An automated procedure for identifying spontaneous otoacoustic emissions

Edward G. Pasanen and Dennis McFadden a)

Department of Psychology and Institute for Neuroscience, Mezes Hall 330, University of Texas, Austin, Texas 78712

(Received 30 December 1999; revised 3 May 2000; accepted 18 May 2000)

An algorithm is described for objectively identifying and measuring spontaneous otoacoustic emissions (SOAEs) using the spectrum that results from transformation of the acoustic waveform measured in the outer ear canal. Prior to spectral analysis, the rms level is calculated for successive short segments of the waveform and only the weakest $25\%$ of the segments are retained for the spectral analysis [the quietest 150 when using 16k-point fast Fourier transforms (FFTs)]. The resulting initial spectrum is scanned for peaks (potential SOAEs) which are then deleted from the spectrum. New values are estimated for the deleted values using linear extrapolations from frequency ranges on either side of the deleted values. The end result is a smooth spectrum devoid of all local peaks. The initial spectrum is then compared peak-by-peak with the smooth spectrum, and those peaks having differences that exceed an objectively determined decision criterion are identified as likely SOAEs. The effects of varying some of the important parameter values of the algorithm are described, and the sensitivity of the procedure is evaluated by measuring the detection rate for a Lorentzian peak of known amplitude added to a spectrum otherwise devoid of SOAEs.

$^a$Author to whom correspondence should be addressed. Electronic mail: mcfadden@psy.utexas.edu

PACS numbers: 43.64.Jb, 43.64.Yp, 43.58.Ta [BLM]

INTRODUCTION

Spontaneous otoacoustic emissions (SOAEs) are narrow-band sounds that are generated in the cochlea and then propagate back through the middle ear into the outer ear canal where they are detectable using a sensitive microphone system. SOAEs are continuously present, relatively stable, and can be found in the majority of ears with normal hearing (see Probst et al., 1991, for a general review). While these sounds can be as strong as 30 dB SPL, the majority are much weaker, and the weakest ones detected are only about 1 dB above the noise floor of the measuring system. As instrumentation and analysis methodology have developed over the years since SOAEs were first detected by Kemp (1979), the reported prevalence and number of SOAEs in human ears has steadily increased (see Talmadge et al., 1993; Zhang and Penner, 1998). It appears that SOAEs are likely to be found in greater numbers as the state of the art in instrumentation and analysis procedures continues to advance.

Historically, SOAEs have typically been identified by an observer simply examining a spectral representation of the waveform collected from the ear canal and noting those peaks that appeared to exceed the background noise floor. However, there are a number of situations where an objective, automated system for detecting SOAEs is desirable. If the number or size of the SOAEs in two or more populations or treatment groups are being compared, the scoring procedure should be free of bias. Identification of SOAEs “by eye” involves a certain amount of practice, and when the task falls upon the same experimenters who generated the hypotheses under test, readers and reviewers can become concerned about possible biases. Furthermore, if there are many ears to be analyzed, automation can save labor. In addition to objectivity and efficiency, an automated system offers the availability of a repeatable, standard procedure, assuring experimenters in different laboratories that their methods are identical, and their results truly comparable.

Although early procedures for SOAE detection could be made somewhat objective (see Whitehead et al., 1993, for example), it was not until the implementation of off-line analyses of digitized ear-canal recordings (see Bilger et al., 1990; Rao and Bilger, 1991) that it was possible to automate the process. Having a relatively long sample of the waveform in the outer ear canal made it possible to establish a noise criterion for acceptance of individual segments for spectral analysis, enabling analysis and averaging of only those samples relatively uncontaminated by noise. Talmadge et al. (1993) reported an analytic procedure for quantifying SOAEs based on digital recordings, and while their detection procedure was not fully automated, it was highly objective. Penner et al. (1993) reported an automated procedure for detection of SOAEs from digital recordings, and she and her colleagues have subsequently explored several approaches to the detection problem (Penner and Zhang, 1997; Zhang and Penner, 1998). Our laboratory began an independent investigation of the problem, motivated by the need to detect and measure SOAEs objectively in a large number of subjects from several different subpopulations (McFadden and Pasanen, 1999). Here, we describe an automated procedure that uses off-line analysis of digital recordings of the output of a commercially available OAE measurement microphone system. An early version of our procedure was presented in
Pasanen and McFadden (1996), and then described briefly in McFadden and Pasanen (1999). Data from the latter study will be used here to explicate and evaluate our SOAE detection procedure, and will be compared with results from Talmadge et al. (1993), Whitehead et al. (1993), and Zhang and Penner (1998).

Although our procedure involves a relatively large number of steps, the computations themselves are mathematically simple, easy to implement, and quite rapid. After the ear canal recording is obtained, the complete analysis for a single ear currently requires less than 30 s of computation time on a relatively fast (in 2000) desktop personal computer (Power Macintosh G4, 400 MHz). It is therefore rapid enough to be accomplished while a subject is still in place, thereby allowing for the collection of replacement data when the need arises.

There are two extreme cases that pose special challenges to any SOAE-detection procedure. First is the problem of detecting weak SOAEs. The SOAE-detection process always involves identification of a peak in a noisy spectrum, and then estimation of the likelihood that the peak in question arose from a process other than the observed noisy background. The value of any SOAE-detection procedure lies to a great extent in the confidence the user has that a relatively small spectral peak can be called an SOAE. The other extreme case challenging any SOAE detector is the ear containing many closely spaced SOAEs. Figure 1 is a spectral plot obtained from such an ear. Identifying the strong SOAEs in this ear would be easy using almost any procedure, but it is a far greater challenge to develop a system for identifying the many weak components present. In the development of the procedure described here, particular attention was devoted to satisfactory treatment of these two extreme conditions of weak SOAEs and multiple, closely spaced SOAEs.

