

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

I-INCE Classification: 6.9

PSYCHOACOUSTIC COMPARISON OF TWO METHODS FOR THE EVALUATION OF PROMINENT DISCRETE TONES

A.C. Balant*, K. Barringer, B.a.*, M. Nobile**

* State University of New York at New Paltz, Hum 18, 75 S. Manheim Blvd., Suite 6, 12561, New Paltz, NY, United States Of America

** IBM Hudson Valley Acoustics Laboratory, Bldg. 704, 2455 South Rd., M/S P226, 12601, Poughkeepsie, NY, United States Of America

Tel.: (914) 435-4962 / Fax: (914) 432-9880 / Email: balant@us.ibm.com

Keywords:
TONES, PROMINENCE RATIO, SOUND QUALITY, TONE TO NOISE RATIO

ABSTRACT
This paper presents the results of two pilot studies comparing the efficacy of the Modified Tone-to-Noise Ratio method and the Prominence Ratio method in assessing the subjective prominence of discrete tones in noise. Ten normal-hearing subjects rated the prominence of synthetic stimuli and real machine noises containing spectral features that are problematic for the two methods. Both methods showed strong positive correlations with the subjective prominence ratings of the synthetic stimuli, and weaker positive correlations with the ratings of the machine noises. Results and implications for future research are discussed.

1 - INTRODUCTION
A task group of the Inter-Committee Working Group (ICWG) on Noise from Information Technology and Telecommunications Equipment (ITTE) is conducting a study on two procedures for evaluating the potential annoyance of prominent discrete tones (PDTs) in the noise emissions of products [1]. These procedures are: the Tone-to-Noise Ratio (TNR) method, standardized in ECMA-74-1997 [2], ANSI S1.13-1995 [3], and ISO 7779-1999 [4], and the Prominence Ratio (PR) method, in ANSI S1.13. These procedures have been shown to correlate well with subjective results in many cases, but some problems remain, especially with certain types of signals [1]. As part of the ICWG initiative, psychoacoustic pilot studies with naive listeners are being conducted. The results of these studies will guide future research efforts directed at solving these problems. This paper presents the results of two pilot studies: one with synthetic stimuli and the other with recorded machine noises, many of which contained spectral features that are known to be problematic for the two procedures.
The history of the development of the procedures used by the ITTE industry for assessing the prominence of discrete tones is outlined in reference [1]. Because of problems identified with the original TNR method, particularly for multiple tones within the critical band [5], two independent approaches were taken, resulting in two different procedures. The *modified TNR method* [2,3,4] assumes that if multiple tones meet a frequency proximity criterion, they should be combined when computing the TNR (which compares the level of the tone to the level of the noise within the critical band centered on the tone). The *PR method* [3], [6] is a fundamentally different approach that compares the level in the critical band centered on the tone, including other tones and/or peaks in the noise, to the levels in the adjacent critical bands. In straightforward situations such as a single audible tone in broadband noise, the two procedures yield similar results. However, they yield different results in many cases: e.g., if the noise spectrum is irregular in the region of the tone; if there are multiple tones in the critical band; if there is a strong harmonic series; or if the tone is of very low frequency. Despite some encouraging psychoacoustic results early on [7,8], the validity of these methods has yet to be fully demonstrated, and both could be improved.

The psychoacoustic research being conducted at SUNY New Paltz is intended to supplement the work of the ICWG task group in finding an objective, joint solution to the continuing problems associated with quantifying the effects of discrete tones in noise. Thus, the goals of these pilot studies are to 1) determine the efficacy of the current procedures with actual noise emissions of ITT products, using the database of real machine noise samples containing discrete tones that is being developed by the ICWG [1]; 2) replicate a subset of the earlier psychoacoustic results to verify consistency with the experimental methods of the current study; and 3) develop the design for a series of definitive experiments that will lead to modification of the current procedures, or to the development of an entirely new procedure.

2 - METHODS

2.1 - Subjects

Subjects were ten female volunteers from the SUNY New Paltz campus community, mean age 24.8 years, range 20-37 years, with no known history of otologic disorders. All subjects passed a pure-tone audiometric screening at 15 dB HL at 500, 1000, 2000, and 4000 Hz.

