**Objective methods to measure vestibular evoked myogenic potential responses saccular tuning curves**

Faten Obeidat and Steven Bell

Abstract

**Objective**: To detect cervical vestibular evoked myogenic potential (cVEMP) responses using objective statistical approaches and to apply this approach to estimate saccular frequency-tuning curves in volunteers and Ménière’s disease (MD) patients.

**Design**: Estimates of cVEMP threshold were carried out by 3 expert raters at 500 Hz and compared to objective threshold estimates (using Hotelling’s T2 [HT2] and Fsp). Saccular tuning curves were objectively estimated.

**Study sample**: Objective and subjective estimates of cVEMP response thresholds were compared for 13 normal hearing adults. Objective measurement of saccular tuning curves was explored in 20 healthy adults and 15 patients with MD.

**Results**: Significant variability was seen between subjective estimates of cVEMP thresholds. Objective analysis with the HT2 test was more sensitive than 2 of 3 experts in detecting responses. The measurement time of cVEMP was considerably reduced with the HT2 test. Objective saccular tuning curves in volunteers showed strongest responses at 500 Hz. A flatter tuning curve was seen for MD patients.

**Conclusions**: There is significant variability in subjective estimations of cVEMP thresholds. Objective analysis methods are more sensitive than subjective analysis, can detect responses rapidly and have potential to reduce variability in threshold estimates, hence they appear well suited to measure cVEMP tuning curves.

# 1. Introduction

In 1935, Von Bekesy was the first to report that high intensity sound stimuli elicited head movements, which he suggested was the result of vestibular stimulation (Welgampola & Colebatch, 2005). Bickford et al. (1964) reported an inion response to loud sounds, which was present in deaf patients with normal vestibular end organ function but absent in deaf patients with vestibular dysfunction. This finding indicated that the response originates in the vestibular system and is not affected by cochlear dysfunction; thus, it could be performed in patients with deafness (Jacobson & Shepard, 2008). More recently, a myogenic response was recorded by Colebatch, Halmagyi et al. using surface electrodes over the contracted sternocleidomastoid (SCM) muscle following vestibular stimulation by high-level air conduction (AC) click stimuli in normal human subjects (Colebatch and Halmagyi, 1992; Colebatch et al., 1994). These responses were labelled click-evoked vestibulo-collic reflex (VCR), which are thought to primarily measure saccular function (Rosengren & Kingma, 2013). Other researchers have described these responses as Vestibular evoked myogenic potentials (VEMPs), as they are myogenic potentials elicited by the stimulation of the vestibular end organs (Zhou & Cox, 2004).

VEMPs are short-latency myogenic potentials elicited by stimulating the ear with high-level air-conducted sound (ACS), bone-conducted vibration (BCV), forehead taps or electrical stimulation and can be recorded by placing surface electrodes over muscles (Rosengren et al., 2010; Young, 2013). cVEMP responses to AC stimulation are primarily considered to be saccular in origin and it has been proposed that cVEMP response frequency will be altered in patients with Ménière’s disease (MD). Previous studies on healthy subjects found that cVEMP exhibits a frequency-tuning curve with best response (frequency tuning) at 500 Hz (Akin et al., 2003; Park et al., 2010; Piker et al., 2013; Rauch et al., 2004; Timmer et al., 2006). Rauch et al. reported that subjects with MD exhibited a different cVEMP tuning pattern from healthy subjects. In their study, subjects with unilateral MD exhibited a higher threshold at all frequencies with greater response at 1000 Hz than 500 Hz on the affected side, while the healthy subjects showed best cVEMP response around 500 Hz. They suggested that an altered cVEMP tuning pattern could be a marker of saccular endolymphatic hydrops. The unaffected ear of MD subjects also showed alterations in cVEMP tuning and a threshold shift comparable with normal subjects; however, the reason was not identified. It is believed that these may be early signs of the development of bilateral MD.

Although the cervical VEMP (or cVEMP) has been widely used in clinical practice as an objective measure of the VCR, the interpretation of data still generally relies upon subjective (visual) interpretations, and so results are highly dependent upon clinical expertise. Different criteria have been used in the literature to visually judge the presence of a cVEMP response. Even when specified criteria are used for visual interpretation of the presence or absence of a cVEMP response, subjective judgment is typically required to decide if the criteria are met.

To reduce cVEMP measurement noise, a number of epochs are averaged (typically around 150). The visual evaluation of the averaged response is problematic when the signal -to- noise ratio (SNR) is poor, due to a small response relative to the physiological background noise (Don et al., 1984). When the SNR is poor, the waveform morphology of the response is affected by noise, and this causes difficulty in identifying the presence of the cVEMP response. Furthermore, replications of significant peak and trough components can be unreliable due to variations in the background noise between testing sessions (Don et al., 1984). For cVEMP tests, using a fixed number of epochs (averages) can never guarantee the repeatability of the waveforms, as SNR and muscle tension often vary between testing runs. More repetitions of the waveform should be conducted for low SNR responses than for high SNR responses to obtain a clear response. However, performing a large number of recordings of the cVEMP may not be possible due to the effect of muscle fatigue (recording 150 epochs at 10 Hz stimulation rate takes 30s, which is around the limit that a patient may maintain good neck tension for). Hence, visual identification is subjective, and considerable variations may occur between experienced audiologists in reporting the presence of a cVEMP response for a given stimulus. In a clinical setting, caution should be exercised when visually identifying the presence and absence of the cVEMP response.

