer an overview of research on listeners' abilities to learn to hear the spectral-temporal details of simple and complex sounds, both speech and non-speech. In the second article, to appear in the October issue of *The Hearing Journal*, we will describe a new training system based on that science that has recently completed clinical validation trials with users of hearing aids and of cochlear implants.[1,2]

DO WE HEAR WHATEVER ENTERS OUR EARS?

It is tempting to believe that within the limits set by our audiograms we hear whatever sounds happen to stray into our ear canals, and that the way we attend to sounds or our past experiences with them do not make a lot of difference to our auditory experience. (It should be noted that speech scientists have long suspected that this belief is probably wrong; the reasons for their suspicions are discussed later.)

In the case of vision, on the other hand, most of us - realize that a particular visual stimulus can be seen more than one way. This becomes abundantly clear when we view a picture, such as one of the panels of Figure 1, and see that it varies in appearance from moment to moment, even though we know that the patterns of ink on paper remain constant. We are only rarely confronted by similarly compelling evidence that the same acoustic waveform entering our ear canals can also yield very different auditory percepts from time to time or person to person—but indeed that is the case.

Objective evidence of the impact of selective attention and experience on the way sounds are heard has been accumulating ever since it became possible to generate and reliably repeat complex auditory test stimuli, roughly the past 50 years. That evidence comes from several different sources, including studies of the development of speech perception in very young

Readers are invited to contact Dr. Watson at watson@indiana.edu or Dr. Miller at jamdmill@indiana.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

infants, the learning of second languages by adults, learning to discriminate and identify complex laboratory test sounds, and the range of individual differences in the auditory capabilities of adults with normal audiograms.

Additional evidence has recently been provided by studies of the effects of auditory training on speech recognition by hearing aid and cochlear implant users.[2-6] In this article we will give a few highlights from studies of auditory perceptual learning and discuss the relevance of this body of knowledge to auditory training for users of hearing aids and cochlear implants.

HOW INFANTS DEVELOP SPEECH PERCEPTION

We learn our native language during the first few years of life, apparently doing so without much effort and at a remarkable rate of acquisition of the lexicon. Our vocabulary actually grows gradually at first, with only 20 to 40 words acquired by 15 months of age. But, by the age of 4 years, we are acquiring a dozen new words a day, and a 7-year-old may learn 20 words a day (more than one per waking hour).[7]

Those early months were spent learning the speech code, and only after it was fairly well mastered could vocabulary building begin in earnest. Acquiring that code meant becoming sensitive to the acoustic contrasts that are significant (phonetic) in one's language, while also learning to ignore other acoustic variations that are not.[8,9]

Evidence collected with laboratory-generated sounds suggests a limit on the amount of information a listener can extract from a complex sound. Thus it is an efficient strategy to learn to attend to the critical features of speech in one's own language, while reducing the processing of non-meaningful variation of speech sounds.

One example of this process was provided in a study of the discrimination of a contrast that is meaningful in Hindi, but is unfamiliar to English speakers, /t(h)a/ versus /d(h)a/.[8] Speakers of Hindi produce and perceive this distinction quite reliably, while English speakers cannot. It was found, using a conditioning paradigm, that 7-month-old infants (children of English parents) could make this discrimination as well as adult Hindis, while adult English speakers failed miserably at it. But it was reassuring to also learn that adult English speakers could be trained to make the discrimination almost as well as adult Hindis. Adults can learn to recognize speech sounds on the basis of cues to which they previously appeared to be insensitive.

SECOND-LANGUAGE LEARNING

There has been a rapid growth of studies of second-language learning, to some degree motivated by the enormous number of non-native speakers of English seeking proficiency in that language. Many foreign students come to the U.S. with excellent literacy skills but poor pronunciation, and they experience great difficulty in understanding speech at normal conversational rates. It has become clear that the development of pronunciation and perception are intimately connected in these students.

