inter.noise 2000
The 29th International Congress and Exhibition on Noise Control Engineering
27-30 August 2000, Nice, FRANCE

I-INCE Classification: 6.9

A COMPARISON OF TWO METHODS FOR THE EVALUATION OF PROMINENT DISCRETE TONES - PHASE 2

R. Hellweg*, M. Nobile, A.C. Balant*****

* Compaq Computer Corp., 200 Forest Street, 01752, Marlboro, MA, United States Of America

** IBM Hudson Valley Acoustics Laboratory, 522 South Road, 12601, Poughkeepsie, NY, United States Of America

*** State University of New York, HUM 24, Suite 6, 12561, New Paltz, NY, United States Of America

Tel.: 508 467 2390 / Fax: 508 467 2435 / Email: bob.hellweg@compaq.com

Keywords:
TONE, PROMINENCE, COMPUTER, PSYCHOACOUSTIC

ABSTRACT
This paper presents an update on the ICWG ITTE study of prominent discrete tones that was reported at INTER-NOISE 99 and presents details of Phase 2, a round robin involving 40 signals. A comparison of the objective and initial subjective ratings from different laboratories has been performed, and results are presented. The relative success of the two objective methods in predicting the engineers' initial subjective ratings also is discussed, as are the underlying issues and recommendations of the task group for the next phase of work.

1 - INTRODUCTION

This paper is a progress report on an initiative taken by the Inter-Committee Working Group on Noise from Information Technology and Telecommunications Equipment (ITTE) to resolve the problems surrounding the identification and evaluation of prominent discrete tones in product noise emissions. The ITTE industry in particular has been active in this area since the early 1970's. Much progress has been made since then, and many of the deficiencies and limitations of early standardized methods for rating the so-called prominence of tonal components in noise have been overcome. Recent efforts by the ITTE industry over the last 10 years have led to the development of two different objective procedures: (1) the Tone-to-Noise Ratio (TNR) method, standardized in ECMA-74-1999 [1] and ANSI S1.13-1995 [2], and (2) the Prominence Ratio (PR) method, also in ANSI S1.13. The TNR method compares the level of the tone to the level of the "masking" noise within the critical band centered on the frequency of the tone. The PR method compares the level of the critical band itself, centered on the tone, to the average of the levels of the two adjacent critical bands. For many product noise emissions, the results of the two procedures are equivalent and correlate well with subjective ratings, but there remain some differences and conflicting results. The primary goals of the ICWG initiative are to understand and resolve these conflicts and to optimize a single method and incorporate it into the relevant standards.
In 1999 an ICWG task group conducted a pilot study of eight "troublesome" signals that were analyzed objectively by the both the TNR and the PR method and rated subjectively. The results of the Phase 1 pilot study were reported at Inter-Noise 99 in Reference 3, which also presents a more complete background and description of the two objective methods.

2 - PHASE 2 ACTIVITIES AND EFFORTS

2.1 - Round robin

Historically round robins were conducted by sending an actual machine around for testing in conformance with a particular standard whose variability one was attempting to quantify. In the early 1990's, a round robin was conducted using recorded tones-in-noise signals on a high-fidelity, PCM digital audio (VHS) tape, in an attempt to control the variability of the source. For the round robin part of this project, test signals were recorded as digital ".wav" files and posted on a private web site. Participating members of the ICWG downloaded the files for both objective and subjective analysis in their labs.

2.2 - Signals

Forty signals were digitally recorded in standard.wav format using a sound card (44.1 kHz, 16-bit mono, about 1.6 MB per signal). The signals had previously existed on either VHS or DAT tapes at the IBM Hudson Valley Acoustics Lab. Twenty-two of the signals represent "synthetic" noise samples where laboratory-generated pure tones were mixed with laboratory-generated pink or shaped noise, and 18 of the signals represented actual ITTE machine noises containing tones, usually caused by air moving devices.