The problem of subjective identification of response thresholds has been recognised for other evoked responses. For example, Vidler and Parker (2004) compared the threshold estimates of 16 professionals on Auditory Brainstem Response (ABR) data. The difference between threshold estimates was in some cases more than 40 dB. However, to date, this variability has not been well studied for cVEMP thresholds, and the statistical methods that have been used for detection of other evoked responses have not been well tested for the measurement of VEMP tuning curves.

Several statistical approaches have been explored and used for automated detection of evoked potentials such as the ABR, Auditory Steady State Responses and the Auditory Late Response. A review of approaches is given in Chesnaye et al. (2018). In the case of ABR, Elberling and Don (1984) established one of the first methods and proposed the Fsp (F value at a single point) statistic as an objective estimate of response quality. This approach (and a variant termed the Fmp) is now available in several commercially available clinical measurement systems. Threshold values for Fsp can be used to indicate when a response is significantly different from noise. It is possible to convert Fsp values to p values, indicating the significance of a response, but only if the degrees of freedom of the signals being measured are known. Elberling and Don (1984) made such an estimate of degrees of freedom for ABR, but the degrees of freedom of VEMP data have not been well explored. One way around this problem of unknown degrees of freedom is to use bootstrap analysis, which is based on random resampling of the data to estimate the null distribution of a Fsp. From this distribution, the p value for a given Fsp value can be established (Lv et al., 2007).

Another objective method that has been well used for auditory evoked response detection is the Hotelling’s T2 test (HT2) (Hotelling, 1931), which is a multivariate extension to the student’s t-test and can be used to evaluate whether the variables’ means are significantly different from hypothesized values. A number of authors have applied the Hotelling’s T2 approach for automated response detection of slow Cortical Auditory Evoked Potentials (CAEPs) (Carter et al., 2010; Chang et al., 2012; Golding et al., 2009; Van Dun et al., 2015; Van Dun et al., 2012). Recently, Chesnaye et al. (2018) investigated and compared the objective detection of ABR response using HT2 statistics and several objective approaches including Fsp and the magnitude squared coherence (amongst others) and established that HT2 was the most sensitive approach.

The goal of this article is to explore whether objective analysis methods can be used to improve the measurement of cVEMP thresholds and hence reduce reliance on subjective interpretation approaches. The objectives were 1) to compare the sensitivity (threshold estimates) of objective detection methods—the Fsp bootstrapping approach and HT2 test—to results obtained through subjective inspection by experienced observers, 2) to compare detection times for the two statistical objective measures when detecting cVEMP responses, and 3) explore applications of the statistical methods for measuring the saccular tuning curve in patients with MD.

This paper presents the experimental findings of two separate studies: 1) a comparison of objective and subjective estimate of cVEMP response thresholds in normal subjects at 500 Hz and 2) objective measurement of saccular tuning curves in normal subjects and patients with MD.

# 2. Methods

## 2. 1 Participants

Experiment 1: For the comparison of objective and subjective estimate of cVEMP threshold, 13 subjects (7 females and 6 males) with normal hearing and balance function in the age range 22-48 years participated in this study. All subjects gave informed consent to participate and had pure-tone thresholds of around or better than 20 dB HL. Subjects who had hearing problems, balance problems, neck/back stiffness or pain were excluded from the study. Responses were recorded ipsilaterally from the left SCM muscle for normal subjects. The experimental protocol for this study was approved by the Human Experiment Safety and Ethics Committee of the University of Southampton.

Experiment 2: For objective measure of saccular tuning curves, 20 normal subjects (11 women and 9 men) aged between 20 and 55 years were allocated to a control group. All subjects gave informed consent to participate. None of the control group had a history of hearing or balance problems. All subjects had pure-tone thresholds of equal to or better than 20 dB HL. cVEMPs were recorded ipsilaterally from either the right or left SCM muscle of normal subjects (chosen randomly) to give a total of 20 ears tests. The clinical test group with MD consistent of 15 patients (nine women and six men), with a mean age of 40 years (range 18-60), who had been diagnosed with bilateral definite MD according to the American Academy of Otolaryngology–Head and Neck Surgery diagnostic criteria (1995 AAO-HNS). Both ears were tested (30 ears). Based on these guidelines, both ears (n=30) of the 15 patients were symptomatic, but unexpectedly left ears were more affected by the disease than the right ears. Testing was conducted at the Middle East Hearing and Balance Centre in Jordan-Amman. Patients had stronger subjective symptoms (tinnitus and ear fullness) and more severe hearing deterioration in the left ears, while they had either tinnitus or aural fullness and less severe hearing loss in the right ears. In line with the AAO-HNS guidelines, AC pure tone averages at 0.5, 1, 2, and 3 KHz frequencies were used to classify the stage of the hearing level. The worst audiometric results during the 6 months period prior to treatment was used for stage classification. Based on these guidelines, left ears were classified as stage 3, while right ears were classified as stage 1 and 2. As the progression of the disease in each ear was different, the left and right Ménière’s ears were sub classified into ‘most’ and ‘least’ affected ears respectively according to the severity of the subjective symptoms (ear fullness, tinnitus) and the stage of hearing level. Patients who were determined to have middle or external ear pathology, neck/back stiffness or pain, or allergy to alcohol swabs were excluded from the study. In addition, patients who underwent surgical treatment were also excluded. Ethics Committees of the University of Southampton (experiments 1 and 2) and the Middle East Hearing and Balance clinic in Jordan-Amman (experiment 2) approved the experimental protocol for the studies.