Among the more familiar contrasts that are difficult for non-native speakers of English is /r/ versus /l/, as in “rake” versus “lake,” for students from China, Japan, and Korea. Training studies have shown that recognition of this contrast can be learned, but that the training regimen must meet certain criteria. The corpus of training material, most importantly, must include tokens of each of the sounds recorded by multiple speakers, since otherwise the learning can be specific to a single speaker.[9] In addition, it was found that the positive effects of learning to recognize a specific contrast survived at least several months after the end of the training period.[10]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

LEARNING TO IDENTIFY COMPLEX (NON-SPEECH) SOUNDS

Three recent lines of research on persons with normal hearing are relevant to the possibility that auditory perceptual training might be an effective means of improving speech perception by hearing-impaired listeners. These are: (1) the time course of auditory perceptual learning; (2) the extent of auditory learning, in terms of the magnitude of the changes in spectral or temporal discrimination abilities that can result from auditory training; and (3) the results of large-scale studies of individual differences in auditory abilities.

Auditory perceptual learning

Reviews of the literature on the learning of non-speech auditory tasks have revealed a remarkable range of times required to achieve the best possible performance for human listeners (i.e., to approach asymptotic discrimination or identification thresholds).[11,12] Figure 2 shows data from a study of the time required to approach asymptotic thresholds in a temporal discrimination task.[13] The listeners heard a sequence of 10 50-ms tones with frequencies from 300 to 3000 Hz, roughly the duration of many polysyllabic words. The sequence was repeated three times and the listener's task was to choose the sequence in which one of the 10 tones was made longer than 50 msec.

The lines labeled A, B, C, and D in Figure 2 represent four different 10-tone patterns. The figure is divided into three sections representing different levels of stimulus uncertainty used in the training. In the condition labeled “high uncertainty,” the tone pattern was varied from trial to trial, sampling from a catalog of 50 such patterns. Under “medium uncertainty” only four patterns were used, with a random selection of pattern A, B, C, or D presented on each trial. Under “minimal uncertainty” the same pattern was used on every trial. About 700-800 trials were run each day and the listeners participated for about 100 consecutive training sessions. By the end of training they were able to detect increments of 8-10 ms in one 50-ms component of a 500-ms tonal pattern. This is a level of temporal acuity that had been documented only for such over-learned sounds as the change in voice-onset time that distinguishes /pa/ from /ba/.

The duration of auditory training required to approach asymptotic threshold performance depends on two variables: the nature of the task and the complexity of the stimuli. As shown in Table 1, the fastest task to learn is simple detection of a sound, while discrimination of one sound from another takes somewhat longer, and the most time-demanding task is stimulus identification. Each of these tasks is accomplished most quickly with simple stimuli (e.g., pure tones), but far more slowly with complex stimuli (e.g., word-length tonal patterns or Morse code).

The fastest training, that for detection of simple stimuli, can be completed in 30 minutes to 2 hours. The slowest, that for identification of complex sounds, can require from 20-30 hours to many months. In a famous auditory training study, published in 1899, the authors kept track of improvements in the recognition of Morse code as trainees in telegraphy progressed from recognizing only 5-10 five-letter sets per minute to 60-70 sets per minute over the course of 6 months of training, 40 hours per week.[14] The telegraph operators' improvement over this period of time was nearly linear and showed little sign of having reached an upper limit.

Such observations suggest an enormous capacity for perceptual learning that has seldom been tapped in auditory research, where 30-40 total hours of training is exceptional. The telegraphy data, some data from the learning of word-length tonal patterns, and the time ESL learners require to learn to recognize the sounds of a new language all point to gradual improvement in auditory identifications skills over at least 200-300 hours of practice. Fortunately, for those attempting to improve speech-recognition skills, although structured training with immediate

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

corrective feedback after incorrect responses may be the most efficient form of training, informal everyday experience may be beneficially combined with formal training to achieve the desired goals.

Magnitude of changes accomplished by auditory training

The data in Figure 2 showed a reduction of the threshold for increments in the duration of a single 50-ms component of a word-length tonal pattern, from 60 ms down to less than 10 ms. There are many such instances of major changes in auditory acuity for specific sounds. A particularly compelling example was one in which listeners were trained to distinguish the presence or absence of a single component of one of the word-length tonal patterns that had been used in the study of temporal discrimination training described above.[13,24] During the first week or so of training the 50-ms component to be detected could not be heard if it was a few dB below the 75 dB at which the other nine components were presented. After 10-15 hours of training the listeners could detect the presence of any one of the components, even when they were reduced by as much as 40-45 dB below the level of the other components.