Whether the SOAEs in question are weak or strong, their detection always involves a criterion—implicitly even when not explicitly. There are in fact situations where it is advantageous to allow the detection criterion to vary in a systematic and quantifiable way. For example, it is known that SOAEs are more prevalent, more numerous, and more powerful in female ears than in male ears (e.g., Talmadge et al., 1993). The use of an overly strict criterion when comparing SOAEs in the two sexes could exaggerate the true sex difference because male ears have weaker and less easily detected SOAEs. If an experimenter were able to vary systematically the criterion for accepting a spectral peak as an SOAE while making comparisons between males and females, any group differences found could be stated more precisely (e.g., the group difference existed independent of criterion, only with a strict criterion, or whatever). Although traditionally it has been desirable to minimize the likelihood of erroneous identification of a spectral peak as an SOAE when it is not (a false alarm), it might sometimes be preferable to accept a very slight increase in false alarms as a small price to pay for the proportionally greater number of SOAEs detected, and (particularly in male subjects, for example) the greater number of ears contributing to SOAE statistics for the group. In regard to the criterion issue, our desire was to develop an automated detection procedure that was “free-standing” or “self-contained,” in that the criterion would be based on the data at hand rather than on normative data collected elsewhere.

Figure 1 illustrates several characteristics of SOAEs: the minimum spacing between SOAEs is approximately proportional to emission frequency (see, e.g., Talmadge et al., 1993; Zwicker, 1990, for reviews) and the bandwidths of emissions at their bases show a similar variation with frequency. Accordingly, the procedure described here involves some parameters that change proportionally with frequency. We believe that this yields a detection procedure that is more effective in detecting weak emissions than any of the more general-purpose smoothing and peak-detection algorithms that are available.

I. PROCEDURE

This section is intentionally lengthy out of the desire to provide sufficient detail for readers to implement or change the detection procedure in their own laboratories. For what follows, the entire spectrum-averaging and SOAE-detection procedure was implemented off-line. However, with advances in computing speed, the procedure is implementable with minimal delay after the recordings are obtained. All of the data shown are from subjects in an experiment reported by McFadden and Pasanen (1999) that was concerned with differences in the SOAEs in people of differing sexual orientations. A complete description of the data-collection procedures can be found in that paper. The raw data were in the form of digitized records of the acoustic waveform obtained from a microphone (Etymotic ER-10) placed in the ear canal of a human subject lying quietly in a soundproofed room. The waveform was amplified and high-pass filtered at 400
Hz in order to remove the noise from bodily movements, etc. The sampling frequency was 22.05 kHz. Typically, the data were obtained as four 30-s recordings taken from each ear, and, typically, the time interval between successive recordings was 10–60 s. All data collection and subsequent analyses were performed on a Macintosh personal computer, using the LabVIEW® programming environment National Instruments, and this platform-independent software is available from the authors upon request. The data acquisition and digital control interface hardware were plug-in boards also manufactured by National Instruments.

The SOAE-detection procedure is broadly divided into five phases. First is the extraction of the initial spectrum from the waveforms recorded in the ear canal. Second is the determination of the smooth, “frequency response” function (called the smoothed baseline) from that initial spectrum. Third is the expression of spectral peaks as deviations from this smoothed baseline. Fourth is identification of SOAEs, and fifth is the assignment of frequency and level to the peaks identified as SOAEs.

**A. Phase 1: The initial spectrum**

The first goal of the procedure was to eliminate the noisiest time periods in the recordings prior to performing spectral analysis. The entire 2-min record was divided up into successive, overlapping segments of approximately 743 ms duration each, with the onsets of successive segments being separated by approximately 186 ms. Given the sampling rate used to make the digital recordings (22.05 kHz), these values correspond to 16,384 sample points per segment, and 75% overlap of successive segments. The rms level was computed for each time segment, and a histogram of rms values was compiled across all the time segments for that ear. For each of the 150 time segments having the lowest rms values, fast Fourier transforms (16k FFTs, Hanning window) were computed and averaged. (That is, only the quietest 25% of the available time segments was used for the analysis.) The resulting averaged spectrum will be called the initial spectrum; the estimates of level (in decibels) are in frequency bins 1.35 Hz in width. An example is shown in...
Fig. 1. This spectrum can be conceptualized as a relatively uniform noise process across the frequency range, with certain positive peaks deviating significantly from that noise background. The general aim of the automated SOAE-detection procedure is to determine for each local peak in the spectrum a measure of the likelihood that it resulted from the broadband noise process that comprised the background. When that computed likelihood is very small, it is assumed that the peak arose from a different process, namely the process responsible for producing an SOAE.

It should be noted that all smoothing and detection operations described here were performed in semilog space; i.e., ordinate units of decibels and abscissa units of frequency in linearly spaced bins. Linear units of power were used only for the final calculation of the total power\(^2\) for an individual SOAE [see Sec. E (phase 5) below] and for the simulation of an emission (Sec. II C).

B. Phase 2: The smoothed baseline

Ignoring the SOAEs for a moment, it is clear from Fig. 1 that the initial spectrum is not flat. The frequency response of the recording system and the spectrum of the noise in the ear canal combine to shape the overall contour seen in Fig. 1. It is useful therefore to model the initial spectrum as a slowly varying, smooth function (the frequency response), upon which is superimposed a rapidly varying, fine-grain process consisting of noise and SOAEs. (The character of this noise process was established to be essentially constant over the entire frequency range under study, and to be similar for male and female ears.)

Obtaining the smoothed baseline involves (a) scanning the initial spectrum for peaks that deviate substantially from the rest of the spectrum in that local region; (b) removing those peaks plus any surrounding deviant values, and substituting estimated values that are in better agreement with the data in that immediate vicinity; and (c) smoothing using conventional rectangular and median filters on the spectrum containing the substituted values. The substitution procedure is illustrated in Fig. 2.

To begin this process, a pass is made through the entire spectrum, bin by bin, and each spectral value is compared with two acceptance limits calculated from neighboring data in order to determine whether the value is sufficiently extreme to be deleted and replaced. The two acceptance limits are obtained by calculating means and standard deviations for two 30-bin frequency regions flanking the spectral value of interest, one on its low-frequency side and one on its high-frequency side. If the spectral value under consideration is greater than the mean plus 3 standard deviations (SDs) for either 30-bin region, that frequency bin is flagged for replacement; otherwise, it is kept, and the algorithm slides up one bin and repeats the process. The original spectral value is replaced by a calculated value in the following way (see Fig. 2). A linear regression line is fitted to the points in each of the sliding 30-bin frequency regions that were used to calculate the means and standard deviations mentioned above, and the two regression lines are extrapolated to the frequency bin containing the flagged point. One extrapolation is upward from the frequency region below the spectral value of interest, and the other extrapolation is downward from the corresponding region above the flagged point. The two extrapolated values are averaged, and that average is substituted for the flagged point. Then, the algorithm slides up one bin and repeats the above process. To ensure convergence of this process (see below), the entire first pass of checking and extrapolation is performed using only the original spectral values. The new spectral values arising from extrapolation are not used until the next step in the procedure.