## 2.2 Stimuli

For experiment 1, 500 Hz 1:2:1 (one cycle rise/fall and two cycles plateau) tone-bursts stimuli were presented using insert earphones (Etymotic ER-3A). The stimulus level was decreased in 3 dB steps from 109 dB A (A-weighted sound level) until the level was significantly below that where a response could be seen (around 86 dB A typically). Each recording consisted of 150 repeats of an 8 ms short tone-burst. A repetition rate of 10 Hz was fixed for all measurements, so total duration for each recording was 15 s. The rate of 10 Hz was found to be the optimal trade-off between recording time and response detection for the majority of subjects in our previous study (Obeidat and Bell, 2017). The order of presentation of stimulus intensities was randomised among subjects.

For experiment 2, tone-bursts (250, 375, 500, 750, and 1000 Hz) stimuli with a one-cycle rise and fall and a two-cycle plateau were presented using insert earphones (Etymotic ER-3A). The number of cycles in the stimulus was kept constant, so the spectral spread as a function of centre frequency was constant, hence stimulus duration changed with frequency from 16 ms at 250 Hz to 4 ms at 1000 Hz. Each recording was 150 epochs. The rate of stimulus presentation of 10 Hz was fixed for all measurements, so total duration for each recording was 15 s. cVEMP thresholds were determined at five frequencies using 3 dB steps from 106 to 85 dB LAS (A-weighted sound level with a slow time constant) (125 to 105 LLpK (peak SPL)). We also measured levels in dB LAeq (A weighted equivalent continuous sound level) which gave almost identical results to LAS. The order of presentation of stimulus intensities was randomised among subjects. To avoid muscular fatigue, a one-minute break was given to each subject after each recording.

For both studies, the calibration of the stimuli was carried out through a Brüel and Kjar (B&K) type 2260 sound level meter (SLM), attached to an occluded ear-canal simulator type 4157 (IEC-711 coupler). The equipment used in this study to deliver the stimuli for cVEMP measurement was Cambridge Electronic Device’s CED 1401 data acquisition system and CED ‘signal’ software (http://ced.co.uk/). A sampling rate (input and output) of 10 KHz was used. The output from the Digital to Analogue Converters (DAC) port was routed through a headphone amplifier (OBH-21) to control the intensity of the stimulus. Amplification of the signals was performed using an isolated amplifier (CED 1902) with a 1-3000 Hz bandpass filter with 1000 times gain.

## 2.3 cVEMP recording

VEMPs from the SCM muscle were recorded ipsilaterally while subjects were seated upright on a chair with their chin turned over the contralateral shoulder to tense the SCM muscle. The electromyographic (EMG) activity of the SCM muscle was recorded using surface electrodes placed on the muscle: active on the belly of the ipsilateral SCM muscle, and reference on the upper sternum of the test side. A ground electrode was placed on lower forehead. The impedance of the electrodes was kept below 10 KΩ. The EMG activity of the SCM muscle was visually monitored on an oscilloscope and kept between 80 and 100 mV.

## 2.4 Detection methods

### Visual inspection of the response

cVEMP thresholds were separately estimated by three experienced audiologists (A, B & C), who were blinded to the experimental conditions. In our work, cVEMP responses were defined as present if the cVEMP was a reproducible biphasic waveform with a positive (p1) peak followed by a negative (n1) peak, the inter-amplitude (p1-n1) of the response was between 20 and 150 μV, based on the criteria presented at the Balance Interest Group of the British Society of Audiology (2012), the waveform should be larger than the rest of the response in the overall average and the acceptable latency range of p1 was 13.9-19.2 ms and n1 was 22.9-30.3 ms. These latency values were determined from a recent study (Blakley and Wong, 2015) of cVEMP evoked by 500 Hz tone-bursts in 48 adults (23-64 years) with no history of hearing and balance problems. If any of these criteria were not met, then the cVEMP response was judged to be absent. An example of a typical cVEMP response to ACS is shown in Figure 1 below. Here all the criteria for response presence have been met, so the response is judged to be present.

-INSERT FIGURE 1-

### Description of the methods

The two statistical methods (HT2 and Fsp bootstrapping, described below) were applied to the responses recorded from otologically normal subjects and cVEMP thresholds were then identified for each statistical method. In this work, the cVEMP threshold was defined as the lowest stimulus level (in dB A) at which a significant response (p ≤ 0.05) is obtained, with a significant response for all higher stimulus levels (see example in Table 1). Each statistical parameter was calculated for the analysis time window 10-30 ms following stimulus onset. These thresholds were compared to those subjectively estimated by three experienced observers.

Table 1 shows the p-values at all intensity levels for the two objective methods used for measuring the cVEMP threshold for one participant. In this example, the minimum stimulus intensity for detection of the response was 103 and 100 for the bootstrap approach and HT2 test, respectively. For all higher levels, the p-values showed a significant response (p ≤ 0.05).