This achievement of learned auditory attentional focus was only remarkable in that it was made to occur in the research lab. Babies and some, but not all, second-language learners and hearing aid users, appear to be able to accomplish similar auditory learning through everyday experience.

Individual differences in auditory skills and the FSR ability

A somewhat surprising result, first reported in the early 1940s, is that listeners with unusually acute spectral or temporal resolving power are no better at speech recognition under difficult listening conditions (e.g., a conversational babble background) than those whose acuity is average or even below average.[27] This finding has now been replicated numerous times, including in a recent study in which a large number of auditory discrimination tests, several speech-recognition tests (nonsense syllables, words and sentences), and a test of the ability to recognize familiar environmental sounds were administered to 338 college students with normal audiograms.28

Performance on the battery of 19 tasks was analyzed (with a principal components analysis and structural equation modeling) to determine how many discrete auditory abilities were required to account for the individual differences among the listeners. The answer was similar to that obtained in previous studies. The listeners appeared to differ in four basic abilities, three of which reflected the spectral and temporal acuity of the auditory system, while the fourth was the ability to recognize familiar sounds, both speech and non-speech.

As in earlier reports, the Familiar Sound Recognition (FSR) ability was statistically independent of (i.e., uncorrelated with) the measures of acuity. One might suspect that this was because listeners with normal hearing do not differ much in their abilities to recognize speech, but this was not the case. The range of SNR thresholds (for 50% correct word recognition in sentences) from the best 10% of the listeners to the worst 10% was about 7-8 dB, which is rather large considering that speech recognition for sentences can shift from near chance to near perfect with a change in SNR as little as 3-4 dB.

What does this have to do with the issue of speech-recognition training for users of hearing aids or cochlear implants? A great deal, because the differences among listeners in spectral-temporal acuity are more likely to reflect limitations imposed by early processing in the auditory system, while the FSR ability appears to involve higher levels of processing. A listener with excellent FSR skills is able to use a limited number of audible fragments of a complex sound to infer the portions not heard and thus successfully identify that sound. Such a person

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

is also more likely to be attending to speech through a set of "cognitive-perceptual filters" that are correctly tuned for the sounds of the language to which she or he is listening. These more central cognitive skills are the ones that show continuous improvement over lengthy auditory training, while measures of acuity using simple stimuli do not.

These conclusions are based on data collected from normal-hearing listeners. Do hearing-impaired persons have the same independence of acuity measures and the FSR ability? The answer to that question is often obscured in research reports by the range of audiograms, since difficulty in hearing the stimuli clearly affects both acuity and recognition tests. However, in those studies in which sufficient numbers of people with similar audiograms were included, speech-recognition and acuity measures seem to be essentially as independent as they are for normal-hearing listeners.[29]

If the range of FSR abilities that has been well-documented for listeners with normal audiograms also characterizes persons with hearing loss—and we know of no reason to believe that it should not— that points to an obvious conclusion: Hearing-impaired persons with poor FSR ability will have significantly worse speech-recognition performance than others with similar audiograms but better FSR abilities. Fortunately, FSR skills can be improved.

CONCLUSIONS

Several lines of evidence thus suggest that individual differences in speech recognition by listeners with similar audiograms are largely a consequence of central rather than peripheral auditory processing limitations. From the learning of speech perception by infants as well as ESL students, we are convinced that one must know the code of syllable constituents for a particular language before one can achieve success in the recognition of polysyllabic words and sentences in that language.

While we might believe that post-lingually hearing-impaired persons "know the code" perfectly well, the considerable range of speech recognition by listeners with very similar audiograms contradicts that belief.[29] It is more likely that some users have learned the new code represented by the sounds of speech as perceived through their aids, while others have not. That hypothesis is supported by the improved performance resulting from auditory training, recently reported by several investigators, including our own group.[2-6]

The lines of evidence discussed here yield at least five criteria for successful (re)training programs. Effective speech-recognition training must include at least:

(1) Right-wrong feedback delivered promptly to the listener after each response, informing her or him of the nature of the error in case of an incorrect response.

(2) A large corpus of recorded stimuli by at least 6-8 different speakers, including male and female and younger and older speakers.

