Psychoacoustic Measures of Tinnitus

James A. Henry*
Mary B. Meikle*

Abstract
This report reviews research from the 1930s to the present that has extended our understanding by investigating the characteristics of tinnitus that can be studied using psychoacoustic techniques. Studies of tinnitus masking and residual inhibition began in the 1970s, leading to the therapeutic use of tinnitus masking and a consequent increase in research devoted to tinnitus measurement. In 1981, the CIBA Foundation symposium on tinnitus advocated general adoption of four tinnitus measures: (1) pitch, (2) loudness, (3) maskability, and (4) residual inhibition. Since then, psychoacoustic research into all four topics has proliferated, yielding many valuable insights and controversies concerning the details of measurement techniques. A consensus has emerged that neither the loudness nor other psychoacoustic measures of tinnitus bear a consistent relation to the severity or perceived loudness of tinnitus. Nevertheless, quantification is needed in clinical trials of proposed treatments and in a variety of other types of tinnitus research. Standardization of techniques for specifying the acoustic parameters of tinnitus thus continues to be an important research goal.

Key Words: Disability, loudness, masking, pitch, psychoacoustics, residual inhibition, severity, tinnitus

Abbreviations: ABLB = alternating binaural loudness balance, DL = difference limen, FCDS = forced-choice double staircase, MML = minimum masking level, PLU = personal loudness unit, RI = residual inhibition

The topic of tinnitus has gained increasing attention throughout this century. This is fortunate for the millions of Americans and the untold millions worldwide who suffer from this invisible auditory disorder. As we approach the turn of the century, tinnitus has attained a level of visibility high enough to attract considerable research attention and augment efforts to provide help for tinnitus sufferers. There are now viable treatment options that were not available 25 years ago, some with reportedly high rates of success. A national organization, the American Tinnitus Association, is dedicated to supporting the tinnitus sufferer in a variety of worthwhile ways. Recently, there has also been increased support for tinnitus research on the part of the U.S. Department of Health and Human Services.

In 1982, the National Academy of Sciences published the report “Tinnitus—Facts, Theories and Treatments,” prepared by the Committee on Hearing, Bioacoustics and biomechanics of the National Research Council and edited by Dennis McFadden (McFadden, 1982). In this far-reaching report, the Committee emphasized the need for basic research into psychoacoustic measures of tinnitus, saying “For theoretical and practical reasons, it is important to obtain characterizations of tinnitus—spectral location, degree of complexity, magnitude, etc.—that are as accurate as possible.” They identified several lines of related research that were “necessary as a basis for establishing a standardized test procedure.” Those topics included comparability of the various existing measurement procedures, test—retest reliability of the data obtained, and the extent to which different measurement procedures could accurately reflect important differences between patients.

Unfortunately, nearly 18 years later, it is still the case that standardized test procedures have not been universally adopted. The available information concerning test—retest reliability is still relatively sparse, and efforts to determine
whether different types of patients respond differently in psychoacoustic tests are also very few. There is some hope, however, that the situation may be improving. The number of tinnitus-related research reports continues to increase year by year, and there is a growing number of researchers whose work (whether clinical or basic) depends on tinnitus identification and characterization. Many of the current tinnitus evaluation techniques, such as sophisticated brain imaging, testing of auditory brainstem responses, measurement of otoacoustic emissions, and so forth, can benefit from more detailed characterization of subjects' tinnitus, as would clinical efforts to identify effective treatment modalities (tinnitus masking as well as “habituation” or “retraining” therapies, electrical stimulation, drug treatments, etc.).

It is therefore the objective of this paper to summarize work to date on the psychoacoustic aspects of tinnitus and to review the available information relating psychoacoustic measures to clinical concerns. Clinics that attempt to provide treatment for tinnitus patients must be able to categorize patients according to their degree of need for treatment, and likewise must be able to compare the post-treatment to the pretreatment status of tinnitus. It is clear that both of these capabilities are heavily dependent on carefully controlled measurement techniques that are feasible to use with a wide range of clinic patients.

EARLY ATTEMPTS AT MEASUREMENT

From the outset, it has been clear that the spectral characteristics of tinnitus are of fundamental importance. As early as 1821, it was recognized that high-pitched versus low-pitched tinnitus was treatable with different types of masking sounds (J.M.G. Itard, quoted in Stephens, 1987). With the advent of electroacoustic equipment, it became technically feasible to establish pitch and loudness characteristics precisely (Jones and Knudsen, 1928), although no description of those early methods exists. The earliest information on psychoacoustic methods for describing tinnitus appears to have been supplied by Wegel (1931) and Josephson (1931). Both investigators used pure tones to measure the loudness and pitch of tinnitus, and Wegel also plotted masking curves by presenting tonal maskers at a number of different frequencies and adjusting their levels to "cover" or mask the tinnitus. These were all ground-breaking studies that laid the foundation for more detailed work to follow.

The name of Edmund Prince Fowler deserves special mention for many of his contributions to tinnitus measurement. He was the first to describe a method for measuring loudness recruitment by balancing the loudness of sounds between ears, which he termed the Alternating Binaural Loudness Balance (ABLB) (Fowler, 1936, 1937). He later applied the loudness balance technique to measure the "effect of tinnitus on hearing acuity" (Fowler, 1938). With this test, the loudness of tinnitus in one ear was balanced with the loudness of a tone in the contralateral ear. The level of the comparison tone, expressed in dB SL, supposedly indicated the tinnitus loudness as experienced by the patient.

Fowler (1940) stated that it was important to duplicate or match the loudness and pitch of tinnitus, using contralateral tones. For pitch matching, he stressed the importance of presenting stimuli at levels equal to the tinnitus intensity.

Fowler (1942, 1943) noted that although patients described their tinnitus as very loud, he found it could usually be matched at only 5 to 10 dB SL. In his opinion, tinnitus, being an "illusion," tends to be exaggerated. The typical loudness matches of 5 to 10 dB SL were "very faint" and did not correspond to the reported severity. Fowler recommended educating patients regarding this discrepancy to understand that their statements about severity did not correspond to "the facts." It is interesting that Fowler, inventor of the loudness balance method for evaluating loudness recruitment, did not apply the concept of recruitment to his measurements in tinnitus patients.

Goodhill (1952) noted that patients have difficulty describing their tinnitus. He believed that the validity of a description depended largely upon a patient's familiarity with musical terms, commenting that only patients who were "scientifically or musically articulate" could match their tinnitus to tones from an audiometer. He therefore developed a "tinnitus identification recording," consisting of 27 different sounds designed to mimic the acoustic sensations described by patients. Patients listened to the recording and selected the sound(s) that most closely resembled their tinnitus.

Reed (1960) conducted the first large-scale study of tinnitus measurement, using 200 subjects. He acknowledged the work of Fowler and others who had developed tinnitus measurement techniques, but noted that their methods had not been described in sufficient detail to permit replication. He therefore described in detail instrumentation and methodology for
matching the “frequency, content, and loudness” of tinnitus. The equipment was capable of presenting a specified bandwidth of noise centered around a specific frequency. The noiseband could be moved up or down the frequency scale by changing the central frequency. If the patient had unilateral tinnitus, stimuli were presented contralaterally. With bilateral tinnitus, stimuli were presented ipsilaterally. Stimulus trials were repeated three times. Using this method, 76 percent of the patients selected tones at > 3000 Hz as a pitch match. The tinnitus loudness was then determined using the loudness balance method of Fowler.

Reed also rated each patient's tinnitus severity as "slight," "moderate," or "severe" according to the impact of tinnitus on the patient's life. Tinnitus severity did not appear to be correlated with any of the psychoacoustic tinnitus measurements: central frequency, bandwidth, or loudness.