2.2 - Materials

Pilot Experiment 1: Synthetic Stimuli. Characteristics of the synthetic stimuli, which were drawn from a large database of synthetic signals that had been recorded previously [7], are shown in Table 1. With the exception of the tones in machine noise, the signals had nominal TNRs (computed considering a single tone only) ranging from −3 to +15 dB. (In the case of 1000 Hz two-tone signals, the frequency separation of 55 Hz met the proximity criterion, and the actual TNRs − computed using the standardized procedure [2,3,4] − exceeded the nominal TNRs by 3 dB). The stimulus duration was 20 sec rather than the 2 sec used previously [7].

	Nom. TNR	250 Hz		1000	
		TNR	PR	TNR	PR
Single Tone in Pink Noise	-3	-2.05	2.14	-2.55	1.51
	3	3.27	5.08	2.61	4.58
	9	9.07	9.77	9.07	9.50
	15	15.32	15.47	14.97	15.03
Single Tone in Machine Noise		-0.79	-4.65	-3.17	-1.81
		5.35	-0.35	3.02	1.19
		10.66	4.16	10.11	6.89
Single Tone on 6-dB Pedestal	3	3.80	11.66	3.09	8.38
	9	9.04	16.60	8.86	13.54
	15	15.14	22.34	15.10	18.96
Single Tone in 6-dB Valley	3	2.79	-2.74	2.97	-0.59
	9	8.67	2.01	8.93	4.02
	15	14.91	7.80	14.88	9.60
Two Tones in Pink Noise	-3	-2.14	2.61	0.36	2.86
	3	3.15	7.39	6.26	6.85
	9	9.07	12.49	12.12	12.27
	15	15.21	18.35	18.00	18.01

Table 1: Characteristics of stimuli used in pilot experiment 1 (the "nominal" TNR values were used for the analysis of variance of the results).

Pilot Experiment 2: Real Machine Noises. These stimuli consisted of 20-sec segments of ITTE machine noises that had been recorded during product development at the IBM Hudson Valley Acoustics Lab. Stimuli were selected to provide a wide range of TNR and PR values, and to provide instances in which those values differed. As indicated in Table 2, many of the stimuli had spectral characteristics that are known to be problematic, including multiple tones in the critical band, harmonic series, and "pedestals" in the background noise.

Code	TNR	PR
aLHT	7.01	5.94
bL	-6.07	3.59
cLP	1.82	8.13
dNH	9.13	5.46
eNL	11.20	7.11
fPV	-1.33	7.12
gV	-0.33	-1.13
hL	6.55	7.32
iLN	-3.31	0.33
jLP	3.22	8.96
kLNP	10.48	13.04
l3V	6.19	8.60
mLP	1.10	4.86
nL	-0.91	-0.03
oL	0.22	-0.11
pHT	2.11	2.99
qHT	5.55	7.62
rL	1.44	2.58
sLT	5.01	2.38
tHT	6.04	7.01
uH	9.60	11.35
vH	10.71	10.88
wH	3.03	6.16
xH	11.09	12.18
yL	-1.23	-0.70
zLPT	4.29	10.28

Table 2: Characteristics of stimuli used in pilot experiment 2; the letter code identifies each signal and also indicates the spectral features: V = very low frequency (< 200 Hz), L = low frequency, (200 - 1000 Hz), H = high frequency (> 1000 Hz), N = harmonic series, P = pedestal, T = two tones in the critical band.