-INSERT TABLE 1-

1. *Bootstrap analysis of Fsp values*

For each cVEMP recording, the Fsp of the coherent average was first calculated. This represents the ratio of the variance of the averaged response to the variance of noise estimated across epochs at a single point (Don et al., 1984). As the degrees of freedom of the VEMP data were not know, it was not possible to convert Fsp values to p values directly, so a bootstrap approach was then used to determine whether the value of Fsp for each recording indicated a significant response (Lv et al., 2007). The bootstrap test was applied by selecting random starting points throughout the original recorded signal to generate an ensemble of non-coherent epochs with lengths similar to the raw signal. Non-coherent epochs are epochs with random starting points that are not aligned with the stimulus timing. The new array of non-coherent epochs was averaged to obtain a non-coherent average, for which the Fsp value was again calculated. This bootstrap process was repeated 500 times, and hence a bootstrap distribution of non-coherent Fsp values was obtained. This gives an estimate of the distribution of Fsp values that would be obtained if there is no significant response (under the null hypothesis). Finally, the Fsp of the coherent average is compared to the distribution of non-coherent Fsp values to calculate a p value. A p value was determined from the proportion of bootstrap values that exceeded the Fsp of the coherent average. For example if only 1% of the non-coherent Fsp values are higher than the Fsp of the coherent average, there is only a 1% change of obtaining such a result by chance, so p=0.01. A cut-off of p<0.05 was used to determine the presence of a response. The bootstrap values of Fsp were obtained for all recordings for all subjects using MATLAB software.

1. *Hotelling’s T2 test (HT2)*

The one-sample Hotelling's T2 test is a multivariate extension to the students t-test, and can be used to test whether the means of N features (in our case N average voltages over time regions of the response - see next paragraph) are significantly different from N hypothesized values. In the present work, it is assumed that the expected values of the features (the N hypothesized values) are zero. The statistic itself is a weighted sum of the N feature means where the weights are determined by the variances and covariances of the features. These weights are furthermore optimal in the sense that they maximize the resulting T2 value, which maximizes the sensitivity of the test when using the N features in question (Simaika, 1941). The weights have the additional property of normalizing the N means, which allows features with different scales and units to be combined appropriately. For more details of the methods, see Chesnaye et al. (2018).

Before applying the HT2 approach, each recorded epoch (n=150) with a duration of 20 ms was reduced to a number of average voltages, with each average having been taken within a time window covering a particular latency range. For example, with 5 windows used, the 5 time windows covered the range from 10 to 30 ms, with each feature being 4 ms wide. We explored the choice of different numbers of time windows from 2 to 20 on threshold estimates. A repeated-measures ANOVA with Greenhouse-Geisser corrected values showed a statistically significant effect of the number of features of the HT2 test on the cVEMP threshold (F (2.144, 25.731) =9.548, P < 0.05). However on paired comparison, only the choice of 2 features raised thresholds compared to the other approaches. 5 features appeared to be the best trade-off between sensitivity and analysis complexity and this number was used in the results section below.

The one-sample HT2 test is a multivariate extension of the ordinary one-sample t-test; it tests a null hypothesis, which has been defined in this work, as the expected averaged value in every time window was zero. Response detection was based on a cut-off p-value of 0.05 obtained from a one-sample HT2 test (a cut-off of p<0.05 results in a false positive rate of 5 %).

*Detection times for statistical method:*

The required time to detect a cVEMP response was measured by finding the number of stimuli (expressed in seconds) required for the p-value to drop and remain below the 0.05 threshold for the remainder of the test. The significance of a statistic was evaluated by taking the first 10 epochs and obtaining the p-value for the HT2 test and the Fsp bootstrapping method, then taking the first 20 epochs and generating the p-value for both objective methods, and so on.

# 3. Results

## 3.1 Subjective inspections of cVEMP threshold

Figure 2 shows the cVEMP thresholds estimated by the three experienced audiologists (A, B & C) for 13 subjects. The inter-observer reliability for detecting a cVEMP threshold was assessed using a Fleiss’ kappa test. Results showed that the proportion of agreement given by Fleiss’ kappa between the three observers was 0.5314. For the reliability of judges to be regarded as high, Fleiss’ Kappa values should be ≥ 0.90, according to Arnold. (1985). Thus, the reliability between observers was not high and this indicates disagreement in the identification of cVEMP thresholds, which is consistent with effects seen for ABR threshold estimation (Vidler and Parker 2004).

-INSERT FIGURE 2-

## 3.2 Comparison of subjective and objective estimates of threshold

Figure 3 shows the mean cVEMP threshold for 13 subjects estimated subjectively by three observers (A, B, & C) and by using the objective methods (HT2 and bootstrap for a significant level of α =5 %) for the same number of averages, n=150. A non-parametric Friedman test showed a highly significant difference between the estimates from the methods [χ2 (4) = 15.118, p <0.001]. Wilcoxon signed-rank testing was conducted to further explore where these differences lay and showed that the HT2 test was sensitive in finding cVEMP thresholds at significantly lower stimulus levels than found by observers A & B (p<0.05). The bootstrap method also detected the cVEMP response at significantly lower thresholds compared to one of the observers (B). Observer C detected the cVEMP threshold at a significantly lower stimulus level compared to observer B. HT2 produced the numerically lowest value of threshold, with a significantly lower threshold than two of the three raters.

-INSERT FIGURE 3-

## 3.3 Comparison of detection time

The average detection time (in seconds) was shorter for the HT2 test compared to the Fsp bootstrapping method. At 109 and 106 dB A, 50 and 110 stimuli, respectively, were required to obtain 100% detection using the HT2 test. In contrast, the bootstrap Fsp method required 120 stimuli at 109 dB A to detect responses in all thirteen subjects, and more than 150 epochs were required to achieve 100 % detection at 106 dB A. On average, the HT2 test can detect cVEMP responses faster than the bootstrap approach, using less than five seconds at 109 and 106 dB A.