(3) Training on the code of syllable constituents of the English language as produced in a variety of phonetic environments and also on the recognition of meaningful sentences, which combines bottom-up and top-down recognition skills.

(4) A method for concentrating training on the specific components of the code of syllable constituents with which an individual listener has difficulty.

(5) A curriculum that guides training in such a way that listeners can appreciate the improvement in their performance over time, and does so in such a manner that they will be willing to continue for the total length of training time that may be required for them to achieve the best level of speech recognition that can be expected, given their audiogram and the properties of their hearing aid or cochlear implant.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

REFERENCES

1. Miller JD, Watson CS, Kewley-Port D, et al. SPATS: Speech perception assessment and training system. J Acoust Soc Am 2007;122(5):3063.
2. Miller JD, Watson CS, Kistler DJ, et al. Preliminary evaluation of the speech perception assessment and training system (SPATS) with hearing-aid and cochlear-implant users. J Acoust Soc Am 2007;122 (5):3063.
3. Fu Q-J, Galvin JJ. Perceptual learning and auditory training in cochlear implant recipients. Trends Amplif 2007;11:193–205. [PubMed: 17709574]
4. Miller JD, Dalby JM, Watson CS, Burleson DF. Training experienced hearing-aid users to identify syllable-initial consonants in quiet and noise (A). J Acoust Soc Am 2004;115(5):2387.
5. Stecker GC, Bowman GA, Yund EW, et al. Perceptual training improves syllable identification in new and experienced hearing aid users. J Rehab Res Dev 2006;43:537–552.
6. Sweetow R, Palmer CV. Efficacy of individual auditory training in adults: A systematic review of the evidence. JAAA 2005;16:494–504. [PubMed: 16295236]
7. O'Grady, W. How Children Learn Language. Cambridge University Press; London: 2005.
8. Werker JF, Tees RC. Phonemic and phonetic factors in adult cross-language speech perception. J Acoust Soc Am 1984;75(6):1866–1878. [PubMed: 6747097]
9. Lively SE, Logan JS, Pisoni DB. Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. J Acoust Soc Am 1993;94:1242–1255. [PubMed: 8408964]
10. Lively SE, Pisoni DB, Yamada RA, et al. Training Japanese listeners to identify English /r/ and /l/ III: Long-term retention of new phonetic categories. J Acoust Soc Am 1994;96(4):2076–2087. [PubMed: 7963022]
11. Watson CS. Time course of auditory perceptual learning. Ann Otol Rhinol Laryngol 1980;89(5pt 2): 96–102. Suppl.
12. Watson CS. Auditory perceptual learning and the cochlear implant. Am J Otol 1991;12(Suppl):73–79. [PubMed: 2069193]
13. Espinoza-Varas B, Watson CS. Temporal discrimination for single components of nonspeech patterns. J Acoust Soc Am 1986;80(6):1685–1694. [PubMed: 3794075]
14. Bryan WL, Harter N. Studies in the physiology and psychology of the telegraphic language: The acquisition of a hierarchy of habits. Psychol Rev 1899;6:345–375.
15. Zwislocki J, Maire F, Feldman AS, Rubin A. On the effects of practice and motivation on the threshold of audibility. J Acoust Soc Am 1958;30(4):254–262.
16. Gundy RF. Auditory detection of an unspecified signal. J Acoust Soc Am 1961;33(8):1008–1012.
17. Watson CS, Franks JR, Hood DC. Detection of tones in the absence of external masking noise. J Acoust Soc Am 1972;52(2B):633–643.
18. Campbell RA, Small AM Jr. Effect of practice and feedback on frequency discrimination. J Acoust Soc Am 1963;35(10):1511–1514.
19. Loeb M, Holding DH. Backward interference by tones or noise in pitch perception as a function of practice. Perception Psychophys 1975;18:205–208.
20. Wright BA, Fitzgerald MB. Different patterns of human discrimination learning for two interaural cues to sound-source location. Proc Nat Acad Sci 2001;98:12307–12312. [PubMed: 11593048]
21. Meyer ML. Is the memory of absolute pitch capable of development by training? Psychol Rev 1899;6:514–516.
22. Tanner WP, Rivette CL. Learning in psychophysical experiments. J Acoust Soc Am 1963;35(11): 1896.
23. Hartman EB. The influence of practice and pitch-distance between tones on the absolute identification of pitch. Am J Psychol 1954;67:1–14. [PubMed: 13138765]
24. Leek MR, Watson CS. Learning to detect auditory pattern components. J Acoust Soc Am 1984;76 (4):1037–1044. [PubMed: 6501698]
25. Drennan W, Watson CS. Sources of variation in profile analysis. I. Individual differences and extended training. J Acoust Soc Am 2001;110(5):2491–2503. [PubMed: 11757938]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