2.3 - Subjective and objective evaluations

Participants were asked to rate the prominence of a discrete tone on a subjective scale, and also to determine its objective TNR and PR according to the ECMA-74 and ANSI S1.13 procedures. (Note: The participants were asked to rate all 40 test signals subjectively but, due to the time involved, to rate only 20 of the signals objectively. Brief descriptions of these 20 test signals, 10 synthetic and 10 actual ITTE tone-in-noise signals, are given in Table 1.) For the initial subjective part of the round robin, participants were asked to listen to the 40 signals via headphones and to rate the prominence of the tones using a 7-point scale (0-inaudible, 1-barely audible, 2-audible but not prominent, 3-slightly prominent, 4-prominent, 5-very prominent, and 6-extremely prominent). For the objective part, participants were directed to follow the ECMA-74 and ANSI S1.13 TNR and PR procedures and to record various parameters needed for the computations. A second round of subjective ratings is to be performed on completion of the objective phase.

Signal	**Description of synthetic signal**
P04	250 Hz tone in pink noise
P05	1000 Hz tone in pink noise
M03	250 Hz tone in machine noise
M05	1000 Hz tone in machine noise
H04	1000 Hz tone centered on a pedestal of noise
V06	1000 Hz tone centered in a valley in noise
T03	2 tones (250 Hz and 285 Hz) within critical band in pink noise
T04	2 tones (250 Hz and 285 Hz), 3 dB higher than T03, within critical band in pink noise
T07	2 tones (1000 Hz and 1054 Hz) within critical band in pink noise
T08	2 tones (1000 Hz and 1054 Hz), 6 dB lower than T07, both within critical band
Signal	**Description of machine signal**
N02	Harmonic series with peak tone at 200 Hz (fundamental = 40 Hz)
N04	464Hz tone on pedestal, "rough" spectrum
N05	Harmonic series, peak tone at 1092 Hz (fundamental = 274 Hz)
N06	Harmonic series with 2 components, 275 Hz and 550 Hz
N11	Narrow band of noise centered at 428 Hz
N20	Three tones in noise, strongest at 2640 Hz, others at 2720 and 2788 Hz
N22	Several tones, strongest at 538 Hz and 506 Hz
N23	Two tones in noise, strongest at 2640, other at 2720
N27	Harmonic series with peak tone at 1600 Hz, (fundamental = 800 Hz)
N29	Two harmonic series with peak tones at 342 Hz and 354 Hz

Table 1: Description of 20 sounds used in ICWG prominent discrete tone round robin.

2.4 - Participants

The participants were twenty-eight ITTE acoustics engineers (or other laboratory personnel) from various IT companies. No formal screening for normal hearing was conducted, nor was the initial subjective part of the round robin conducted in a controlled manner (headphone type, listening level, and listening order were not specified). (Note: A parallel effort has been completed using formal psychoacoustical techniques with non-engineers rating the same 40 test signals subjectively [4].)

3 - PHASE 2 INITIAL RESULTS

At the time of writing of this paper, responses to the initial subjective rating part of the round robin had been received from 28 participants. The objective results have not been received from all participants at the time of writing because of a later deadline (due to the time required to perform objective analyses). The following results focus primarily on the subjective responses to the 20 signals used in the objective study. However, in order to compare the subjective ratings to the TNR and PR objective metrics, the authors' two labs (Compaq and IBM) completed the objective round robin, and their responses were averaged.

Table 2 presents the mean initial subjective responses of the 28 engineers to the 20 test signals along with the associated standard deviations. Also given are the computed TNR and PR following the ECMA-74 and ANSI S1.13 procedures. Since there were only two labs performing the objective analyses, a difference ("delta") is reported between them for the TNR and PR values rather than a standard deviation.