Graham and Newby (1962) set out to determine whether there were different tinnitus characteristics in four groups of 25 subjects each: one group with normal ears who experienced tinnitus in a quiet room and three groups with hearing loss (sensorineural, conductive, or mixed). For pitch matching, a 10 dB SL tone was presented at each of 11 test frequencies. To measure tinnitus loudness, they used monaural loudness balancing. Subjects were asked to judge the loudness of their tinnitus and the loudness of a 1-kHz, 20 dB SL tone, using a subjective magnitude scale.

Some noteworthy conclusions of Graham and Newby's study were (1) subjective descriptions of tinnitus were not related to any of the psychoacoustic measurements and (2) tinnitus pitch matches in the conductive hearing loss group (tinnitus usually described as "roaring") were significantly lower in frequency than for the other two hearing loss groups.

ADVENT OF TINNITUS MASKING STUDIES

In 1971, Feldmann initiated the first systematic studies of tinnitus masking. He noted that conventional tone-on-tone masking resulted in a characteristic masking pattern: when masker intensity is plotted as a function of frequency, the intensity of the masking tone is lowest when its frequency is close to that of the tone being masked (Wegel and Lane, 1924). Feldmann hoped that a similar situation would prevail in the case of tinnitus. If so, this method could supply an alternative method for determining the pitch of tinnitus. Feldmann developed a protocol that succeeded in establishing masking patterns for 200 tinnitus patients. He observed five different types of masking patterns, none of which resembled conventional masking curves. When masking levels were referenced to the subject's hearing thresholds at the various frequencies, these patterns could be described as follows: (1) "congruence," in which masking occurred at low levels at all frequencies; (2) "distance," with masking at high levels across frequencies; (3) "persistence," which indicated that tinnitus could not be masked; (4) "convergence," when masking occurred at high levels at low frequencies, converging to low levels at high frequencies—generally close to the tinnitus frequency; and (5) "divergence," the opposite pattern of "convergence" and generally seen with low-frequency, Meniere's-type tinnitus.

Feldmann found that tinnitus could not be masked by any sound in 11 percent of his subjects and that for 32 percent any weak sound could mask it effectively. He also demonstrated that tinnitus could be inhibited following termination of a stimulus, an effect that was later termed "residual inhibition" (Vernon, 1977). He further demonstrated contralateral inhibition of tinnitus. Feldmann concluded that the neural activity responsible for tinnitus differed from the neural activity resulting from stimulation with external sound. This report thus contains the first statement that the perception of tinnitus does not seem to follow the same psychoacoustic principles that apply to the perception of external sounds.

A QUARTER CENTURY OF RESEARCH INTO TINNITUS MEASUREMENT

Following the pioneering work done in the early years, the stage was set for a rapid growth of interest and work in the area of tinnitus. Investigations into psychoacoustic measurement of tinnitus increased substantially from the 1970s on, with increasing emphasis on accurate and reliable methods for quantifying tinnitus sensations. Spurred by the development of tinnitus masking as a therapeutic procedure, which brought many hundreds of patients in to try the new masking therapy, a number of studies focused on techniques for specifying the pitch and loudness of tinnitus. Test–retest reliability became an important issue, and a number of investigators sought to
understand the unique ways in which tinnitus responded to masking.

Early in 1981, the CIBA Foundation in London sponsored one of its prestigious scientific symposia in order to promote international cooperation in tinnitus research. Participants developed general guidelines for the clinical evaluation of tinnitus, which they hoped would be used widely. The recommended measures covered pitch matching, loudness matching, the maskability of tinnitus, and residual inhibition (Evered and Lawrenson, 1981). These recommendations were based on the techniques pioneered by Fowler, Reed, Graham, Feldmann, and others. A set of procedural details for these tests was offered by Vernon and Meikle (1981). A central concept in the CIBA symposium was that standardization of tinnitus measures would advance international understanding and facilitate work on tinnitus.

Studies of Tinnitus Pitch

Measuring the pitch of tinnitus sounds straightforward enough in theory, but in practice there are several factors that complicate matters. First, about 46 percent of patients have complex tinnitus consisting of more than one type of sound (Meikle et al, 1995). Although they may be asked to match only the predominant one or two tinnitus sounds, it may not always be possible for them to ignore the competing effects of additional tinnitus sounds. Second, a substantial number of tinnitus patients (36%) report that their tinnitus exhibits pitch or other variations of sound quality—in some cases, as often as several times a day (Meikle and Griest, 1991). In studies designed to evaluate the reliability of particular pitch matching methods, it is important to select subjects with constant, unvarying tinnitus to avoid excessive response variability due to fluctuations in the tinnitus. Third, as noted by several different investigators, tinnitus that is not very loud tends to be quite labile and easily masked by many sounds, including audiometric test tones and tones presented for comparison during pitch matching (unpublished observations of present authors). As a result, individuals with mild tinnitus may not be able to perform pitch or loudness matching with a high degree of reliability.

As we all know, pitch is the psychological percept corresponding to the frequency of acoustic stimuli. It is awkward, however, to keep repeating the term “the frequency matching the pitch of the tinnitus.” For convenience, in what follows, we will use the shorter terms “the frequency matching the tinnitus” or even “the tinnitus frequency” when discussing work on tinnitus pitch and loudness matching.

General Comments about Pitch Matching

Although the pitch of tinnitus may seem to be one of the easiest attributes of tinnitus to quantify, in fact, there are some traps for the unwary. Inexperienced subjects may tend to confuse pitch and loudness, causing them to reject a tentative frequency match for the tinnitus on the basis of loudness differences or to accept an inappropriate match because its loudness is equivalent to the tinnitus (Vernon and Meikle, 1981). It is important, therefore, during pitch matching to adjust the loudness of comparison stimuli very carefully to make them equal to the loudness of the tinnitus (Fowler, 1940; Vernon & Meikle, 1981).

Another potential source of error is the possibility of “octave confusion,” in which the subject may at first select tones an octave above or below the frequency that they later identify as the “best” tinnitus pitch match. Octave confusions occur quite commonly even for normal-hearing people making comparisons between external tones. The fact that pitch matches for tinnitus might be affected by octave confusion came to the attention of clinicians working with tinnitus patients in the 1960s (Graham and Newby, 1962) and was emphasized in the 1980s by Vernon et al (1980a). Subsequently, the tinnitus testing procedures recommended in 1981 by the CIBA symposium on tinnitus included a test for octave confusion as an integral part of the test battery (Evered and Lawrenson, 1981).

The CIBA recommendations advocated use of the contralateral ear for pitch matching if the tinnitus is unilateral. The idea is that it is less confusing to the patient if the comparison stimuli are presented to an ear that is free from distracting sound sensations (Vernon et al, 1980a). In practice, the majority of patients are found to have bilateral tinnitus (Meikle and Walsh, 1984; Meikle and Griest, 1991), and, in most cases, it appears to be feasible to match the pitch of the tinnitus in each ear separately (if they differ) by presenting the comparison tones ipsilaterally (Tyler and Conrad-Armes, 1983b; Vernon and Fenwick, 1984).

Another testing consideration is the desirability of adjusting the loudness of comparison tones by starting below the subject's threshold and increasing the intensity in an ascending
series of small steps (e.g., 1–2 dB) (Vernon, 1982). An ascending series is used to minimize the potential for masking or residual inhibition caused by the test tones (and such effects could conceivably occur at loud sound levels even when the comparison tones are presented contralaterally).

A number of additional methodological details have been evaluated experimentally, as will be discussed below. First, however, we will summarize the body of evidence that has accumulated regarding the range of tinnitus pitches observed in clinical work with tinnitus patients.