26. Leek MR, Watson CS. Auditory perceptual learning of tonal patterns. Percepion Psychophys 1988;43:389–394.

27. Karlin JE. A factorial study of auditory function. Psychometrika 1942;7:251–279.

28. Kidd GR, Watson CS, Gygi B. Individual differences in auditory abilities. J Acoust Soc Am 2007;122 (1):418–435. [PubMed: 17614500]

29. Humes LE. Factors underlying the speech-recognition performance of elderly hearing-aid wearers. J Acoust Soc Am 2002;112(3):1112–1132. [PubMed: 12243159]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

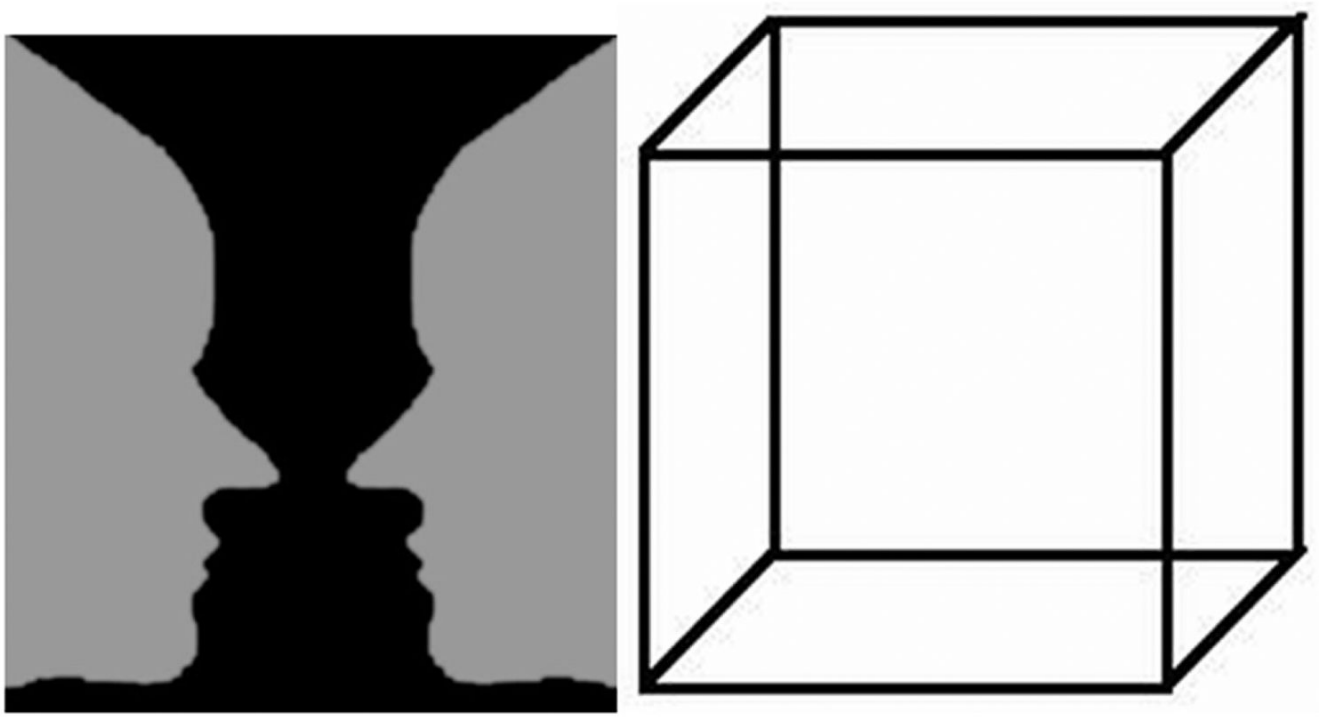

Figure 1.
Classic ambiguous figures, the Rubin vase/faces and the Necker cube are examples of the way in which a constant visual stimulus can elicit different perceptual experiences.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