Signal Number	**Frequency, Hz**	**Average TNR, (n = 2), dB**	**Delta TNR, dB**	**Average PR, (n = 2), dB**	**Delta PR, dB**	**Mean Initial subjective response, n = 28**	**Std Dev of Initial Subj. Response, n = 28**
P04	250	15.3	0.03	15.5	-0.08	2.9	1.18
P05	1000	2.8	-0.31	4.6	0.02	1.8	0.82
M03	250	-0.9	0.12	-4.5	-0.26	0.9	1.08
M05	1000	10.4	-0.52	6.8	0.15	3.3	1.05
H04	1000	15.1	-0.09	19.0	-0.15	5.1	0.71
V06	1000	3.1	-0.22	-0.6	0.09	0.7	0.77
T03	250	9.2	-0.22	12.5	-0.11	2.5	0.94
T04	250	15.3	-0.20	18.4	-0.12	3.2	1.24
T07	1000	18.3	-0.68	18.0	0.01	4.4	1.13
T08	1000	6.5	-0.38	6.9	-0.10	1.6	0.96
	Synthetic	**Average**	**-0.25**	**Average**	**-0.05**		
		Std. Dev.	**0.24**	**SD**	**0.12**		
N02	200	6.9	0.15	5.8	0.25	3.0	1.45
N04	464	1.5	0.59	8.1	-0.01	2.3	1.65
N05	272	5.2	-0.65	2.6	-0.22	4.8	1.84
N06	274	10.7	0.97	7.1	0.01	3.3	1.28
N11	428	-2.8	5.99	9.1	-0.25	3.0	1.26
N20	2640	6.1	-1.06	7.7	-0.15	2.6	1.17
N22	505	4.4	1.19	2.5	-0.24	1.7	1.09
N23	2640	6.0	0.04	7.0	0.09	3.3	1.08
N27	1600	11.0	0.17	12.1	0.08	3.7	1.01
N29	342	3.8	0.91	10.5	-0.40	3.0	1.56
	Machine	**Average**	0.83	**Average**	-0.08	3.0	1.22
		SD	1.94	**SD**	0.20		
	Machine	**Avg (wo N11)**	**0.26**	**Avg (wo N11)**	**-0.07**		
		SD(wo N11)	**0.75**	**SD(wo N11)**	**0.20**		

Table 2: Average objective and initial subjective ratings of 20 sounds.

3.1 - Objective responses

The objective results from the authors' two labs show good agreement, especially for the synthetic signals. In general, PR shows less variability since the frequency of the tone defines all the necessary parameters. For the TNR, "human decisions" are required in bracketing the tone(s), and the results are affected by the bandwidth used in the analysis which can lead to lab-to-lab differences. This is illustrated dramatically for signal N11, which has a relatively broad peak at 428 Hz rather than a well-defined tonal spike (see figure 4).

3.2 - Subjective responses

The mean initial subjective responses show a range from "barely audible (1)" to "very prominent (5)" indicating that the signals provide a good sampling across the prominence spectrum. However, the key finding here was the high standard deviations for some of the signals, especially the real machine noises. A s of 1.45 around a mean of 3.0 (as was observed for signal N02) shows a range from between "barely audible" and "audible but not prominent" to between "prominent" and "very prominent". The actual range, of course, is even greater, with several signals being rated both "inaudible" and "extremely prominent" by different listeners (the results from controlled psychoacoustic experiments show almost as much inter-subject variation [4]). Given that our primary goal is to develop an objective metric that correlates closely with subjective responses, this large variability in the subjective responses is a concern.

3.3 - Correlation

Detailed regression lines for the mean initial subjective ratings for both TNR and PR and correlation coefficients will be presented after the objective part of the round robin is completed. However, correlation coefficients were determined for the 40 mean initial subjective ratings with the objective results from one laboratory. Mean subjective response (MSR) *vs.* TNR for synthetic signals yielded R=0.84; MSR *vs.* PR for synthetic signals, R=0.84; MSR *vs.* TNR for real machine noise signals, R=0.73; MSR *vs.* PR for machine noise signals, R=0.38. Refer to the companion paper [4] for the results of psychoacoustic studies on a larger set of stimuli with naïve subjects. Comparison of the initial subjective results from this round robin with a matching subset of the results from the psychoacoustic study [4] showed a high correlation between the engineers' and the non-engineers' ratings (R=0.94 for synthetic signals and R=0.90 for machine noises).

3.4 - Specific results

The following observations were determined from comparing the initial subjective ratings to the computed objective ratings in Table 2.