**Pitch Matching Data**

Most workers agree that the pitch of tinnitus is typically related in some way to the individual's audiometric configuration. There is, however, some question concerning the exact nature of the relationship. Wegel (1931), Josephson (1931), Mortimer et al (1940), and Fowler (1942) all believed that the pitch of tinnitus corresponds to a frequency region in which the audiogram shows either a point of maximum hearing loss or else the transition from normal to abnormal hearing. In a recent discussion of tinnitus mechanisms, pitch matching data from a sample of over 1000 tinnitus patients were summarized to illustrate the point that there is great variability in the precise location of tinnitus frequencies (cf. Figs. 14–3 and 14–5 in Meikle, 1995). This was also seen earlier by Graham and Newby (1962). It is clear that pitch matches for tinnitus can occur practically anywhere in frequency regions where there is hearing loss.

In general, the pitch of tinnitus tends to be high. The largest available surveys of tinnitus pitch matches show that in the majority of patients, the pitch is at or above 3000 Hz (Reed, 1960; Meikle and Walsh, 1984; Meikle et al, 1995). High-frequency hearing loss is the most common audiometric configuration encountered in the tinnitus clinic; thus, it is not surprising that surveys have found relatively low percentages of patients with tinnitus pitches below 1000 Hz (Meikle and Walsh, 1984; Meikle, 1985). The recent summary of audiometric data from the 1000 tinnitus patients referred to above illustrated a striking and very orderly inverse relationship between the pitch of tinnitus and the severity of hearing loss (Meikle, 1995). After dividing the group into five successively higher pitch categories, the mean audiometric curves for the five groups were compared. Statistical analysis showed that patients with the lowest tinnitus pitch matches tended to have the greatest hearing loss, both in terms of the extent of loss and the frequency range affected. Conversely, those with the highest pitch tinnitus tended to have the least hearing loss. These differences were highly significant (p < .0001). At present, however, there is no explanation for the fact that severe tinnitus occurs in a small number of patients who appear to have perfectly normal hearing (Reed, 1960; Graham and Newby, 1962; Cahani et al, 1983). Thus, although it is clear that the pitch of tinnitus tends to correspond to frequency regions of significant hearing loss, there are important exceptions to that rule.

The pitch of tinnitus has been investigated in relation to causal factors for tinnitus, although the situation is complicated in many patients by the presence of more than one causal factor (e.g., patients with noise-induced hearing loss in addition to Meniere's disease). Nodar and Graham (1965) compared pitch matches for subjects with Meniere's disease, conductive hearing loss, and sensorineural hearing loss. All Meniere's subjects had pitch matches below 1000 Hz (median 320 Hz), which agreed with the findings of Walsh (1956), Caprosa (1963), and Day (1963). The conductive hearing loss group had a median pitch match of 490 Hz (range = 90–1450 Hz), which differed significantly from the sensorineural group (median = 3900 Hz, range = 545–7500 Hz), in agreement with the findings of Graham and Newby (1962).

Douek and Reid (1968) attempted to determine if a clinical audiometer could be used to obtain pitch matches that had diagnostic value. They measured tinnitus pitch in 200 patients using Josephson's (1931) method. The patients were divided into eight groups according to clinical diagnosis of auditory pathology, and their pitch matches showed general correlations with the various auditory pathologies. The authors noted that thresholds seemed elevated in the region of the tinnitus frequency, which they termed an "audiometric artifact."

Using a pitch matching procedure described by Graham and Newby (1962), Cahani et al (1983) determined the tinnitus frequencies in 56 noise-exposed subjects with and without hearing loss. The 29 normal-hearing subjects were found to have significantly higher frequency tinnitus than the 27 hearing loss subjects. Because of the different distributions of pitch matches, the authors suggested the possibility of different neural processes being responsible for tinnitus generation for noise-exposed individuals with normal hearing compared to those with hearing loss.
Several interesting attempts have been made in the past to evaluate the pitch of tinnitus in subjects exposed experimentally to noise levels sufficient to produce temporary threshold shift (Atherley et al., 1968; Loeb and Smith, 1967). Most, if not all, subjects experienced tinnitus. The studies disagreed as to the audiometric frequency at which maximum threshold shifts occurred, and it is not certain that the measurement methods used then were capable of yielding reliable pitch matches in the face of thresholds that were varying over time during the postexposure recovery period. For ethical reasons, such studies are difficult or impossible to replicate.

**Methodology of Tinnitus Pitch Matching**

Tyler and Conrad-Armes (1983b) evaluated tinnitus pitch using three psychophysical methods (Adjustment, Limits, and Adaptive) for each of 10 subjects. Comparing the tinnitus frequencies obtained using the three methods, no significant differences were found in either the group means or the standard deviations. For clinical use, the authors recommended either their Adaptive or Adjustment Method, which required only 1 to 2 minutes to obtain a single pitch match. They suggested presenting matching tones in the ipsilateral ear to avoid problems with binaural diplacusis (differences in pitch perception between ears). They also recommended seven to nine pitch matches from each subject because of the large variability in patients' abilities to provide repeated matches. Thus, for unreliable patients, increasing the number of pitch matches tended to improve the accuracy of the final match.

Vernon and Fenwick (1984) emphasized that tinnitus patients easily confuse loudness and pitch; thus, it is necessary to present matching tones at levels previously matched to the tinnitus loudness to minimize this confusion. A two-alternative forced-choice procedure, using discrete pairs of frequencies separated by 1000 Hz, was advocated for pitch matching. Matching tones could be presented either contralaterally or ipsilaterally to the tinnitus, determined by patient preference, but the authors concluded that the final pitch match should be made in the ipsilateral ear because of the potential for binaural diplacusis.

Penner and Bilger (1992) investigated the use of a forced-choice double-staircase (FCDS) procedure employing 100-Hz step sizes for obtaining pitch matches. Subsequently, the technique was modified (Penner and Klafter, 1992) because the step sizes were considerably larger than the frequency difference limen (DL) for normal-hearing subjects. Penner and Klafter therefore used the FCDS procedure with 0.2 percent step sizes for frequency and found that the DL for the tinnitus matches was comparable to that for external tones. The authors surmised that this good correspondence could have been due to the FCDS procedure's ability to accurately measure tinnitus that is stable, or, alternatively, that subjects may have matched to an "imagined tone." They evaluated this latter notion with a second experiment and found that about half of the subjects had larger DLs for matches to imagined tones than to external tones, whereas the other half did not. This result was difficult to interpret and may have indicated that half of the subjects were matching to imagined tones and not to their tinnitus.

An unusual technique was proposed by Ohsaki et al. (1990). They noted that a standard audiometer could not obtain an accurate tinnitus pitch match, nor could it measure small variations in tinnitus pitch. They therefore developed a novel testing device called a "heptachord generator." This device was capable of presenting tones according to the standard musical scale (do-re-mi). Two repeated matches at the same frequency were required to identify a pitch match. They found that matches repeated within 1 week were no further apart than the adjacent test frequency for 95 percent (18) of the 19 ears tested. Pitch matches made in the morning and afternoon of one day were repeatable within one frequency for 100 percent of 10 ears. The authors concluded that their heptatonic method to obtain tinnitus pitch matches was more accurate than using a conventional audiometer. However, the pitch of tinnitus in many patients is above 4000 Hz, the frequency at which musical tonality begins to break down. Thus, the use of musical intervals would seem to be difficult to apply in testing the majority of tinnitus patients.