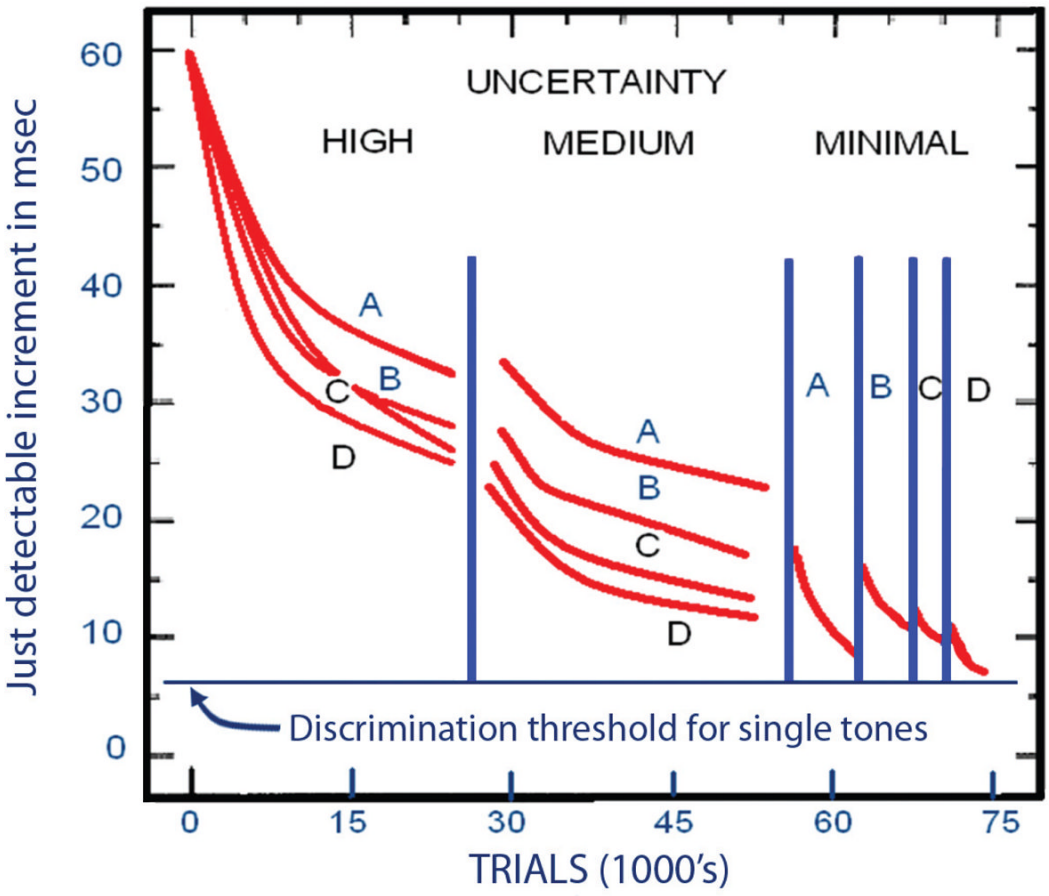

Figure 2.

Time course of auditory perceptual learning for the detection of an increment in the duration of one 50-ms component of a 500-ms tonal pattern.13 A-D represent four different patterns. The 75,000 trials required about 100 hours of training. (Figure adapted from ref. 9)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

The approximate duration of training required to approach asymptotic performance in auditory detection, discrimination, and identification tasks. (An extension of earlier reviews.[11,12])

Stimuli	Task		
	Detection	**Discrimination**	**Identification**
Single tones, other simple stimuli	1.6 h[15] <1.0 h[16] <1.0 wk[17]	4 h[18] 4 h[19] 2-7 h[20]	24-28 wks[21] 20 h[22] 4 wks[23]
Complex sounds, tonal sequences	<14 h[24] <30 h[24]	>20 h[25] >40 h[13]	>40 wks[14] >60 h[26]