## 3.4 Objective estimates of VEMP tuning curve

Figure 4 shows cVEMP thresholds (in dB LAS) for normal subjects’ ears (n=20), and the ‘most’ (n=15) and ‘least’ (n=15) affected ears of bilaterally affected patients with MD, which were objectively detected by the HT2 test for a significance level of α =5 % (150 sweeps) as a function of stimulus frequency. Across normal subjects (control group), there was some variation in individual thresholds and the pattern of the saccular tuning curve varied a little, but in general it was U shaped with the best frequency response between 375 and 500 Hz, on average. The data was not normally distributed. For the control group, a Friedman test showed a statistically significant difference in cVEMP thresholds between frequencies (P < 0.001). The Wilcoxon signed-rank testing revealed that the frequency of 500 Hz produced a significantly lower cVEMP threshold than all of the other frequencies (p < 0.05), except 375 Hz.

The Mann-Whitney tests were used to compare pairs of results from different groups (control and MD). The Wilcoxon tests compared results in the same group (e.g. within control). Compared to normal ears (n=20), the Mann-Whitney tests showed that the ‘most’ affected MD ears (n=15) had a significantly higher threshold for tone-bursts of 375-1000 Hz (p-values=0.001-0.006). The saccular tuning curve for the ‘most’ affected ears with MD no longer clearly indicated its best frequency response at 500 Hz, showing that the 500 Hz frequency tuning had either been lost or shifted. The Mann-Whitney test also showed that the MD ‘least’ affected ears had a significantly higher threshold for tone-bursts of 375 and 500 Hz (p-values= 0.01and 0.03), compared to normal subjects. For MD ‘most’ and ‘least’ affected ears, Wilcoxon signed-rank testing showed that MD ‘most’ affected ears produced significantly higher cVEMP threshold at 500 and 750 Hz tone-bursts (p-values= 0.007 and 0.012), compared to MD ‘least’ affected ears. The MD ‘least’ affected ears had less shift in the cVEMP threshold than the MD ‘most’ affected ears. The frequency tuning in the MD ‘least’ affected ears was not changed, suggesting that the 500 Hz frequency tuning had not been either lost or shifted. The largest difference between the groups was seen at 500 Hz. It appears that 500 Hz may be the best frequency to differentiate normal ears from MD affected ears based on threshold alone.

-INSERT FIGURE 4-

Figure 5 shows the peak-to-peak amplitudes (p13-n23) of cVEMP at 103 dB LAS as a function of tone-burst stimuli for normal subjects’ ears (n=20), MD ‘most’ affected (n=15) ears and MD ‘least’ affected (n=15) ears. The amplitude values of the cVEMP recordings that showed a significant p-value on the HT2 test at 103 dB LAS were subjectively measured from plots of response waveforms. Across normal subjects (control group), a Friedman test showed that there was a significant effect of frequency on the peak-to-peak amplitude [χ2 (4) = 29.519, p <0.000]. Wilcoxon signed-rank testing was conducted for multiple comparisons and revealed that the frequency of 250 Hz produced significantly lower cVEMP amplitude than all of the other frequencies.

Compared with normal subjects, the Mann-Whitney test revealed that the MD ‘most’ and ‘least’ affected ears showed significantly lower amplitudes for tone-bursts (375-1000 Hz) (p values=0.001-0.008), with more differences in the MD ‘most’ affected ears than in the ‘least’ affected ears. For MD ‘most’ and ‘least’ affected ears, the Wilcoxon signed-rank testing showed that the frequency of 250 Hz produced significantly lower cVEMP amplitude than the 500, 750, and 1000 Hz frequencies. On average, normal ears showed the largest cVEMP amplitude at 500 Hz tone-burst; however, differences between 500, 375, 750 and 1000 Hz were not significant, possibly due to limited statistical power to detect this. The MD ‘most’ affected ears showed loss or shift of frequency tuning in comparison with normal subjects.

Both amplitude and threshold measures showed cVEMP frequency tuning, with the strongest response at 500 Hz. Similarly, both measures showed impairment of both ears for bilaterally MD patients, with more alteration in the most-affected ear. However, cVEMP amplitudes showed more variance than threshold measurements, and thus threshold appears to be the best measure to use to discriminate the groups.

-INSERT FIGURE 5-

# 4. Discussion

Although the cVEMP is considered an objective measurement of saccular function, the objectivity is diminished by visual analysis of the responses. Subjective analysis of response presence remains the most common approach for cVEMP threshold estimates. The motivation behind this work was to detect cVEMP responses objectively in order to eliminate the high variability of visual judgements and to increase the sensitivity of response detection. Our work established that cVEMP tuning curves can be estimated objectively using statistical approaches, which is a novel contribution of the present study.

The objective estimate of response presence with the HT2 test was the most sensitive method, identifying cVEMP responses at lower stimulus levels than either visual detection or the bootstrap Fsp method. There was a considerable level of disagreement between the three experienced audiologists (A, B and C) in their decisions regarding identification of response thresholds of cVEMP, as reflected by a moderate reliability kappa value. This is consistent with the subjective variability that has been found between different experienced professionals in subjective threshold determination of ABR data (Vidler and Parker, 2004). The advantage of using objective methods over subjective identification of cVEMP responses is that they are not affected by the subjectivity of visual judgement. However, for clinical use, we envisage that some subjective inspection of waveforms is still helpful to check if the morphology of the response is normal or not, as this could have clinical implications. This could be done for only the highest stimulation levels, where the VEMP response should be clear.