**Tinnitus Loudness Matching**

**Loudness Matching at the Tinnitus Frequency versus at a Normal-Hearing Frequency**

Vernon (1976) noted the previous findings of Reed (1960), Fowler (1942), and Graham and Newby (1962) showing that tinnitus was typically matched to tones only a few decibels above threshold. The question was how tinnitus of "such feeble intensities" could be so disturbing. Vernon pointed out the obvious possibility...
that tinnitus patients might be experiencing abnormally rapid loudness growth (recruitment) that could account for sounds being perceived as much louder than their low sensation levels would indicate. He proposed a series of experiments to test that hypothesis, including one in which tinnitus loudness matching measurements should be made at both the tinnitus frequency and at a nontinnitus frequency. Because of the increased possibility of recruitment in the frequency region of the tinnitus, he stated that “it is essential that the portion of the ear involved in the tinnitus be compared to the portion that is not” (p. 18).

Goodwin and Johnson (1980b) were the first of many investigators to evaluate the effect of recruitment by matching tinnitus loudness at both the tinnitus frequency and at a normal-hearing frequency. For their subjects, loudness matches for tinnitus were obtained at higher sensation levels when measured at frequencies where hearing was normal (mean = 24 dB SL) than when measured at the tinnitus frequency (mean = 7 dB SL). These results suggested that tinnitus loudness is significantly underestimated when measured at the tinnitus frequency. The authors concluded that loudness recruitment at the tinnitus frequency was probably responsible for the differences they observed, and therefore that measurement of loudness matches at the tinnitus frequency was not appropriate.

Tyler and Conrad-Arms (1983a) also conducted a study to compare loudness matches obtained at the tinnitus frequency to those obtained at a normal-hearing frequency; results of their study were similar to those of Goodwin and Johnson (1980b). Tyler and Conrad-Arms, however, concluded that the loudness matches were still too small at the normal-hearing frequency relative to the patients’ subjective tinnitus complaints. They proposed a method for transforming the loudness matches from dB SL to sones, as discussed in the next section below.

Jakes et al (1985) were interested in studying the relationship between tinnitus loudness and different dimensions of complaints due to the tinnitus. Their 1-kHz loudness matches were unrelated to most aspects of tinnitus complaints, except slightly (r = .34) with the “intrusiveness of tinnitus.”

Penner (1986a) used a magnitude estimation procedure to measure slopes of loudness growth functions for each subject at the tinnitus frequency and at 1 kHz, where hearing was normal. For most of the subjects, the mean slope at the tinnitus frequency was steeper than the mean slope at 1 kHz, indicating that growth of loudness for an external tone was more rapid at the tinnitus frequency than at 1 kHz. Citing Robinson and Dadson’s (1956) equal-loudness contours in normal-hearing subjects, Penner noted that this “recruitment” at higher frequencies was not representative of a pathologic condition. She concluded that, for most patients, tinnitus loudness is not “paradoxical” in that it mimics the loudness of external high-frequency tones.

Risey et al (1989) emphasized the importance of a loudness match procedure from which results would correlate with self-reported loudness and severity scales. They obtained loudness matches at the tinnitus frequency and at 1 kHz. Tinnitus severity ratings were also obtained using a 1 to 10 scale. Compared to loudness matches at the tinnitus frequency, the 1-kHz matches were significantly larger, in dB SL, and showed better test–retest reliability. The 1-kHz matches also showed a small but significant correlation with the tinnitus severity ratings (r = .224; p < .05), whereas matches at the tinnitus frequency did not (r = .055; p > .05).

In a second project, Risey et al conducted the same tinnitus evaluations in subjects who were administered furosemide intravenously. For the subjects whose tinnitus was reported to decrease, their 1-kHz loudness matches were reduced an average of 5.75 dB SL, and their severity ratings dropped an average of 1.96 points. These changes were significantly correlated (r = .78; p < .05). The remaining subjects experienced no change in their tinnitus and also no changes in either their loudness matches or their rating scales. The authors concluded that loudness matches obtained at 1 kHz are sensitive to changes in loudness induced by pharmacologic agents and correlate well with self-reported ratings of tinnitus severity.

**Conversion of Loudness Matches to Different Units of Measurement**

As described above, many investigations have consistently demonstrated that tinnitus loudness matches tend to be larger in frequency regions where hearing sensitivity is normal, supporting the premise that recruitment might be responsible for the small size of loudness matches obtained at the tinnitus frequency. Nevertheless, Tyler and Conrad-Arms (1983a) noted that loudness matches at the “normal” frequencies were still considered small relative to patients’ complaints. Thus, tinnitus loudness expressed in dB SL may not be meaningful even
when measured at a frequency where hearing is normal. They therefore devised a method to transform the loudness matches from dB SL to sones, using the psychophysical equations developed by Stevens (1955) and adjusting for the slopes of loudness growth functions for persons with cochlear hearing impairment. In some subjects, conversion to sones seemed to be in agreement with the perceived loudness of tinnitus, whereas in other subjects there was poor agreement. Although results were inconsistent, the authors suggested that sones might be the more appropriate way to represent tinnitus loudness. They acknowledged that the loudness growth formulas might be too general for application to specific individuals, highlighting the need to establish individualized functions at each loudness matching frequency.

Penner (1984) attempted to evaluate subjects' tinnitus magnitude by determining equal loudness measurements across frequencies and obtained results that were comparable to those expected for normal listeners. These data suggested that tinnitus loudness matches expressed in dB SL are an inverse function of the absolute threshold—the higher the threshold, the smaller the loudness match. Penner concluded that tinnitus loudness might better be expressed in sones, as suggested by Tyler and Conrad-Armes (1983a), and that tinnitus matched to a small value in dB SL might reasonably be expected to seem “loud” to the patient.

Hinchcliffe and Chambers (1983) pointed out that tones calculated to be 1 sone in loudness may be perceived at different loudness levels by different individuals and that a sound perceived as the same loudness may be acceptable to one individual and not to another. They therefore attempted to construct individual psychophysical functions that would take into account individual differences in loudness growth and in “loudness acceptability.” Their loudness functions were based on a mathematical function described by Scharf and Stevens (1959) to describe the growth of loudness near threshold.

Hinchcliffe and Chambers used a loudness function at 1 kHz as the reference function, but instead of using 40 dB SL as unity (the 1 sone standard advocated by Stevens), they used the “most comfortable loudness” level. They termed this level one “personal loudness unit” (PLU) and used it as the comparison standard (“unity”) for subjective magnitude estimation. Subjects adjusted the level of an external sound to be one-half unity (0.5 PLU) and twice unity (2 PLU). In this way, they constructed individual loudness functions for each subject. Although they suggested that this method might be usable for clinical tinnitus studies, their method has not been adopted, perhaps because it is time consuming and may be difficult for some patients.

Hallam et al (1985) applied the method of Hinchcliffe and Chambers (1983) to calculate PLUs to represent the tinnitus loudness for their subjects. They determined PLUs at the tinnitus frequency and at 1 kHz, and also expressed the measurements in dB HL, dB SL, and sones. Like others, they found that measures expressed in dB HL or dB SL were poorly correlated with psychological scales of tinnitus severity. Unlike Tyler and Conrad-Armes (1983a), transformation to sones did not improve the correlation. Conversion of the measures to PLUs, however, increased the correlations (to r's of .30-.42) between the PLUs and specific items from their psychological scales. It should be noted, however, that these improvements were only obtained when half of the subjects were removed from the analysis due to their difficulties performing the tasks.

Jakes et al (1986) noted that although considerable effort had been expended to refine tinnitus loudness match measures, little work had gone into relating these measures to self-reported loudness. Those few studies showed small correlations, even when significant (Hazell, 1981, Tyler and Conrad-Armes, 1983a; Hallam et al, 1985). Without such a correlation, the authors stated, a loudness match method is invalid for representing loudness. They obtained loudness matches at the tinnitus frequency and at 1 kHz and expressed the loudness matches in dB HL, dB SL, and PLUs, as described by Hinchcliffe and Chambers (1983). Patients also indicated their tinnitus loudness using five different self-report scales. Only mild to moderate correlations were found between any of the loudness match metrics and the various loudness scales. The PLU metric resulted in the best correlation (largest r = .44; p < .01). Not surprisingly, all of the correlations were “markedly improved” when subjects who had difficulties with the test procedures were excluded (largest r = .87; p < .001).