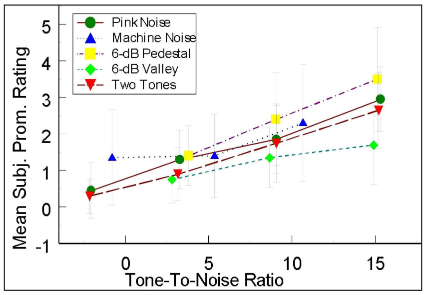
2.3 - Procedures

The stimuli were recorded using a Tascam DA-40 digital audio tape (DAT) recorder. During testing, the stimuli were played back from the DAT through the line input of a Starkey AA30 audiometer which was set at 55 dB HL. Subjects were tested in an Eckel sound-treated audiometric test suite. Stimuli were presented monaurally (right ear) using Telephonics TDH-49 headphones. Each subject listened to a set of recorded instructions that explained the concept of prominence. The subjects were asked to perform prominence ratings on the following scale: *0-Inaudible, 1-Barely audible, 2-Audible but not prominent, 3-Slightly prominent, 4-Prominent, 5-Very prominent, 6-Extremely prominent.* They were instructed to base their judgements on the tonal component that seemed to be most prominent if there was more than one audible tone. A response sheet with columns of equal width was used to encourage interval ratings of the stimuli. The stimuli were presented in random order with an interstimulus interval of 10 seconds. Each subject rated all stimuli on two occasions separated by two to ten days.

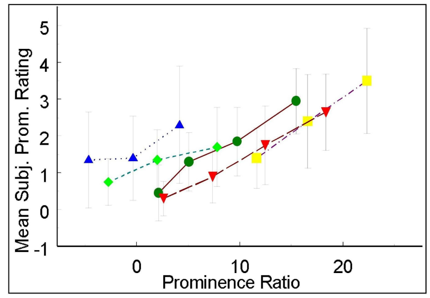
3 - RESULTS

Results of Pilot Experiment 1: The mean subjective prominence ratings of the 250 Hz stimuli are shown as a function of TNR in Fig. 1a and as a function of PR in Fig. 1b. The ratings of the 1000 Hz stimuli are shown as a function of TNR in Fig. 1c and as a function of PR in Fig. 1d. Overall, the signals containing 250 Hz tones were rated less prominent than those containing 1000 Hz tones. At the highest TNRs (15 dB) the mean subjective prominence ratings of the 250 Hz signals ranged from audible (2) to slightly prominent (3), while the ratings for the 1000 Hz signals ranged from slightly prominent (3) to very prominent (5). The stimuli with tones on "pedestals" of noise were rated more prominent, and those with tones in "valleys" of noise were rated less prominent, than the tones in pink noise.

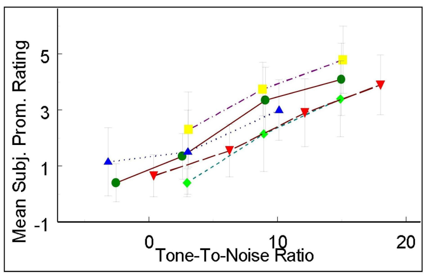
To assess the significance of these observed differences in the mean subjective ratings, a repeated-measures analysis of variance (ANOVA) with nested subject groups was conducted on a subset of the results. The variables were: nominal TNR (T) with three levels (3, 9, and 15 dB), frequency (F) with two levels (250 Hz and 1000 Hz), noise type (N) with four levels (pink, pink with 6-dB pedestal, pink with 6-dB valley, and two tones), replications (first vs. second measurement) and subjects (S) with 10 levels. The significant main effects included T, F, N and S (all with $p < 0.001$). There was a significant interaction between T and F ($p < 0.001$) (i.e., the growth of the prominence ratings with increasing TNR depended on the frequency) and between T, F, and N ($p = 0.0075$). Post-hoc testing revealed that the mean rating for the pedestal signals was significantly higher than the mean ratings for the pink noise and two-tone signals. The latter ratings did not differ significantly from each other, but were significantly higher than the ratings for the valley signals (Tukey HSD method, $p<0.01$).



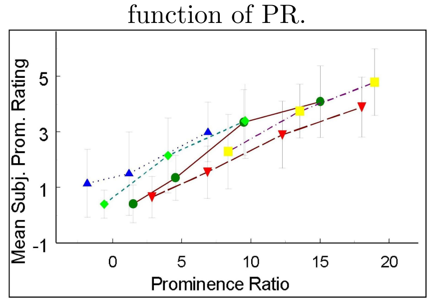
(a): Mean subjective prominence ratings of the 250 Hz stimuli as a function of TNR.