On average, the HT2 test can detect cVEMP responses quickly, in less than 5 seconds at 109 and 106 dB A (when the cVEMP is clear). This can considerably reduce the duration of the test and consequently reduce the fatigue of the subject by reducing the time needed to maintain neck tension. The bootstrap approach performs significantly worse than HT2 in terms of test duration, but it is still more sensitive than visual inspection of response presence. Therefore, for people with normal vestibular function, using the HT2 test at these stimulus intensities (109 and 106 dB A) can considerably reduce the duration of the measurement. This requires fewer sweeps and can thus lessen the subject’s fatigue from having to maintain neck tension. The results of the present study are consistent with a previous study on ABR by Chesnaye et al. (2018), who found an advantage in terms of sensitivity and detection time for the HT2 test over other statistical approaches when detecting ABRs. With respect to how the number of features of the HT2 test affects the cVEMP thresholds, in the current work using the HT2 test with five features appeared to be a good trade-off between sensitivity and analysis complexity.

Objective estimate of the saccular tuning curve showed cVEMP frequency tuning with lowest thresholds at 500 Hz on average, and this is broadly consistent with the subjective estimate of response tuning found in previous studies (Akin et al., 2003; Park et al., 2010; Piker et al., 2013; Rauch et al., 2004). It appears that objective detection methods can be used to measure saccular tuning curves (which has not been previously reported in the literature). cVEMP tuning curves that were affected by the presence of MD showed flatter tuning than those in the control group. While others reported that Ménière's ears exhibited a shift upward from 500 to 1000 Hz with a rise in the thresholds of all frequencies (e.g. Rauch et al., 2004; Timmer et al., 2006). In previous tuning studies, tuning curves were measured using quite high-level sounds, especially when testing frequencies away from the minimum of the saccular tuning curves. A concern of using such high stimulus levels is noise exposure, especially in patients with tinnitus. In our study, out of 30 ears, 20, 13, 8, 10, and 6 ears showed absence of cVEMP waves at 250, 375, 500, 750, and 1000 Hz frequencies, respectively. If we had increased the stimulus intensity then we may have been able to measure more responses at high frequencies. However this was not done in the current study due to concerns over acceptable noise exposures for subjects.

In the current paper we have focussed on measurement of cVEMP threshold. Such threshold measurements have been used in several previous studies as a diagnostic indicator: for example in Rauch et al. (2004) patients with unilateral Meniere's disease had significantly increased cVEMP thresholds compared to unaffected ears or normal ears. In (Streubel et al., 2001) patients with [superior canal dehiscence](https://www.sciencedirect.com/topics/medicine-and-dentistry/superior-canal-dehiscence) syndrome demonstrated that cVEMP threshold from the affected side was significantly lower than unaffected ears. However an alternative diagnostic approach is to measure threshold asymmetries between the right and left sides (similar to a canal paresis type of measurement). \*\*

In future work we intend to compare the sensitivity and specificity of objective saccular tuning curves with other objective approaches to detect MD, such as Electrocochleography.

Surprisingly, for all of the 15 patients in the present study, left ears were more affected than right ears. The probability of detecting patients with MD in all 15 left ears is highly unlikely to be a chance result: if MD were equally prevalent in left and right ears there is only a 1 in 16384 chance that the threshold would be being higher in the left ear of all subjects. At present, there is not good evidence for a side of lesion preferences to MD. We are unsure if this an unusual feature of the test population recruited in the study, although a retrospective study conducted by Devaiah et al. (2003) on eight patients with possible MD defined by AAO-HNS criteria found that left ears were symptomatic in seven of the eight patients, again suggesting higher prevalence on the left side.

# 5. Conclusions

There is significant variability between subjective estimates of VEMP threshold by experienced raters. Objective detection methods can reduce the variability of subjective estimation of cVEMP thresholds and offer a sensitive and rapid approach to measure saccular frequency-tuning curves. cVEMP-tuning curves provided by objective detection were broadly similar to those provided by visual estimation in previous studies. cVEMP tuning curves were affected by the presence of MD and showing flatter tuning than for the control group. MD most affected ears showed higher thresholds than MD least affected ears, which in turn showed higher thresholds than the control group.

**Declarations of interest**

None

**References**

Akin, F. W., Murnane, O. D., & Proffitt, T. M. 2003. The effects of click and tone-burst stimulus parameters on the vestibular evoked myogenic potential (VEMP). Journal of the American Academy of Audiology, 14(9), 500–509. <http://doi.org/10.3766/jaaa.14.9.5>.

Arnold, S. A. 1985. Objective versus visual detection of the auditory brain stem response. *Ear and hearing*, 6(3), 144-150. <http://doi.org/10.1097/00003446-198505000-00004>.

Blakley, B. W., & Wong, V. 2015. Normal values for cervical vestibular-evoked myogenic potentials. *Otology & Neurotology*, 36(6), 1069-1073. <http://doi.org/10.1097/MAO.0000000000000752>.

Bickford, R. G., Jacobson, J. L., & Cody, D. T. R. (1964). Nature of average evoked potentials to sound and other stimuli in man. *Annals of the New York Academy of Sciences*, *112*(1), 204-218.