1. About one-third of the 20 signals represent cases where both the TNR and PR are acceptable in predicting subjective ratings of prominence (P05, H04, N02, N20, N22, N23, and N27). These represent a cross-section of signal types (single tones, low frequencies and high frequencies, tones on pedestals, harmonics, multiple tones, synthetic and real, etc.). For example, signal N27 shown in Figure 1 has a mean response of 3.7 ("prominent") and a TNR=11.0 and PR=12.1. (Recall the threshold of prominence according to the objective metrics is TNR $\geq$6.0 and PR $\geq$7.0).

2. There are cases (N04) where TNR seems to do better than PR (PR overestimates); cases (N06, M05) where PR does better than TNR (TNR overestimates); and cases (P04) where both overestimate the subjective response. The latter is plotted in Fig. 2 and shows a very strong synthetic tone at 250 Hz in pink noise. From previous research, we would have expected consistent ratings of at least "very prominent" for this type of signal.

3. There is also at least one case (N05) in which both TNR and PR significantly underestimated the subjective response. This signal contained a harmonic series in which the harmonics were stronger than the fundamental (see spectrum in Figure 3). It received one of the highest ratings "very prominent", yet both metrics would predict "not prominent" when rating the fundamental frequency of 274 Hz. (We plan to perform objective analyses for the harmonics.) This signal received several comments about its clearly annoying aspect ("raspy", "grating"), but there was some uncertainty whether or not it was "tonal".

4. Another example of neither metric correlating particularly well is N29 (two tones on a pedestal of noise) with TNR underestimating and PR overestimating the subjective response.

5. Trends of increasing objective levels seem to correspond to increasing subjective ratings (T03, T04, T07) even though the objective levels seem to overestimate the subjective ratings − not only for these signals, which are two tones in the critical band, but also more generally.

6. Signal N11, plotted in Fig. 4, illustrates an interesting case of a narrow band of noise, perhaps sharp enough to be perceived as "tonal". Participants rated this as "slightly prominent" (3.0), but the computed metrics differed. The large inter-laboratory difference in computed TNR resulted from each lab using significantly different frequency bandwidths to determine the level of the 428-Hz "tone".

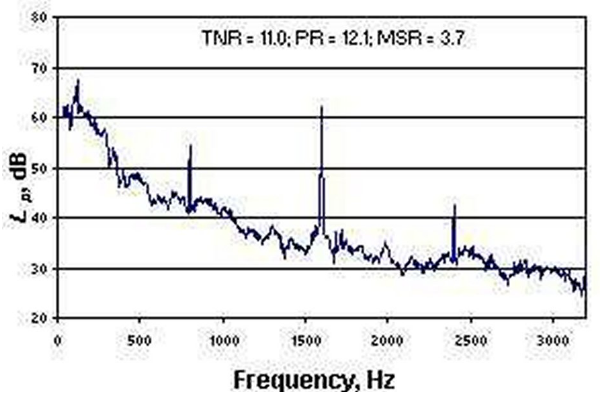

Figure 1: N27 with 1600 Hz tone.

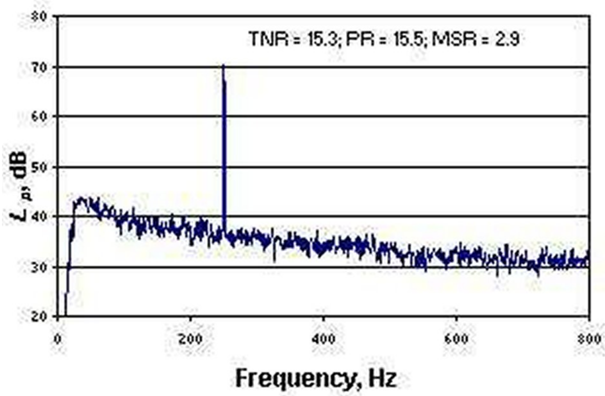

Figure 2: P04 with 250 Hz synthetic tone.

4 - CONCLUSIONS AND FUTURE DIRECTIONS

The ICWG Task Group Phase 2 effort has generated many questions yet provided several insights, and the authors are optimistic and feel that this effort has been positive. First of all, we now have a viable method of quickly and uniformly conducting further round robin testing in this area (the.wav files and electronic circulation). Second, we have begun to identify the aspects of computing the *objective* metrics that must be unambiguously specified in order to minimize lab-to-lab variability.