(b): Mean subjective prominence ratings of the 250 Hz stimuli as a function of PR.



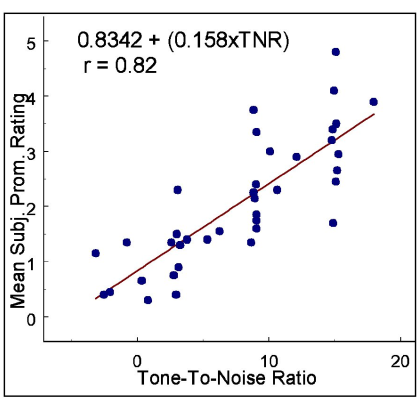
(c): Mean subjective prominence ratings of the 1000 Hz stimuli as a function of TNR.



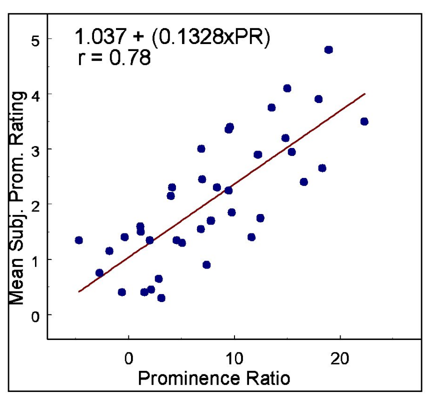
(d): Mean subjective prominence ratings of the 1000 Hz stimuli as a function of PR.

Figure 1: Results of pilot experiment 1.

Figure 2 shows the Pearson Product Moment correlations between the mean subjective prominence ratings and the TNR and PR values of all stimuli used in Pilot Experiment 1. The TNR and the PR procedures yielded comparable strong, positive correlations with the subjective ratings. Regression lines and correlation coefficients are shown in the figure.



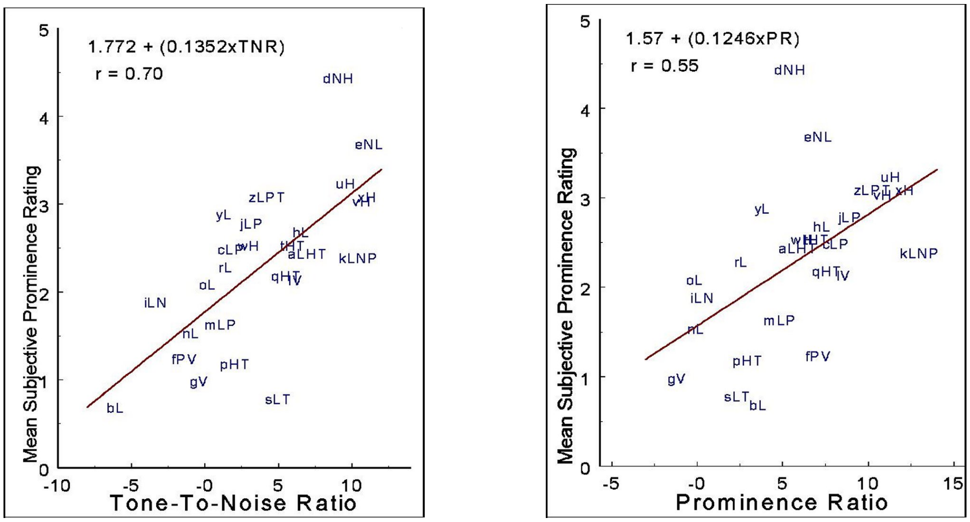
(a): Correlation of mean subjective prominence ratings of synthetic stimuli with TNR.



(b): Correlation of mean subjective prominence ratings of synthetic stimuli with PR.

Figure 2: Results of pilot experiment 1.