<https://doi.org/10.1111/j.1749-6632.1964.tb26749.x>

British Society of Audiology Balance Interest Group. 2012. Information Document Performing Cervical Vestibular Evoked Myogenic Potential Measurements, (January 2012).https://www.thebsa.org.uk/wpcontent/uploads/2014/04/VEMP\_Guidance\_v1.1\_20121.pdf.

Carter, L., Golding, M., Dillon, H., & Seymour, J. 2010. The detection of infant cortical auditory evoked potentials (CAEPs) using statistical and visual detection techniques. *Journal of the American Academy of Audiology*, 21(5), 347-356. <https://doi.org/10.3766/jaaa.21.5.6>.

Chang, H. W., Dillon, H., Carter, L., Van Dun, B., & Young, S. T. 2012. The relationship between cortical auditory evoked potential (CAEP) detection and estimated audibility in infants with sensorineural hearing loss. *International Journal of Audiology*, 51(9), 663-670. <https://doi.org/10.3109/14992027.2012.690076>.

Chesnaye, M. A., Bell, S. L., Harte, J. M., & Simpson, D. M. 2018. Objective measures for detecting the auditory brainstem response: comparisons of specificity, sensitivity and detection time. *International journal of audiology*, 57(6), 468-478. <https://doi.org/10.1080/14992027.2018.1447697>.

Colebatch, J. G., & Halmagyi, G. M. (1992). Vestibular evoked potentials in human neck muscles before and after unilateral vestibular deafferentation. Neurology, 42(8), 1635-1635.

<https://doi.org/10.1212/WNL.42.8.1635>

Colebatch, J. G., Halmagyi, G. M., & Skuse, N. F. (1994). Myogenic potentials generated by a click-evoked vestibulocollic reflex. *Journal of Neurology, Neurosurgery, and Psychiatry*, *57*(2), 190–197.

<https://jnnp.bmj.com/content/jnnp/57/2/190.full.pdf>

Committee on Hearing and Equilibrium. 1995. Committee on Hearing and Equilibrium guidelines for the diagnosis and evaluation of therapy in Ménière’s disease. *Otolaryngology–Head and Neck Surgery*, 113(3), 181-185. <https://www.ncbi.nlm.nih.gov/pubmed/7675476>.

Curthoys, I. S., & Grant, J. W. (2015). How does high-frequency sound or vibration activate vestibular receptors? *Experimental brain research*, *233*(3), 691-699.

[https://link.springer.com/content/pdf/10.1007%2Fs00221-014-4192-6.pdf](https://link.springer.com/content/pdf/10.1007/s00221-014-4192-6.pdf)

Don, M., Elberling, C., & Waring, M. 1984. Objective detection of averaged auditory brainstem responses. *Scandinavian audiology*, 13(4), 219-228. <https://doi.org/10.3109/01050398409042130>.

Devaiah, A. K., Dawson, K. L., Ferraro, J. A., & Ator, G. A. 2003. Utility of area curve ratio electrocochleography in early Meniere disease. *Archives of Otolaryngology–Head & Neck Surgery*, *129*(5), 547-551. <http://doi.org/10.1001/archotol.129.5.547>.

 Elberling, C., & Don, M. 1984. Quality estimation of averaged auditory brainstem responses. *Scandinavian audiology*, 13(3), 187197. <https://doi.org/10.1016/j.medengphy.2006.03.001>.

Golding M., Dilon H., Seymour J., Carter L. 2009. The detection of adult cortical auditory evoked potentials (CAEPs) using an automated statistic and visual detection. *International Journal of Audiology*, 48:833-842. <https://doi.org/10.3109/14992020903140928>.

Hotelling H. 1931. The Generalization of Student’s Ratio. Ann. Math. Statist, 2 (3), 360-378. <https://doi.org/10.1214/aoms/1177732979>.

Jacobson, G.P., and Shepard, N.T. (Eds.). (2008). *Balance function assessment and management*. San Diego: Plural Publishing

Lin, M. Y., Timmer, F. C., Oriel, B. S., Zhou, G., Guinan, J. J., Kujawa, S. G., ... & Rauch, S. D. (2006). Vestibular evoked myogenic potentials (VEMP) can detect asymptomatic saccular hydrops. *The Laryngoscope*, *116*(6), 987-992.

 <https://doi.org/10.1097/01.mlg.0000216815.75512.03>

Lv, J., Simpson, D. M., & Bell, S. L. 2007. Objective detection of evoked potentials using a bootstrap technique. *Medical engineering & physics*, 29(2), 191-198. <https://doi.org/10.1016/j.medengphy.2006.03.001>.

Obeidat, F. S., & Bell, S. L. 2018. The effect of stimulation rate on cervical vestibular evoked myogenic potential quality. *Clinical Neurophysiology Practice*, 3, 24-27. <https://doi.org/10.1016/j.cnp.2017.11.001>.

Park, H. J., Lee, I. S., Shin, J. E., Lee, Y. J., & Park, M. S. 2010. Frequency-tuning characteristics of cervical and ocular vestibular evoked myogenic potentials induced by air-conducted tone-bursts. *Clinical Neurophysiology*, 121(1), 85–89.http://doi.org/10.1016/j.clinph.2009.10.003.

Piker, E. G., Jacobson, G. P., Burkard, R. F., McCaslin, D. L., & Hood, L. J. 2013. Effects of age on the tuning of the cVEMP and oVEMP. *Ear and Hearing*, 34(6), e65–73. <http://doi.org/10.1097/AUD.0b013e31828fc9f2>.