Matsuhira et al (1992) proposed a method to correct tinnitus loudness matches that are underestimated due to the influence of loudness recruitment, using only data that would normally be obtained during a standard clinical audiologic evaluation. Their objective was to provide clinicians with a method that would more accurately indicate the perceived loudness of a patient's tinnitus without requiring additional tests. They
developed an “averaged loudness function,” based on loudness growth functions previously derived by other investigators for the typical cochlear-impaired ear. To correct the size of patients’ tinnitus loudness matches for abnormally rapid loudness growth, they modified the averaged loudness function for each patient using data obtained from testing that individual. The results, however, were highly variable, as would be expected because of the great variability that exists in regard to abnormalities of loudness perception. From their work, it appears difficult or impossible to determine an “average loudness abnormality” that could be generalized widely.

Henry (1996) used the ABLB technique to evaluate loudness growth in unilateral tinnitus patients with normal or near-normal hearing in the contralateral ear. Tinnitus loudness matches, as well as growth of loudness for external sounds using ABLB, were obtained with a series of tones. Analysis of group results showed that variance in the rate of loudness growth (i.e., variance in regard to recruitment) accounted for only about 25 percent of the variance in the size of the loudness matches, leaving about 75 percent of the variance unaccounted for. These observations suggest that additional factors, as yet unknown, influence the small size of tinnitus loudness matches.

**Tinnitus Loudness Matching Curves**

In a manner similar to that used by Feldmann for obtaining tinnitus masking curves (1971), obtaining tinnitus loudness matches at a series of test frequencies yields several different types of loudness matching curves (Vernon et al, 1980b; Vernon and Meikle, 1981). Recalling his previous statements, Vernon (1976) noted that recruitment is the probable cause of the paradoxically small tinnitus loudness matches. He pointed out that the size of the loudness match, in dB SL, tends to decrease as a function of frequency, up to the tinnitus frequency. He therefore suggested that when loudness matches greater than 10 dB SL are observed, it is likely that the loudness was matched incorrectly (i.e., to a frequency below the actual tinnitus frequency).

Vernon and Meikle (1981) observed that the loudness matches for tinnitus clinic patients fell into two distinct groups when loudness matches in dB SL were plotted as a function of test frequency. In one group, the loudness matches were large at lower frequencies and became progressively smaller as the frequency of the test tone increased and approached the tinnitus frequency. In the other group, the loudness matches were independent of frequency (and most commonly were only a few dB above threshold). There appeared to be no other distinguishing characteristics for the two groups. The authors suggested that the shape of the loudness matching curve might predict an individual’s tinnitus maskability curve (recall the analogous masking curves observed by Feldmann, 1971). To date, no one has tested that interesting possibility.

Most recently, Meikle et al (1996) collected data from four separate loudness matching studies in order to review tinnitus loudness balance data from 121 patients. All four studies showed three major patterns of loudness matching curves accounting for 80 percent of the subjects: one pattern was termed “merging,” one “parallel,” and one “transitional.” They used linear regression to calculate a best-fit line for each subject’s loudness curve. The slopes of the best-fit lines were generally negative, forming a continuum of values ranging from about 0.6 to 0 in 80 percent of the subjects (20% exhibited loudness matching curves that were too complex to fit to a single line). It is an interesting conjecture that the loudness matching patterns might be relatable to Feldmann’s masking curves. In any case, the various types of loudness curves may ultimately prove useful for categorizing clinically relevant dimensions of tinnitus.

**Reliability of Tinnitus**

**Pitch and Loudness Matches**

Most investigators agree that pitch matches for tinnitus are not as reliable as loudness matches and that there are large individual variations in regard to the ability of patients to match their tinnitus pitch. Excellent pitch matching capabilities are typical for people with musical training or who work in acoustics. At the other extreme are patients with significant hearing loss who may have reduced frequency resolving ability. In our own clinical experience, we have encountered patients for whom all frequencies above a certain value (e.g., 3000 Hz) sound similar. Clearly, such individuals cannot make reliable frequency judgments about their tinnitus pitch.

Additional sources of unreliability exist in patients with multiple tinnitus sounds or whose tinnitus varies over time in regard to pitch, timbre, or loudness. Variations of the sound quality of tinnitus occur in approximately one-third of the tinnitus patient population (Meikle et al,
Such individuals have more difficulty monitoring how their predominant tinnitus pitch relates to the external comparison tones. Although tinnitus that exhibits spectral variations is a clinical reality that cannot be ignored, individuals with variable tinnitus are not appropriate subjects for research evaluating the reliability of various pitch matching protocols.

Early in his work with tinnitus patients, it was stated by Vernon that tinnitus loudness matches were "inordinately reliable," being reproducible in most cases to within 1 dB when testing was repeated on the same day (Vernon et al, 1980b). Similar results were reported by Bailey (1979). In later work, Vernon studied test-retest reliability in nine tinnitus patients selected as subjects because they had constant, severe tinnitus that they described as tonal and free from loudness fluctuations (Vernon, 1985). They were tested four times on each of 4 days separated by at least 1 week. Using the ear contralateral to the predominant tinnitus, he found that the subjects' loudness matches varied by no more than 2 dB when measures were repeated on the same day. (Intervals of 10-15 minutes were interpolated between trials.) When loudness matches were repeated after a week or more, the variability was greater but, in most cases, the differences were no larger than 5 dB, although a few showed differences up to 10 dB.

Vernon has emphasized that consistent measurement of tinnitus loudness matches depends on (1) use of ascending stimulus intensities to avoid producing residual inhibition; (2) increasing the comparison tones in small increments (2 dB to determine threshold, after that 1 dB to establish the loudness match); and (3) determining the threshold each time the loudness match is measured; "otherwise the normal variability of threshold measures will introduce spurious differences between the threshold and the loudness match."

Penner (1983) and Burns (1984) commented that the high reliability of tinnitus loudness matches reported by Vernon et al (1980b) and by Bailey (1979) was greater than might be expected even for normal-hearing subjects. They also commented that response averaging techniques used by some investigators (Reed, 1960; Graham and Newby, 1962; Goodwin and Johnson, 1980a, b), in which tinnitus pitch and loudness matches were required to fall within a specified range a criterion number of times, did not take into account the full distribution of responses outside of the specified range, leading to a false impression of reliability. Burns cited studies

Penner et al, 1981; Penner, 1983) indicating pitch matching variability "several orders of magnitude higher" than that obtained from normal listeners matching to external tones. He mentioned Penner (1983) as the only study providing actual data on loudness match variability, and that her data showed it to be large. Burns also found "extremely large" variability for the tinnitus measurements, especially pitch matches, relative to external stimuli.

Penner (1986b) suggested that the reliability of tinnitus measurements was poor because tinnitus acted as a fluctuating internal noise. She tested that idea by investigating the slope of the psychometric function (probability of correct signal detection as a function of signal level) in three frequency regions: below, in, and above the tinnitus frequency. The slopes were flattest in the tinnitus region, supporting her theory that tinnitus is a source of unstable internal noise, which would cause variability in the measurements.

Smith et al (1991) obtained repeated measures of tinnitus pitch and loudness matches and minimum masking levels (MMLs). Pearson correlation coefficients between the first and second replications were at least 0.95 for the loudness matches and MMLs. For the three repeated pitch matches, however, the correlations were poor between the first and the second matches (r = .49), as well as the first and third replications (r = .26), but improved between the second and third (r = .80). The authors noted that subjects found pitch matching to be a difficult task and that the data indicated a practice effect.