The smoothing process begins with the 300-Hz bin. For that bin, the 30-bin acceptance and extrapolation intervals begin 3 frequency bins below and above the 300-Hz bin. This 3-bin interval (3.04 Hz) is called the lag, and is varied proportionally to the frequency of the spectral point under evaluation (as is the frequency separation of actual SOAEs—see Zwicker, 1990; Probst et al., 1991 for reviews). For example, the lag at 900 Hz is 9 bins (12.1 Hz), and at 9000 Hz is 90 bins (121.1 Hz). As noted above, the output of the ear-canal microphone was high-pass filtered at 400 Hz before being amplified and digitized, to reduce the problem of low-frequency noise overloading the recording system. Because there is often substantial biological noise remaining, SOAEs are difficult to detect reliably at frequencies immediately above 400 Hz. Here, only SOAEs above 550 Hz are considered. Smoothing begins at 300 Hz to ensure that no transient artifacts from the initiation of the smoothing process are present above 550 Hz.

During this initial smoothing stage of the procedure, every point in the original spectrum is evaluated for its consistency with neighboring values, and a new, somewhat smoother spectrum is generated as a result, with the larger peaks absent or at least reduced considerably. Note that a log is kept of those bins in the new spectrum that now contain substituted values.

Next, several additional passes are made through the new spectrum in order to smooth it further. The procedure used is slightly different from the preceding stage in order to avoid the problem of recursive substitution. In our algorithm, newly substituted values are never immediately used for extrapolation during the current pass. (A recursive substitution procedure can, in some instances, generate a function that diverges substantially from the shape of the initial spectrum.) Following the initial substitution pass, the spectrum is smoothed using a 19-point median filter. A second substitution pass is then made upon the new, smoothed spectrum, again starting at the lowest frequency. All previously flagged and substituted points (but only these points) are revisited, and each flagged point is again replaced, this time using only the linear extrapolation from the 30-bin interval above the point. Thus, each logged bin is again assigned a new value, based on the (recently smoothed) neighboring data points. A third substitution pass is then performed, but it proceeds from high frequency to low, and uses only the 30-bin extrapolation interval below each logged point. The 19-point median filtering is then repeated. The two one-sided substitution passes just described are then performed again, followed by a final smoothing of the entire spectrum using a 51-point rectangular window that works from low to high frequency. (This procedure of
C. Phase 3: Spectral peaks

The next stage in the analysis is to identify all positive peaks in the initial spectrum and obtain an estimate of each peak's deviation from the noise floor. For the purposes of this stage, a spectral peak is defined simply as any point in the initial spectrum where the slope changes from positive to negative with increasing frequency. Typically, this is equivalent to a point having a value greater than either of the two immediately flanking points. Under such a broad definition, it follows that there will be many peaks detected in the initial spectrum, typically more than 1500 across the 9-kHz range of frequencies considered here. All such peaks are identified, their locations on the frequency axis noted, and for each peak, a peak deviation is calculated. The peak deviation is the mean of the five spectral values centered at the identified peak in the initial spectrum, minus the mean of the corresponding five values in the smoothed baseline. The set of all peak-deviation values, each corresponding to a unique spectral peak, is assembled into an array. The elements in this array of peak-deviation values are checked to determine whether the spectral bin containing the peak had been flagged during the smoothing operation above, indicating that the peak was sufficiently deviant to give rise to a substituted value during the smoothing process. The set of all unflagged peaks, those not corresponding to substituted points, is now assembled into a second array, and the mean and standard deviation are computed for this set of peak-deviation values. Note that these latter peaks can be traced to regions in the initial spectrum where, in the first smoothing pass, it was determined that an SOAE was not likely to be found. Put in signal-detection terms, this set of unflagged peak-deviation values forms an estimate of the "noise-alone" distribution of deviation values, against which deviation values of observed peaks can be compared.

D. Phase 4: Identification of SOAEs

The next step is to determine whether a given peak in the initial spectrum is to be called an SOAE. Using the mean and standard deviation of the set of unflagged peaks (just described above), each entry in the array of peak-deviation values for all spectral peaks (flagged and unflagged) is now expressed as a multiple of standard-deviation units above the mean. That is, 5-bin peak-deviation values for the initial spectrum are expressed as multiples of the standard deviation of the peak deviations obtained from spectral regions judged likely not to contain SOAEs. It should be clear that the greater the magnitude of the standard-deviation multiple for a particular 5-bin region surrounding a peak in the initial spectrum, the greater the likelihood that the values in that region arose from a process other than that which generated the set of unflagged peaks. Said differently, the larger the standard-deviation multiple associated with any peak, the greater the confidence one has in declaring that peak to be an SOAE. In signal-detection terms, the standard-deviation multiple is the decision variable. Once a criterion value has been selected for the standard-deviation multiple, any peak in the initial spectrum can be tested against that criterion value and an objective, binary decision made about its status as an SOAE.

Because the definition of a peak was simply "a change in slope from positive to negative," there may be many more peaks identified as SOAEs than truly are. For example, a spectral peak associated with a true SOAE may well have secondary peaks along the skirts of the main peak, and some of these might also satisfy the decision criterion. Hence, a final cleanup is necessary. This final stage also takes into account the general belief that two true, independent SOAEs cannot be located within about 0.1 octave of each other (Zwicker, 1990). The cleanup begins by ranking the set of peaks identified as SOAEs in order of their standard-deviation multiples. The peak having the highest multiple of all is accepted finally as an SOAE, and, along with any other peaks within 0.1 octave of that peak, is deleted from the set. From the remaining peaks, the one having the highest multiple is also finally accepted as an SOAE, and those within 0.1 octave are deleted, and so on, until the entire set is checked in this way. This cleanup eliminates secondary peaks occurring by chance in the immediate vicinity of a real SOAE, some intermittent SOAEs, and a portion of the SOAEs that are actually distortion products generated by the interaction of other SOAEs (see Burns et al., 1984; Probst et al., 1991; van Dijk and Wit, 1998). (Note that, in this version of our algorithm, no explicit attempt is made to determine whether identified SOAEs might be distortion products.)