We feel the primary conclusion from the subjective part of this round robin (and from the formal psychoacoustic testing [4]) is that the substantial variability in the subjective ratings must be understood. Perhaps some subjects have different understandings of "prominence" or even "tone" itself. Perhaps the psychoacoustic percept of "prominence" itself is not as well defined as, say, loudness. Possibly the level of playback, sequence of playback, and/or the quality of the headphones affected judgements, as none of these were controlled here. In the case of the machine noise signals, subjects may have responded to tones other than those for which the objective measures were computed, or they may have been responding to the overall tonality of the signals, especially those that contained multiple tones and/or harmonics. Perhaps listening to a "reference" tone (as recommended in ECMA 74) might help ensure that all participants are indeed rating the same thing that is being modeled by the objective TNR and PR metrics.

In general, it appears that the values that are currently selected for the "threshold of prominence" for the objective metrics (TNR $\geq$6.0 and PR $\geq$7.0) may be too low. Many times both TNR and PR were computed much higher than these values but subjects would rate the signal as only "slightly prominent". Another important conclusion, also supported by the companion paper [4], is that the objective metrics should implement some type of frequency weighting for the computed TNR and PR. These results suggest that low-frequency tones are not perceived as prominent as higher frequency tones at the same TNR or PR.

The authors are looking forward to the completion of the objective phase of the round robin. The information that will be gained about the implementation of the two existing objective procedures will

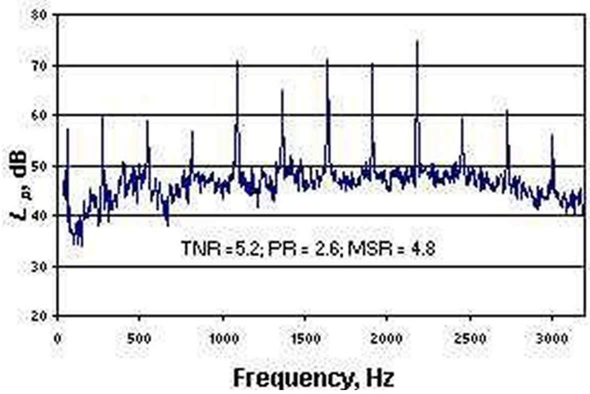

Figure 3: N05 with harmonics.

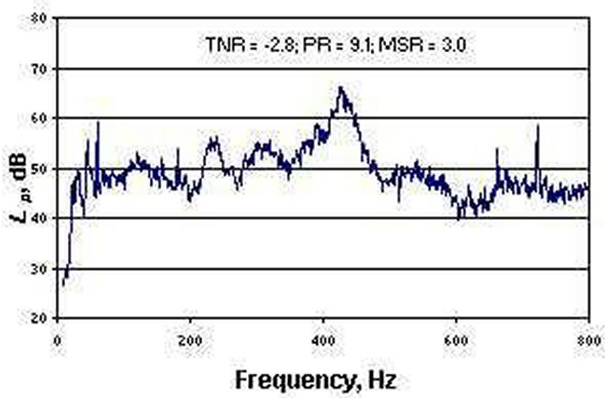

Figure 4: N11 with 428 Hz narrow band peak.

be combined with the results of current and planned psychoacoustic studies in developing a revised method for assessing the prominence of discrete tones in noise.

ACKNOWLEDGEMENTS
The authors would like to thank the participating labs in Phase 2 of this study: Acoustic Systems, Compaq, Dell, Hewlett Packard, IBM, Lexmark, Matsushita, and Unisys.

REFERENCES

1. **ECMA-74-1999**, *Measurement of Airborne Noise Emitted by Information Tech. and Telecomm. Equip.*1999

2. **ANSI S1.13-1995**, *Measurement of Sound Pressure Levels in Air*

3. **A. Balant, R. Hellweg, M. Nobile and L. Wittig**, A comparison of two methods for the evaluation of prominent discrete tones, In *INTER-NOISE 1999*, 1999

4. **A. Balant, K. Barringer, and M. Nobile**, A comparison of two methods for the evaluation of prominent discrete tones - Psychoacoustic results, In *INTER-NOISE 2000*, 2000