Penner and Bilger (1992) noted that (1) previous procedures used to obtain tinnitus pitch matches generally resulted in unreliable matches; (2) patients often report that a single pure tone does not simulate tinnitus; thus, pitch matching with pure tones may be inappropriate; and (3) because matching tones in an adjustment method are presented monotonically, bias due to sequential effects may occur (Woodworth and Schlosberg, 1954). The authors investigated the use of the adaptive FCDS procedure (Jesteadt, 1980) as a method to force patients to choose each stimulus independently and to allow the patient to classify tones with respect to their tinnitus (e.g., determining whether a tone was higher or lower in pitch than the tinnitus).

The FCDS and a method of adjustment were used to obtain repeated within-session tinnitus pitch and loudness measures. The standard deviations of pitch matches obtained with the FCDS procedure were comparable to those obtained with external tones (chosen to mimic...
tinnitus), but the adjustment method showed relatively poor reliability. For loudness measures, the FCDS and adjustment methods both provided reliable responses comparable to use of an external tone. For a subset of subjects, all measures were repeated over 20 sessions, and all methods provided approximately equivalent results. In particular, the reliability of pitch matches was much poorer between than within sessions. The authors concluded that the FCDS procedure “offers hope that an optimal means of tinnitus measurement has at last been found. However, despite the best technique that psychoacoustics has to offer, between-session measures of the tinnitus fluctuate” (p. 699).

Mitchell et al (1993) developed a computer program for semiautomated evaluation of tinnitus pitch and loudness matches and minimum masking levels (MMLs). All test stimuli were presented according to algorithms under computer control, and subjects responded by pushing a switch located on the armrest of their chair. Instructions for responding were provided verbally by the examiner. The measures were obtained from patients with tinnitus and from a control group. For the nontinnitus subjects, a low-level tone at 2150 Hz was used to simulate tinnitus. Measures were repeated over two sessions to evaluate reliability because “Reliable measures of tinnitus are needed to allow treatment-related changes in tinnitus to be studied” (p. 150). Loudness match variability was reported to be comparable to previous reports showing good reliability (Bailey, 1979; Vernon et al, 1980b; Tyler and Conrad-Armes, 1982) but contrasted with the results of Penner (1983). The variability of pitch matches was comparable between tinnitus and external tones, which also disagreed with others’ findings (Penner, 1983; Tyler, 1991). MMLs were obtained only in the subjects with tinnitus, and MML reliability was only slightly poorer than for the loudness matches.

Henry et al (1996b) addressed the need to develop standardized techniques for the clinical quantification of tinnitus. They conducted a pilot effort to demonstrate the feasibility of obtaining reliable tinnitus loudness and pitch matches using a fully automated computerized system. With their procedure, the entire testing protocol was under computer control, including instructions for responding. The testing algorithm was designed to closely replicate the clinical testing protocol for tinnitus loudness and pitch matching, as described by Vernon (1987). Hearing threshold evaluation, tinnitus loudness matching, and tinnitus pitch matching were sequenced to ensure that pitch matching tones were presented only at levels that were previously matched to the tinnitus loudness. Ten subjects with tinnitus showed good between-session reliability for tinnitus loudness matches (all r’s ≥ .768; p’s ≤ .0094 for 12 test frequencies) and pitch matches (r = .786; p = .007).

Between-session reliability of loudness matches at multiple frequencies was compared between the fully automated system and Vernon’s (1987) manual procedure (Henry et al, 1996a). With the manual procedure, 22 patients were evaluated over two sessions, and Pearson r’s were ≥ .70 at all test frequencies (p’s < .001). With the automated system, 20 subjects showed improved between-session reliability as indicated by Pearson r’s ≥ .89 (all p’s < .0001). The authors suggested that an automated method such as theirs may ultimately provide optimal measurement reliability for clinical purposes because procedural variations are reduced with automation. Work with the automated system is ongoing, and further data confirming the reliability of loudness matches with this technique have recently been reported (Henry et al, 1999).

Psychoacoustic Studies of Tinnitus Masking

Clinical use of tinnitus masking has involved bands of noise, primarily because tonal maskers are not well tolerated by most tinnitus patients. It should be emphasized that masking of tinnitus using synthesizer-generated noise bands produces complete “coverage” of the tinnitus (i.e., renders it inaudible) in 90 percent of tinnitus patients (Vernon and Meikle, 1988). An additional 5 to 6 percent experience masking that is “partial,” that is, the tinnitus is reduced in loudness. These clinical observations make it clear that masking-induced reduction or elimination of tinnitus perceptions is nearly a universal phenomenon.

Psychoacoustic investigations of tinnitus have been concerned mainly with the use of tonal maskers. Several lines of investigation have characterized such studies, including the extent to which tinnitus masking exhibits frequency dependence, possible adaptation phenomena, and diotic versus dichotic effects.

Lack of Frequency Dependence for Tinnitus Masking

As discussed in an earlier section, the frequency relationships of tinnitus masking effects received their first systematic investigation by Feldmann (1971), who evaluated tinnitus mask-
ing in 200 subjects. In more recent times, Formby and Gjerdingen (1980) described tinnitus masking curves from one individual, using masking tones at 14 frequencies between 521 and 3629 Hz. The shapes of the masking curves resembled those typically obtained with tone-on-tone masking (Wegel and Lane, 1924), leading the authors to suggest that such tuning might provide an efficient means to identify the tinnitus frequency. These results, however, contradicted most other works, including that of Feldmann (1971), and have not been replicated since.

Additional studies addressing the frequency dependence (or lack thereof) of tinnitus masking were conducted by Mitchell and his colleagues (reported in Vernon et al, 1980b; and in Mitchell, 1983). A total of 32 subjects were tested and found to exhibit the same types of masking curves described earlier by Feldmann. “Convergent” masking curves (Feldmann Type 1) were found in 52 percent of Mitchell’s subjects; “congruent” curves (Feldman Type 3) occurred in 19 percent; curves similar to Feldmann’s Type 4 (“distance” type) occurred in 22 percent; and in the remaining cases, no masking could be obtained (similar to Feldmann’s Type 5, “persistence” type). Similar observations have also been reported by Hazell (1981), Shailer et al (1981), Tyler and Conrad-Armes (1984), Penner (1987), Penner and Klafter (1992), and Mitchell et al (1993). All investigators agreed that the usual marked frequency dependence seen in conventional tone-on-tone masking is not characteristic of most attempts to mask tinnitus. The conclusion from the various studies is that tinnitus masking appears to be categorically different from the conventional masking of one external tone by another.

A similar statement applies to the concept of critical bands for masking, as stated by Shailer et al (1981). In that study, masker levels were determined as a function of masker bandwidth in 12 tinnitus subjects. An additional six control subjects were presented with a 4-kHz tone. The control subjects displayed the typical critical band function, as did four subjects with tone-like tinnitus. The other three subjects with tonal tinnitus and the five subjects with noise-like tinnitus all had masked thresholds that were independent of bandwidth. Thus, for eight of the tinnitus subjects, the effectiveness of masking was independent of the masker bandwidth. However, for the four subjects who showed critical masking bands for their tinnitus, their tinnitus seemed to “behave” similarly to an acoustic pure tone when subjected to maskers of differing bandwidths.

The lack of critical masking bands in many cases of tinnitus masking was also noted by Johnson and Mitchell (1984), who commented: “The width of the masking signal appears not to be a determining factor regarding the effectiveness of masking tinnitus. For many patients, the tinnitus can be masked more effectively with pure tones or very narrow bands of noise than it can for broad bands of noise....”