E. Phase 5: Assignment of frequency and level to identified SOAEs

The frequency actually assigned to a peak identified as an SOAE is simply the frequency bin that contained the peak in the initial spectrum. The level of each SOAE detected in a given ear is estimated using the following summation procedure performed on the initial spectrum. Beginning with the spectral peak itself and moving toward higher frequencies, values in successive bins are summed (in units of power) as long as the addition of each additional point increases the sum (total power) by at least 1.0%. Then, the process is continued beginning with the bin immediately below the spectral peak and moving downward. Finally, the sum is converted from units of power back to SPL.
II. EVALUATION OF THE PROCEDURE

Two different objective approaches have been used to evaluate our SOAE-detection algorithm. First, default values for the numerous parameters in the smoothing and detection processes were determined; then the various parameters in the algorithm were varied, the data for a fixed group of subjects were reanalyzed, and the effect of that parameter variation on the summary statistics for that group was observed. The summary statistics used consisted of one global measure on the summary statistics for that group was observed. The decision criterion selected by McFadden and Pasanen (1999) was a standard-deviation multiple of 5.0. That value was chosen by comparing, for a large number of ears, the SOAEs identified by the algorithm to those obtained by visual scoring by experienced observers, in ignorance of the subject group in each instance, and also in ignorance of the plots of Fig. 3. The value of 5.0 was chosen as a slightly conservative compromise; SOAEs so identified by the algorithm were readily accepted as SOAEs after visual inspection of the spectra. When the criterion was set to 4.0, there were a few peaks identified as SOAEs that a human observer might question, but none that an observer would reject outright. However, at a criterion level of 3.0, there were peaks identified as SOAEs that, on inspection, would be considered false alarms. As the criterion value was raised to 7.0 and higher, there were many spectral peaks rejected by the algorithm that an observer would accept.

A. Selection of decision criterion

The most salient of all the parameters in the detection procedure is the standard-deviation multiple used as the decision criterion for acceptance of a spectral peak as an SOAE. That criterion will have a direct effect on almost any summary statistic involving SOAEs. Figure 3 shows three summary measures plotted as a function of decision criterion level, for the right ears of male and female subjects from McFadden and Pasanen (1999). The top panels show the effect of SOAE decision criterion on the proportion of right ears having at least one SOAE detected by the algorithm for three groups of male and female subjects: heterosexual, homosexual, and bisexual (bisexual males, a very small sample, was omitted for clarity). For criterion values in the range 8.0 to 5.0 (strict to lax), the proportion of emitting ears is quite stable, with a sharp break in the curves as criterion was set lower than 5.0 (4.0 for females). For the three groups of females shown in the left-hand panel, the curves are very nearly parallel, suggesting that any conclusions drawn about differences between groups were not dependent on the value of decision criterion selected (provided it was not too lax, of course). The pattern for the middle panels, showing average number of SOAEs per emitting ear, is also parallel and very stable, with the sharp break in the curves occurring at the same criterion value (4.0 SDs) for both sexes. The bottom panels show the effect of decision criterion on the average total SOAE power per emitting ear, and the pattern is somewhat different. Although the absolute values of total power vary (and often substantially) as criterion is varied, the plots are roughly parallel. The difference in total SOAE power between heterosexual and nonheterosexual females, for example, is relatively constant for all values of the decision criterion, even for the most lax values at the extreme right. The data from these three summary measures suggest that a decision criterion somewhere above 4.0 would be desirable.

The proportion of ears exhibiting SOAEs (top panels), the average number of SOAEs per emitting ear (middle panels), and the strength of the SOAEs identified (bottom panels) plotted as a function of the decision criterion adopted (expressed as a multiple of the SD of peak deviation scores) and shown separately for males and females (right ears only). Symbols denote the data from different subject groups. Clearly, all experimental conclusions would be the same for any decision criterion greater than about 4. Data are from subjects in McFadden and Pasanen (1999).
B. Smoothing parameters

In the spectral smoothing stage of the algorithm, the size of the extrapolation and smoothing windows, and the lag in bins separating the extrapolation window and the spectral point under consideration, can in some cases affect the fit of the smoothed baseline to the spectrum in those regions where no SOAEs are present. The success of the procedure depends upon a good fit in those regions because an SOAE is indicated by the very lack of a good fit in a small local region. Any procedure must balance the competing needs to fit well in the broad sense but not so well locally that the smoothed baseline function follows the contours of relatively broadband, weak SOAEs, making them impossible to detect. If the smoothed baseline fits the spectrum poorly over a wide range of frequencies, the mean peak-deviation score for the unflagged points will be inflated, resulting in an artificially strict decision criterion. If the region of poor fit is small, a moderate but broad elevation in that spectral region might be incorrectly identified as an SOAE.

To test the robustness of the algorithm to variations in certain parameter values, a group of heterosexual female subjects from McFadden and Pasanen (1999) was analyzed several times, with a single parameter varied from its default value each time. Four parameters were varied: extrapolation window (20, 30, or 40 bins), lag (2, 3, or 4 bins at 300 Hz and the corresponding proportions at higher frequencies), smoothing window (41, 51, or 61 bins), and the acceptance criterion for substitution and extrapolation in the initial spectrum (2.5-, 3.0-, or 3.5-s.d. multiples). The default value for each parameter was the middle one of each triplet. Three summary statistics for the group—proportion of ears having SOAEs, average number of SOAEs per ear, and average total SOAE power per emitting ear—were computed for each analysis. As each parameter was varied, the overall change in the three summary statistics averaged about ±1%, with a maximum of 1.8% for lag value, smoothing, and extrapolation windows. The analysis was somewhat more sensitive to changes in the criterion for substitution and extrapolation, where the summary statistics varied as much as 2.9%. Because this latter parameter determines directly the set of flagged peaks in the initial spectrum, thereby directly affecting the decision criterion, it is not surprising that selecting the value for that parameter is so critical. The important point is that the summary statistics changed relatively little over fairly wide ranges for these parameters. Unfortunately, there is no “gold standard” for goodness of fit that would guide the parameter-selection process. As noted, the values finally chosen as default for these parameters came from examining spectra from a large number of ears having many SOAEs, and visual assessment of goodness of fit in the neighborhood of strong, closely spaced SOAEs.