Rauch, S. D., Zhou, G., Kujawa, S. G., Guinan, J. J., & Herrmann, B. S. 2004. Vestibular evoked myogenic potentials show altered tuning in patients with Ménière’s disease. *Otology& Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, 25(3), 333–338. <http://doi.org/10.1097/00129492-200405000-00022>.

Rosengren, S. M., Welgampola, M. S., & Colebatch, J. G. 2010. Vestibular evoked myogenic potentials: past, present and future. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 121(5), 636–51. <http://doi.org/10.1016/j.clinph.2009.10.016>.

Simaika, J. B. 1941. On an optimum property of two important statistical tests. Biometrika, 32(1), 70-80.

<https://doi.org/10.1093/biomet/32.1.70>.

Streubel, Phillip D. Cremer, John P. Carey, Noah Weg, Lloyd B. Minor, S. O. (2001). Vestibular-evoked myogenic potentials in the diagnosis of superior canal dehiscence syndrome. *Acta Oto-Laryngologica*, *121*(545), 41-49.

 <https://doi.org/10.1080/000164801750388090>

Timmer, F. C. a, Zhou, G., Guinan, J. J., Kujawa, S. G., Herrmann, B. S., & Rauch, S. D. (2006). Vestibular evoked myogenic potential (VEMP) in patients with Ménière’s disease with drop attacks. *The Laryngoscope*, *116*(5), 776–779.

<https://doi.org/10.1097/01.mlg.0000205129.78600.27>

Todd, N. P., Rosengren, S. M., & Colebatch, J. G. 2009. A utricular origin of frequency tuning to low-frequency vibration in the human vestibular system?. *Neuroscience letters*, 451(3), 175-180.

 <https://doi.org/10.1016/j.neulet.2008.12.055>.

Van Dun, B., Dillon, H., & Seeto, M. 2015. Estimating hearing thresholds in hearing-impaired adults through objective detection of cortical auditory evoked potentials. *Journal of the American Academy of Audiology*, 26(4), 370-383. <https://doi.org/10.3766/jaaa.26.4.5>.

Van Dun, B., Carter, L., & Dillon, H. 2012. Sensitivity of cortical auditory evoked potential detection for hearing-impaired infants in response to short speech sounds. *Audiology Research*, 2(1). <https://doi.org/10.4081/audiores.2012.e13>.

Vidler, M., & Parker, D. 2004. Auditory brainstem response threshold estimation: subjective threshold estimation by experienced clinicians in a computer simulation of the clinical test. *International Journal of Audiology*, 43(7), 417-429. <https://doi.org/10.1080/14992020400050053>.

Welgampola, M. S., & Colebatch, J. G. (2005). Characteristics and clinical applications of vestibular-evoked myogenic potentials. *Neurology*, *64*(10), 1682-1688.

<https://doi.org/10.1212/01.WNL.0000161876.20552.AA>

Young, Y. H., Huang, T. W., & Cheng, P. W. 2003. Assessing the stage of Ménière’s disease using vestibular evoked myogenic potentials. *Archives of Otolaryngology–Head & Neck Surgery*, 129(8), 815-818. https://doi.org/ 10.1001/archotol.129.8.815.

Young, Y. H. 2013. Potential application of ocular and cervical vestibular‐evoked myogenic potentials in Ménière’s disease: A review. *The Laryngoscope*, 123(2), 484-491. <https://doi.org/10.1002/lary.23640>.

Zhou, G., & Cox, L. C. (2004). Vestibular evoked myogenic potentials: history and overview. *American Journal of Audiology*, *13*(2), 135-143.

<https://www.ncbi.nlm.nih.gov/pubmed/15903139>

**Tables**

Table 1. An example of p-values for the two objective methods at different stimulus levels for one participant. The marked p-value shows the cVEMP threshold for one subject.

**Figures**

Figure 1. Example of a typical biphasic clear cVEMP response to ACS, consisting of a positive-negative component that arises at 13-23 ms after stimulus presentation (P13-N23 or P1-N1).

Figure 2. cVEMP thresholds for 13 subjects with normal hearing and balance function, as determined through subjective detection by three experienced audiologists (A, B, & C). For each participant, the three bars show the cVEMP threshold estimated by observers A, B, and C respectively.

Figure 3. The mean cVEMP threshold for 13 subjects estimated subjectively by three observers (A, B & C) and objective methods (HT2 and Fsp bootstrapping) for a significance level of 5 %. Subjective and objective thresholds were determined for 150 sweeps. The error bars represent ±1 SE of the mean.

Figure ‎4. Mean cVEMP thresholds at different tone-burst stimuli for ‘most’ and ‘least’ affected ears of bilaterally affected patients with MD (n=15), and for normal subjects’ ears (n=20). The error bars represent ±1 SE of the mean. For the analysis, the threshold value of recordings that did not show a significant response on the HT2 test was set to 109 (3 dB step above the maximum stimulation level used in this study).

Figure 5. Mean peak-to-peak amplitude of cVEMP as a function of tone-burst frequencies for normal subjects (n=20), and for ‘most’ and ‘least’ affected ears of bilaterally affected patients with MD (n=30) at 103 dB LAS (122 LLpK (peak)). Error bars represent ±1 SE of the mean. The amplitude values of recordings that did not show a significant response on the HT2 test at 103 dB LAS were set to zero for the analysis of amplitude values.

Figure 1



Figure 2



Figure 3



Figure 4



Figure 5



Table 1