Tyler and Conrad-Armes (1984) had noted that the abnormal masking patterns that generally resulted from tinnitus masking experiments might be attributable to the reduced frequency resolution often associated with hearing impairment, as manifested by broad psychoacoustic tuning curves. They therefore studied whether poor frequency specificity seen with tinnitus was a property of the tinnitus itself or a result of hearing impairment. Using 10 subjects with tinnitus and sensorineural hearing loss, masking of tinnitus with pure tones was compared to masking of an external tone matched in loudness and frequency to the tinnitus. Results revealed that most subjects had poor frequency resolution for both types of masking curves and that different masking patterns emerged from individual subjects.

**Adaptation of Tinnitus Masking**

Penner et al (1981) suggested that neural adaptation of broadband noise and an external tone being masked by the noise should occur at the same rate, which would indicate that the level of noise would not have to be raised over time to provide continuous masking. In 20 subjects with tinnitus, their tinnitus was the “tone” being masked, and the noise had to be increased to continuously mask the tinnitus. Penner concluded that tinnitus responds differently from an external tone when masked by broadband noise; thus, the tone and the tinnitus operate by different neural mechanisms. Penner later hypothesized that the noise might have exacerbated the tinnitus, causing it to become louder and thus more audible during masking (Penner, 1988). She tested this hypothesis using a different task in which the masking level was fixed and subjects were asked to adjust the level of a pulsed tone so as to remain at the same loudness as the tinnitus. When the pulsed tone was presented ipsilaterally, it had to be increased substantially, but when presented contralaterally, the change was minimal. This contradiction raised the interesting question of which matching tone—ipsilateral or contralateral—best represented the tinnitus loudness.
Penner and Bilger (1989) attempted to determine whether increases in masker levels needed to "cover" the tinnitus were comparable in magnitude to decreases in loudness due to adaptation (i.e., tone decay). They noted that even moderate tone decay (10-15 dB) observed in patients with sensorineural hearing loss (Dirks et al., 1974) could not account for the 30-dB shift in masker level necessary to mask tinnitus. They therefore tracked the intensity of a pure tone required to mask tinnitus in six subjects. In the same subjects, they also measured adaptation by tracking the intensity of a pure tone required to maintain constant loudness. In three of the subjects, stimulus changes required to compensate for adaptation were similar to changes needed to maintain masking; in the remaining three subjects, there were large differences in the stimulus levels required for maintaining constant masking versus constant loudness.

Johnson and Hughes (1992) addressed the question of whether some patients can be masked more effectively when the masking noise is presented diotically rather than dichotically. They tested patients with bilateral tinnitus using special equipment that could deliver masking noise either diotically (the same noise signal delivered to both ears) or dichotically (different noise generators with identical spectral characteristics independently delivering noise signals to the two ears). Ten of the patients were not candidates for comparing the two types of noise (four could not be masked at all and six were maskable only in one ear). Of the remaining 18 patients, 3 could be masked dichotically but not diotically, 8 could be masked both dichotically and diotically, and 7 could only mask each ear monaurally (i.e., bilateral masking was not effective under either condition). There were no subjects who could be masked diotically but not dichotically. Measurement of MMLs showed that the levels of noise needed to accomplish effective masking were considerably greater for the diotic than for the dichotic condition. The investigators concluded that the success of tinnitus masking is greatly influenced by central neural influences and that the superiority of dichotic masking was consistent with what is known about Masking Level Differences (Jeffress et al., 1952).

**Psychoacoustic Studies of Residual Inhibition**

Some degree of residual inhibition (RI) following tinnitus masking can be observed in nearly 90 percent of all tinnitus patients. Such a high prevalence rate suggests that RI is a fundamental aspect of tinnitus masking phenomena. With current masking techniques, RI exhibits high variability between subjects both in regard to its duration and magnitude. Measurement of RI durations in a large sample of tinnitus clinic patients (Meikle et al., 1995) yielded durations varying from less than a minute (57% of subjects) to more than 10 minutes (3% of subjects). If the duration of RI could be extended, it could provide substantial relief for many tinnitus sufferers. Techniques to increase the magnitude of RI in subjects who experience incomplete suppression of tinnitus (i.e., those experiencing "partial" RI) would also extend the therapeutic benefit. For that reason, most of the psychoacoustic research concerned with RI has focused on ways to render it longer lasting and more complete.

In his pioneering work on tinnitus masking and residual inhibition, Feldmann (1971) demonstrated that RI induced by short bursts of interrupted masking stimuli was influenced by the intensity of the masker in some subjects but not in others. Later, Feldmann (1983) showed that patients could control the duration of masking bursts and length of interburst intervals so as to completely eliminate their perception of tinnitus. That technique for producing RI has not been pursued by other investigators, although the results clearly presented a potentially effective approach to tinnitus treatment.

Parametric studies of factors influencing RI have been undertaken by several investigators. In 1982, Vernon described the effects of altering the intensity of the masker upon RI duration using narrow bands of masking noise centered at the tinnitus frequency. In eight subjects, masker intensities were varied in 5-dB steps from MML + 5 dB to MML + 25 dB. It appeared that increasing the masker intensity tended to increase RI durations, but the inter-subject variability was very large (the measured trend was not statistically significant). Manipulation of masker intensities was also performed by Bailey (1979) using masking tones at various frequencies. He too found that the duration of RI increased with increasing masker intensity, although the reported observations involved only a few subjects.

Increases in masker intensity were studied in a group of 10 subjects by Tyler et al. (1984). In general, longer durations of RI resulted when the masker intensity was elevated.

Terry et al. (1983) studied RI in 32 subjects, obtaining a number of interesting results. They
found significant effects of varying the duration of masking in the range of 10 sec to 10 min, with RI durations linearly related to the logarithm of masker duration. Bailey (1979) and Tyler et al (1984) also examined the effect of extending the masker duration. They concluded that longer durations of masking tended to produce longer lasting RI, although the results reported by Bailey are difficult to interpret because it is not clear whether group measures or data from a single patient are presented.

The results of varying the masker frequency were quite discrepant between the various studies cited above. Tyler et al (1984) and Vernon (1985) reported little effect of frequency variations; Bailey (1979) claimed that there was a very sharply tuned dependence of RI on masker frequency (but data from only one subject were shown), and Terry et al (1983) found that the optimal frequency for producing RI was usually lower than the individual’s tinnitus frequency.

Tyler et al (1984) and Vernon and Meikle (1988) studied the recovery patterns of RI. They reported that several different recovery patterns could be observed as the tinnitus returned to its normal level. It is not clear to what extent these recovery patterns exhibit test–retest reliability.

There has been little or no systematic evaluation of the variability of RI measures, but there is some evidence that there is considerable variation from trial to trial within one subject even when masking parameters are held constant (Meikle, unpublished data). The topic of RI remains an important and challenging one for further study.

RELATING PSYCHOACOUSTIC MEASURES TO TINNITUS SEVERITY AND HANDICAP

From the time of Reed (1960), investigators have noted poor correlations between psychoacoustic measures of tinnitus and the subjective dimensions of tinnitus reported by tinnitus patients. Graham and Newby (1962) reported that none of their psychoacoustic measurements related to subjective descriptions of tinnitus. Meikle et al (1984) analyzed psychoacoustic data (pitch, loudness, and maskability) from nearly 2000 tinnitus patients. Patients also rated the severity of their tinnitus on a scale from 1 to 10, with 10 representing “extremely severe.” There were no significant correlations between the psychoacoustic measures of tinnitus and the subjective ratings of tinnitus severity.

Hallam et al (1984) reported tinnitus psychoacoustic data on two groups of 13 tinnitus patients each. For one group, the tinnitus was a main complaint, and in the other group it was incidental to either impaired hearing or dizziness. Measures of tinnitus pitch, loudness, and maskability were almost identical between groups.