It could be argued that, given a value for the decision criterion which results in very infrequent false alarms, the procedure should be optimized for the highest rate of SOAE detection possible. In fact, the default set of parameter values did yield the maximum number of SOAEs detected per ear and the maximum number of ears having at least one emission (or very nearly so), when looking at summary statistics for the various subject groups in McFadden and Pasanen (1999).

C. Detection of simulated SOAEs

SOAE detection can be viewed as an example of the classic Neyman–Pearson decision problem (Van Trees, 1968). In any given ear, the a priori probability of an SOAE occurring at any selected frequency is not known. The best that an observer can do is establish a criterion for identification of an SOAE such that the probability of a false detection is acceptably low. One way to evaluate a detection algorithm under such conditions is by simulation. For example, spectra that by other analyses, or by inspection, observers agree contain no SOAEs can be submitted to the detection algorithm for analysis, and any peaks identified as SOAEs by the algorithm can be viewed as false detections. For an additional test, a spectral peak can be synthesized and added to such spectra to mimic a real SOAE. The sensitivity of the algorithm can be inferred from the likelihood of that known added peak being detected as its peak amplitude is varied. Both of these approaches were used.

Because the noise floor of the recording system is not flat as a function of frequency, any characterization of the probability of detection of an SOAE simply as a function of its SPL is meaningless. However, as noted above, local fluctuations in the noise floor about the smoothed baseline are very similar in magnitude over the entire frequency range under study. It follows that the likelihood of detecting a given SOAE, expressed as a fixed increment above the noise floor (or smoothed baseline) will be approximately constant.

The ears of subjects from McFadden and Pasanen (1999) having no SOAEs detected by the algorithm when the decision criterion was set to 5.0 s.d.’s were rechecked by eye to obtain 60 (16 female, 44 male) having no identifiable SOAEs. Averaged spectra were obtained, based on the 150 quietest time intervals (phase 1 described above). A Lorentzian peak of fixed height and bandwidth was added, in units of power, to various regions of the spectra obtained from each ear. For each ear a starting location was selected at random in the vicinity of 550 Hz. The peak was added to the spectrum at that location, and the resulting spectrum was submitted to the smoothing and detection process. Any peak identified as an SOAE within a 50-bin window (67.5 Hz) centered at the point of addition was scored as a “detec- tion.” The center point of addition of the Lorentzian was then moved 50 bins higher in frequency than the initial random location, and the smoothing and detection procedure was repeated. This 50-bin stepping procedure was repeated so that the addition occurred at locations covering the entire spectrum up to 9 kHz in each of the 60 ears. Each spectral bin was therefore sampled only once on any individual simulation run, ensuring that any aberrant peak in the initial spectrum that might be (incorrectly) identified as an SOAE would contribute only once to the total count of detections. This simulation was repeated for a number of amplitudes of the added Lorentzian peak, at bandwidths typical of actual SOAEs. Repeated over all 60 spectra, the total number of peak additions was about 7500 for each condition for each combination of bandwidth and amplitude. The number of
times the addition resulted in a detection divided by the total number of peak additions provided an estimate of the probability of detection of an SOAE of the selected amplitude and bandwidth.\(^5\)

The peak addition was performed in the following way. Our desire was to find a Lorentzian peak that when added to the spectrum produced a simulated SOAE rising a fixed number of decibels above the noise floor. The noise floor in the range of peak addition was estimated by computing the mean of the spectral points in the smoothed baseline over that range, in units of power. The height of the desired peak in decibels was converted to a linear (power) increment above the noise floor. This provided the height of the resulting peak in power units, called the target peak power. The noise floor power was subtracted from the target peak power to obtain the height of the actual Lorentzian peak to be added to the initial spectrum. The Lorentzian at that height and a selected bandwidth was computed, over a range of 200 bins, and added to the initial spectrum, in units of power. After conversion back to decibels, the resulting initial spectrum was submitted to the smoothing and peak-detection algorithm. The 200-bin interval over which the Lorentzian was computed was selected to be large relative to the peak bandwidth in order to eliminate any transient at either end of the interval resulting from peak addition. For the points most remote from the peak’s maximum, the increment to the initial spectrum was at most 0.01 dB, for the peak height and bandwidth values used in the simulation.

For every added Lorentzian peak (and activation of the detection algorithm), the decision criterion for acceptance of a peak as an SOAE was varied, from very lax to strict. Assuming that at least one peak per ear was identified when the criterion was set to its most lax value, the criterion was then adjusted upward in increments of 0.5 SD, with detected peaks recorded at each stage. The process was stopped when the decision criterion value became so strict that no peaks were identified, and the values of the standard-deviation multiples for each such detected peak were recorded.

Figure 4 shows the rate of detection as a function of the magnitude of the decision criterion for Lorentzian peaks of several heights. The curve at the far right in the left panel is the limiting case, where the peak height was zero, and it offers a qualitative estimate of the false-alarm rate. The vertical dotted line indicates a criterion level of 5.0 SDs, noted above to be judged a good conservative compromise for the data of McFadden and Pasanen (1999). Figure 4 shows that the probability of detecting a peak when none in fact was added is asymptotically low for a criterion value of 5.0 SDs, but unfortunately, using real data, it is not practical to obtain a reliable estimate for a false-alarm rate that is so vanishingly small. A different approach to estimating false alarms is described in Sec. D below.

It is useful to compare the plots of detection rates for the range of peak heights where the detection rate is nonzero but not perfect for a selected value of the decision criterion, such as 5.0. A 1.0-dB peak is only rarely identified at that criterion, while a 2.5-dB peak is identified nearly every time. Simulations were run at several peak heights in that range, and the detection rate for a criterion of 5.0 s.d.’s was noted for each peak height. These data are summarized as a probability of detection function in the right-hand panel of Fig. 4.

The likelihood of the added Lorentzian peak being detected depended upon its bandwidth as well as its height, and the effect of this variable is shown in Fig. 5. For the simulation summarized in Fig. 4, the Lorentzian bandwidth was 10.8 Hz (8 bins). For this exercise, the detection rate was measured as the bandwidth at half-power was varied from 2.7 to 18.8 Hz (2 to 14 spectral bins), with the peak height fixed at 1.8 or 2.0 dB. Detection improved as bandwidth increased, and was nearly asymptotic with a bandwidth of 13.5 Hz. Detection performance was clearly poor for the narrowest peak width (2.7 Hz, or 2 bins). However, for the group of heterosexual female subjects studied in McFadden and Pasanen (1999), the mean bandwidth for SOAEs that were less than 3 dB above the noise floor was 13.8 Hz.

---

**Fig. 4.** (Left): The probability of a Lorentzian peak added to the initial spectrum being classified as an SOAE as a function of the decision criterion adopted (expressed as a multiple of the SD of peak-deviation scores). The parameter on the curves is the height of the added peak (dB), and the curve marked ‘no peak’ corresponds to a peak height of 0 dB. (Right): The probability of a peak being classified as an SOAE as a function of peak magnitude for a fixed decision criterion of 5.0 SDs. The bandwidth of the added peak at half-maximum was 10.8 Hz for each curve.

**Fig. 5.** The probability of an added Lorentzian peak being classified as an SOAE as a function of its bandwidth at half-maximum. Peak height was either 1.8 or 2.0 dB as indicated. Decision criterion was 5.0 SDs. The single open circle is the probability for a 2.0-dB peak when the peak-deviation score is based on 3 spectral bins centered on the peak, instead of 5, as normally used.
suggesting that weak SOAEs are rarely as narrow as 2.7 Hz. Of course, it is possible that this observed mean bandwidth was elevated simply because the narrowest emissions of low amplitude were not being detected by the algorithm. Recall that identification is based on the deviation score for a spectral peak, defined as the mean of the 5 spectral bins centered at the peak, and this definition works against the detection of narrow-band peaks. Therefore, the data for that group of female subjects were reanalyzed, using peak-deviation scores based on only 3 points (instead of 5) centered at the spectral peak, with detected peaks fitted by a Lorentzian function to estimate bandwidth. As a further test, the simulation was repeated for a 2-dB narrow-band peak (2.7-Hz bandwidth), using this 3-point definition for the deviation score, and the detection rate is plotted as the single open circle in Fig. 5. In the simulation, detection performance improved, from about 5% to nearly 50%. However, the reanalysis of real subjects’ data revealed that, for narrow-band (5.4 Hz or less) SOAEs less than 3 dB above the noise floor, only one additional emission (out of 14) was detected. This reanalysis confirms that very weak SOAEs having extremely narrow bandwidths are not common. Furthermore, basing the peak deviation score on 3 points instead of 5 resulted in poorer overall detection of SOAEs (692 vs 712 total emissions in 114 ears).

D. Estimation of false-alarm rate

It is clear from Fig. 4 that the likelihood of falsely identifying a spectral peak as an SOAE was very low for decision criteria greater than about 3.5 SDs. Unfortunately, the estimation of these extremely low probabilities is not a simple matter. In the simulation, the rate of false detections was less than 0.001 for criterion values of 4.25 or higher (and of course was zero for a criterion value of 5.0 SDs) for the sample of 60 nonemitting ears. A theoretical approach using the distribution of peak-deviation scores for unflagged peaks does lead to a reasonable estimate. This distribution of peak deviations is surely not normal because the range of possible negative values is not unbounded. However, the set of deviations lying above the mean did comprise very nearly 50% of all unflagged peaks, and when this upper half of the distribution of deviation scores was compared to the upper half of a normal distribution, the two were found to be very similar indeed. Specifically, the distributions of peak-deviation scores for the set of 60 nonemitting ears used in the simulation above were normalized and compiled for 16 values of standard-deviation multiple above the mean, in steps of 0.25. The results from all 60 histograms were pooled, and then compared with a normal density function with mean zero and variance 1.0. For the upper halves of the distributions, after combining the extreme bins to obtain a count of at least 1, chi-square was computed for the remaining 14 abscissa values. The chi-square sum was 0.793, revealing that the upper half of the distribution is very close to the normal distribution. Priestley (1981) pointed out that a spectral estimate of the nature described here is in fact asymptotically normal. Assuming that the upper half of the distribution is normal, an estimate of false-alarm rate can be obtained from the area in the tail of the normal density function, for values greater than 5 standard deviations. This area is approximately 2.87E-7, the probability that any single spectral peak would be falsely identified as an SOAE. This is close to the estimate Talmadge et al. (1993) reported for their method of requiring five consecutive spectral points lying at least 2 SDs above the noise floor. In our procedure, there are about 1500 peaks in a typical spectrum (16k FFTs), meaning that the expected false-acceptance rate is about 0.00043 emissions per ear (using a decision criterion of 5.0 SDs instead of 5).

<table>
<thead>
<tr>
<th>Females</th>
<th></th>
<th>Males</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevalence</strong> (No. subjects)</td>
<td><strong># Emissions per emitting ear</strong></td>
<td><strong>Prevalence</strong> (No. subjects)</td>
<td><strong># Emissions per emitting ear</strong></td>
</tr>
<tr>
<td>Present study (16K FFTs)</td>
<td>91% (N=57)</td>
<td>6.0 (0.40)</td>
<td>59% (N=56)</td>
</tr>
<tr>
<td>Present study (4K FFTs)</td>
<td>89% (57)</td>
<td>5.9 (0.39)</td>
<td>59% (56)</td>
</tr>
<tr>
<td>Present study (32K FFTs)</td>
<td>91% (57)</td>
<td>5.6 (0.38)</td>
<td>62% (56)</td>
</tr>
<tr>
<td>Talmadge et al. (32K FFTs)</td>
<td>83% (36)</td>
<td>8.3 (0.38)</td>
<td>63% (40)</td>
</tr>
<tr>
<td>Zhang and Penner (32K FFTs)</td>
<td>82% (44)</td>
<td>4.1 (0.38)</td>
<td>68% (39)</td>
</tr>
</tbody>
</table>
E. Number of points per FFT