Jakes et al (1985) found no relationship between measures of tinnitus loudness and various measures of tinnitus complaint, except for a “slight” relationship ($r = .34$) between the loudness match at 1 kHz and tinnitus “intrusiveness.”

Hazell et al (1985) reported outcome measures from 472 tinnitus patients from three hospitals who were fitted with maskers. Correlations between psychoacoustic measures of loudness and the loudness ratings were low. Also, the relationship between residual inhibition and masker effectiveness was not significant. It was generally concluded that tinnitus tests could not predict treatment benefit, nor did they correlate with measures of subjective tinnitus severity.

Hallam et al (1985) found that loudness matches expressed in dB HL or dB SL were poorly correlated with tinnitus severity and that transformation to sones did not improve the correlation. They noted modest improvement in the correlations when these measures were converted to PLUs, which took into account individual loudness growth.

Risey et al (1989) showed that loudness matches at 1 kHz were significantly correlated with tinnitus severity ratings, whereas those at the tinnitus frequency were not. The size of the correlations at 1 kHz were $r = .224$ in linear regression analysis and $r = .458$ when curvilinear regression was used. These values reveal that the loudness match accounts for roughly 20 percent of the variance in subjective severity ratings. They also demonstrated that when patients reported a decrease in their tinnitus after furosemide treatment, the 1-kHz loudness matches and severity ratings were reduced, and these changes were significantly correlated. It should be noted, however, that their 1 to 10 severity scale indicated “1” as the “softest or mildest tinnitus imaginable” and “10” as the “worst tinnitus imaginable.” That wording may have led subjects to interpret the severity ratings as a scale of perceived loudness and not of tinnitus-related disability or handicap. We have already commented on the ambiguity inherent in the word “severity.”

The Tinnitus Handicap Questionnaire developed by Kuk et al (1990) was administered to
patients who provided subjective loudness ratings (on a scale from 1–100) and psychoacoustic measures (loudness matches in sones, obtained at 500 Hz). There was a moderately high correlation between the subjective loudness ratings and the degree of tinnitus handicap (r = .57; p = .0001) but a low correlation between the loudness matches (in sones) and tinnitus handicap (r = .27; p = .04). The authors suggested that the psychoacoustic measures and the subjective ratings did not test the same attribute(s) of tinnitus.

Halford and Anderson (1991) developed the Subjective Tinnitus Severity Scale, designed to yield a single score that reflected how "intrusive, prominent and distressing" tinnitus was to the patient. Tinnitus loudness matches were obtained from these patients at 1 kHz and at the tinnitus frequency. Scores from the Severity Scale showed modest but significant correlations with the loudness matches at 1 kHz, expressed in dB HL (r = .48; p < .01) and dB SL (r = .36; p < .05), and with the loudness matches at the tinnitus frequency in dB HL (r = .41; p < .05).

Dineen et al (1997) obtained a profile of audiologic and psychological characteristics in patients prior to initiating a program of tinnitus management. Tinnitus measures included pitch matches, loudness matches at 1 kHz and at the tinnitus frequency, MMLs, and a 10-point visual analog scale for tinnitus loudness. The patients were divided into two groups: those who had previously sought professional help for tinnitus (and reportedly received no benefit) and those who had not previously sought help. There were no significant differences between groups for the various psychoacoustic measures of tinnitus. The previously helped group did, however, show higher scores on the visual analog scale of tinnitus loudness than the no-help group. The authors concluded that a patient's decision to seek professional help for their tinnitus is not significantly affected by the psychoacoustic characteristics of tinnitus.

CONCLUSION

From this review, it is evident that a great deal of innovative work has gone into the development of psychoacoustic measures for tinnitus. Attempts to quantify the pitch, loudness, and maskability of tinnitus have become increasingly sophisticated, as have the protocols for evaluating the validity and reliability of such measures. Unfortunately, current techniques for tinnitus measurement are almost as numerous as the investigators who employ them, highlighting the need for standardization of measurement methods.

At this point, it might be helpful to review what we know. We know that tinnitus sensations can be described in acoustic terms and that many attempts to match tinnitus sensations with external acoustic stimuli have been successful, although data concerning the reliability of these measures have been incomplete. We know that appropriate masking sounds can reduce or eliminate the perception of tinnitus in approximately 95 percent of all patients (Vernon and Meikle, 1988). Despite the highly individualized nature of tinnitus perceptions, masking curves reveal that for many individuals, the effectiveness of tonal maskers is independent of frequency. We know that the phenomenon of residual inhibition occurs for the large majority of tinnitus, indicating a promising area for future research into techniques that could prolong residual inhibition.

There is general agreement that tinnitus does not behave like external sounds in several ways, despite the fact that tinnitus is characterized by dimensions such as pitch, timbre, loudness, and localization that mimic the perceptual dimensions of external sounds. It is not yet clear whether the ways in which tinnitus differs from external sound perceptions result from hearing impairment and related abnormalities such as recruitment, diplacusis, and tone decay or whether they result from special neurologic mechanisms, as yet unknown, that may give rise to tinnitus.

For all that is known, it is the unknown and the issues of disagreement that currently prevent development of more uniform methods for testing tinnitus. An example is the disagreement concerning the amount of variability in regard to tinnitus loudness matches and pitch measures. Some of the observed differences undoubtedly reflect true differences in the reliability of testing techniques. However, other differences result from the differing ways in which subjects are selected — tests that clinic patients receive as part of the intake procedure, for example, are likely to reveal more response variability than tests of individuals invited to participate in the more rigorous efforts of a psychoacoustic laboratory. Other potential factors contributing to observed differences in variability are differences between subject populations (e.g., severe versus mild tinnitus, or differences in regard to the nature and extent of hearing impairment). It is our hope that by
highlighting the types of disagreement characterizing past research, the present review will encourage tinnitus researchers to work together to develop more uniform testing techniques.

Psychoacousticians are justified in emphasizing the need for greater attention to the measurement technique and also for advocating psychoacoustically rigorous techniques. However, clinicians are justified in emphasizing the difficulty of applying psychoacoustic techniques in the clinic, where many patients cannot perform the type of test that would be done in a psychoacoustics laboratory, and time constraints prevent lengthy procedures. We believe that workable compromises can be reached between the psychoacousticians and the clinicians. It should be recalled that it also took some years for standardization of audiometric tests, and in some areas of concern to audiologists the tests are still being refined (e.g., measures of central auditory processing). Tinnitus is about 40 years behind hearing impairment in regard to standardization of primary testing techniques.

When a patient seeks a hearing assessment from an audiologist, a fairly uniform testing protocol is conducted. Test results from one clinic can be readily interpreted by an audiologist in another clinic—across the country and even around the world. However, when a patient comes to the clinic with the primary complaint of tinnitus, there is no standard protocol for tinnitus evaluation. Consequently, the clinics that provide tinnitus services, which are relatively few, do so in an independent manner. Test results remain proprietary and often uninterpretable by other clinicians.

What should be done? Fifteen years ago, a plea was made for standardizing tests of tinnitus, and that need has yet to be met. It is clear that tinnitus investigators should strive for standardization of psychoacoustic techniques as an important source of information needed by investigators in many areas of tinnitus research including brain imaging, development of animal models, and clinical trials of proposed new tinnitus therapies. It is our hope that these techniques can be developed collaboratively by leaders in the field, thus promoting comparison of results between different laboratories and clinics and providing a consistent set of guidelines that can be applied widely in further tinnitus research.

Acknowledgment. Funding was provided by the Veterans Affairs Rehabilitation Research and Development Service.

REFERENCES