The data, analyses, and parameter selections reported so far in this paper were based on FFTs of 16k points, but full analyses were also performed using FFTs of 4k and 32k points. Summary data comparing the three analyses for the male and female heterosexual subjects from McFadden and Pasanen (1999) are shown in Table I. The prevalence estimates and the number of SOAEs per emitting ear tabled suggest a slightly better detection performance for the 16k FFT analysis than for the 4k analysis, at least for the female group. In addition, the total number of SOAEs detected for both male and female groups (not shown) increased somewhat when the larger window was used. For a recording of fixed length (in this case, 2 min), there is an obvious tradeoff between FFT window size and maximum number of FFTs available to be averaged. However, even when the number of spectra averaged was a fixed proportion of available records (the quietest 25%), the 16k FFT analysis (150 records) afforded a slight advantage over the 4k FFT analysis (300 records). For the group of heterosexual female subjects in McFadden and Pasanen (1999), the 16k analysis found emissions in 84% of the left ears and 88% of the right ears; the 4k analysis found emissions in 79% of the left and 81% of the right ears. The average number of emissions per emitting ear was very slightly higher for the 4k analysis, with 5.87 (4k, left ear) and 5.67 (16k, left ear), with a similar pattern for the right ear. Of course, the additional emitting ears detected by the 16k analysis would most likely have had relatively few emissions per ear, resulting in a reduced average number per emitting ear.

In the third row of Table I are shown the results of the 32k analysis of the data. For comparability of noise criterion, the quietest 75 samples of 32k-point duration were used for the averaged FFTs. The 16k analysis yielded a total of 584 emissions detected for this subgroup of female subjects, and the 32k analysis yielded a total of 555; for males, the emissions totaled 200 and 185, respectively. On the other hand, the prevalence estimate for males increased slightly with the 32k-FFT analysis. There are certainly a number of factors accounting for the reduced sensitivity of the 32k analysis, but a major one must be the reduction of the number of spectra averaged from 150 to 75. It must be added that a full exploration of the parameter space was not performed for the 32k analysis, nor was the simulation involving the added Lorentzian peak conducted. It is possible that parameter settings could be optimized for slight further improvement in SOAE detection.

Many authors favor analyses based on FFTs of 32k points or more, and Talmadge et al. (1993) pointed out the clear necessity for the finer frequency resolution afforded by such large window sizes when estimating SOAE bandwidth. Estimates of prevalence of SOAEs and number of SOAEs per emitting ear obtained by Zhang and Penner (1998) and Talmadge et al. (1993), both based on 32k FFTs, are also shown in Table I for comparison with our results. The sampling rate used for collecting our data was 22.05 kHz, providing a frequency resolution of 1.35 Hz (16k FFT), identical to that used by Talmadge et al. (1993), and similar to that of Zhang and Penner (1998). Although there would be a theoretical advantage from finer spectral resolution in ability to detect weak SOAEs of very narrow bandwidth, it is the strong emissions, not the weak, that typically have the narrowest bandwidths, and their detectability is never in question. In fact, from our experience, very weak narrow-band spectral peaks are often artifacts (harmonics of 60 Hz, for example), and they are fortuitously rejected by our algorithm. Inspection of the averaged spectra from the 32k FFTs revealed that the lower-frequency regions (below about 1 kHz) were often noticeably noisier than the corresponding 16k-FFT averages. Coupled with the higher prevalence estimate for males shown in Table I, this observation suggests that, where feasible, the 32k-FFT analysis might well be preferable, provided that the time record available for analysis is long enough to provide sufficiently many quiet samples (see Sec. G below). From the above we conclude that, for purposes of SOAE detection, 16k-FFT analysis using our algorithm performs at least as well as the other procedures cited, and the savings in computation time and memory when using the smaller FFT window may prove to be an advantage in some situations.

The number of emissions found per emitting ear shown in Table I varies somewhat across the three studies, and differences in the rules for SOAE acceptance probably account for some of that variation. Talmadge et al. (1993) reported all spectral peaks meeting their detection criterion, irrespective of minimum SOAE frequency spacing or frequency relationships with other SOAEs. Zhang and Penner (1998) apparently discarded emissions which could be distortion products of other emissions. In the present study, we did not omit possible distortion products, but we did eliminate those spectral peaks that were within 0.1 octave of another peak of greater magnitude (about 0.44-mm separation along the basilar membrane using the Greenwood map; Greenwood, 1990, 1991). When we relaxed this requirement (using 16k FFTs), to a minimum frequency spacing of only 3% (0.043 octave), the mean number of emissions increased to 7.3 per emitting ear (median: 6.5) for females, and 4.3 (median: 3.0) for males, in close agreement with Talmadge et al. (1993). Relaxing the minimum spacing to 3% provided for elimination of any secondary spectral peaks riding on the flanks of a stronger SOAE, but allowed for inclusion of all “freestanding” peaks strong enough to satisfy the decision criterion.

F. Elimination of noisy samples

Detection performance of our algorithm was slightly poorer when analyzing 32k-FFT averages than when analyzing 16k FFTs (see Table I), and possible reasons for this difference were mentioned in the previous section. Halving the number of FFTs averaged produced the expected doubling of the variability of the unflagged spectral peaks about the smoothed baseline, probably obscuring some of the smallest SOAE peaks. However, the elevation in noise floor
observed in the low-frequency regions of many spectra suggests that the pool of sufficiently quiet recording segments was exhausted for some ears. This situation could possibly be improved if a more sophisticated noisy-sample rejection scheme were adopted, such as that of Zhang and Penner (1998) or Talmadge et al. (1993). The Talmadge et al. (1993) procedure is based on peaks in the original time sample, and uses a different rejection criterion for each of three spectral regions, resulting in a ‘composite spectrum,’ where the lowest frequency range is based on the fewest samples averaged (the strictest criterion, the quietest 25%), and the other two higher-frequency ranges are based on successively more samples. The quietest 150 samples (for 16k FFTs) described above corresponds very closely to 25% of the available samples. For the purposes of the McFadden and Pasanen (1999) study, it was not considered appropriate to treat different spectral regions differently because evidence already exists that SOAE prevalence may vary across spectral regions for different populations (Whitehead et al., 1993). Zhang and Penner (1998) used an iterative procedure based on both number of zero crossings and the expected Gaussian distribution of the energy in short segments in their recordings. It is possible that their procedure might produce an improvement over ours, but again, for the purposes of our study, it was considered important to have the averaged spectrum for every subject based on the same number of samples; theirs does not appear to do so.

The requirement of finding windows of 32k contiguous sample points (almost 1.5-s duration) that are sufficiently quiet will always pose a serious difficulty, but an improvement could be achieved by the placement of the middle part of sample windows within those periods in the time record where there is relatively lower noise, and allowing noisier parts to fall in the first or last quarter, where they would be attenuated by the window function. The noise criterion we used for rejection of noisy samples was based on the rms value calculated for each time sample, and this is probably adequate when the samples are reasonably short (0.74 s in our 16k analysis) because the noise generated by a human subject is not likely to occur in extremely short bursts. When the sample duration is much longer, however, the rms value may underestimate the degree of contamination of that sample. Based on our experience, it appears that a recording sample somewhat longer than 2 min would be the simplest solution for those wanting to use 32k FFTs.

Our data were collected mainly from cooperative college-age subjects, and therefore were about as quiet as could be expected. Other subject populations might warrant recorded samples totaling more than 2 min. Our procedure was to record continuously for only 30 s, with a break of 15 s or so between recordings. The subject was told via intercom when a 30-s segment was finished, given about a 10-s break, and then a 5-s warning when a new recording was about to begin (although the new recording could be delayed as long as necessary if the subject was not ready when the warning was given). We believe this procedure reduced the noisy episodes in the recordings considerably. For other subject populations, such as children, shorter recording segments might be advisable.

III. SUMMARY

We have described here an automated procedure for detection of SOAEs. The intention was, first, to provide a more complete description of the procedure used by McFadden and Pasanen (1999) [initially reported by Pasanen and McFadden (1996)], and second, to provide sufficient detail to allow implementation of the procedure by others. Adoption of a standard SOAE-detection procedure, whether ours or not, would be quite helpful in allowing comparisons of data from different laboratories, and would enable collection of SOAE data by nonexpert researchers or clinicians. The procedure described here has several advantages:

(i) It has high sensitivity combined with a low false-acceptance rate.
(ii) It allows direct manipulation of SOAE acceptance criterion over a meaningful range.
(iii) It is computationally simple, and rapid enough to be accomplished while a subject is still in place, allowing for collection of replacement data when the need arises.
(iv) The minimum height needed for a peak to be characterized as an SOAE at a specific level of statistical confidence is set according to the quality of the data collected from that ear.
(v) The smoothing process is tailored to reflect the bandwidth and frequency spacing unique to SOAEs.
(vi) Although there are a number of parameter values to be set, the procedure is quite robust to those choices. Nevertheless, users interested in optimizing the algorithm’s performance for specific conditions can change parameter values easily, and exploration of the parameter space to optimize the algorithm’s performance can be done in a simple and straightforward manner.

The entire algorithm and other data analyses were programmed using the LabVIEW® graphical programming language (National Instruments), but the computations are conceptually simple, and could be easily implemented in other programming environments.

ACKNOWLEDGMENTS

This work was supported by a research grant from the National Institute on Deafness and other Communication Disorders (No. NIDCD 00153). The authors thank C. A. Champlin, L. K. Cormack, and J. C. Loehlin for their helpful comments on an earlier version of the manuscript. Interested readers are invited to contact the authors directly to obtain a LabVIEW® version of the SOAE-detection software.

1The bandwidth of an SOAE has typically been taken as the width at half-power of a Lorentzian peak fitted to the SOAE (e.g., Talmadge et al., 1993; Van Dijk and Wit, 1998), but this estimate does not necessarily reflect accurately the width of the SOAE near the base of the peak. The SOAEs of the group of heterosexual female subjects in McFadden and Pasanen (1999) were fitted with Lorentzian functions to estimate bandwidth. Narrow-band SOAEs (5.4 Hz or less) were ignored because the spectral frequency resolution was inadequate for accurate bandwidth estimates. The correlation of SOAE frequency with bandwidth at half-power for the remaining SOAEs was 0.40.
To select appropriate bandwidths for simulation, the bandwidth of each
5
The actual value of criterion for accepting a peak as an SOAE was obtained
4
The 19-point median filter replaces each point of the spectrum with the
3
The 60 spectra described in Sec. II C.
0-2
Technically, level is the correct term for describing values expressed in
power bandwidth of the fitted function as the estimate of SOAE bandwidth. ~
points contributing to the total power described above, and using the half-
SOAE was estimated by fitting a Lorentzian function to that set of spectral
~
point of the spectrum with the arithmetic mean of the 51 original bin values
median value of the 19 original bin values centered on the selected point.

The 51-point rectangular smoothing window is a filter which replaces each
point of the spectrum with the spectrum of the fitted function as the estimate of SOAE bandwidth.

The mean plus 5 s.d. values ranged across groups from about 1.33
to 1.40 in the left ear, and 1.32 to 1.41 in the right, implying that the
distributions of nonflagged peaks were very similar across groups.

To select appropriate bandwidths for simulation, the bandwidth of each
SOAE was estimated by fitting a Lorentzian function to that set of spectral
points contributing to the total power described above, and using the half-
power bandwidth of the fitted function as the estimate of SOAE bandwidth.

As Talmadge et al. (1993) pointed out, such a bandwidth estimate is only appropriate if it is at least four times the frequency resolution of the spectrum. For the present analysis, with a frequency resolution of 1.35 Hz, it follows that bandwidth estimates below about 5.4 Hz are not accurate. However, for evaluating the detection performance of the algorithm, this bandwidth inaccuracy is not a problem because such narrow-band SOAEs are typically quite strong and easily detected. Our data (see Sec. II C) and the data of Talmadge et al. (1993) indicate that weak emissions (within 10 dB of the noise floor) typically have bandwidths substantially greater than 5 or 6 Hz.