Results of Pilot Experiment 2: Figure 3 shows the Pearson Product Moment correlations between the mean subjective prominence ratings and the TNR and PR values for all stimuli used in Experiment 2. Regression equations and correlation coefficients are shown in the figure. The symbols on the graphs indicate the characteristics of the stimuli, including harmonics (N), pedestals (P), multiple tones (T), and the frequency range of the tone (V, L, or H). Taking all signals into account, there is a mid-positive correlation between the subjective ratings and the TNR, and a weaker positive correlation between the ratings and the PR.



(a): Correlation between mean subjective prominence ratings and TNR measured on the 26 recorded machine noises.

(b): Correlation between mean subjective prominence ratings and PR measured on the 26 recorded machine noises.

Figure 3: Regression lines and Pearson Product Moment correlation coefficients are indicated on the figure; H = high frequency (>1000 Hz), L = low frequency (200 Hz − 1000 Hz), V = very low frequency (<200 Hz), N = harmonic series, P = pedestal, T = two tones in critical band; the letter codes are cross-referenced in Table 2.

4 - DISCUSSION

Pilot Experiment 1: Despite the small sample size, limited stimulus set, and differences in method relative to previous research (which used a more open-ended magnitude estimation technique with short-duration stimuli) the results of Experiment 1 were generally consistent with previous research [7,8]. Tone frequency was a significant factor in determining prominence, such that a 1000 Hz tone in pink noise at a TNR of 9 dB had a mean subjective prominence rating comparable to a 250 Hz tone at a TNR of 15 dB. Future studies should include a wider range of tone frequencies. As might be anticipated from the earlier work [7,8], the tones on the 6-dB pedestals of noise were consistently rated more prominent than tones in pink noise and/or two-tone signals at the same TNRs, while tones in 6-dB "valleys" were rated less prominent. In contrast to previous findings, both the PR method and the TNR method seemed to overestimate the prominence of multiple tones in the critical band. This was true both at 250 Hz, where the proximity criterion was not met, and at 1000 Hz, where it was met. Further investigation will be required to determine the source of this discrepancy, which may be related to the differences in psychoacoustic methods.

Pilot Experiment 2: As can be seen from the correlation plots in Fig. 3, many of the responses cluster fairly close to the regression lines. In general, stimuli with lower-frequency tones tended to be rated less prominent, consistent with the results of Experiment 1. Stimuli with pedestals tended to fall above the regression line when plotted according to TNR but were clustered closer to the regression line for PR. Stimuli with strong harmonic series tended to be rated more prominent, especially if a harmonic other than the fundamental was dominant (stimulus dNH). In this limited sample the TNR was more highly correlated with subjective ratings than the PR, but even the TNR method accounted for only one-half of the variation in the subjective ratings ($r^2 = 0.49$). One source of variability may be the fact that many of these stimuli contained multiple tones in different critical bands. Since "search tones" were not used, subjects may in fact have been responding to different tones than the ones for which the TNR and PR were computed, or their responses may have been influenced by the overall tonality of the signals.

5 - SUMMARY AND PLANS FOR FUTURE RESEARCH

Despite the small sample size and limited stimulus set, the pilot studies reported here support the view that both TNR and PR show promise for assessing the psychoacoustic impact of prominent discrete tones in noise. The data suggest that future research efforts should be aimed at: 1) developing a correction to account for the dependence of subjective prominence on frequency; 2) assessing the efficacy of the two-tone correction of the modified TNR method; and 3) developing a correction to account for the effects of harmonic series.

REFERENCES

1. **A.C. Balant, R.D. Hellweg, Jr., and M.A. Nobile**, A comparison of two methods for the evaluation of prominent discrete tones, In *INTER-NOISE 1999*, pp. 1373-1378, 1999

2. **ECMA**, *ECMA 74 - 1997 Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment*, 1997

3. **ANSI**, *ANSI S1.13-1995, American National Standard Methods for the Measurement of Sound Pressure Levels in Air*, 1995

4. **ISO**, *ISO - 7779, Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment, Second Edition*, 1999

5. **L.E. Wittig and R.D. Hellweg, Jr.**, Problems with identification of prominent discrete tones, In *NOISE-CON 87*, pp. 733-738, 1987

6. **M.A. Nobile and G.R. Bienvenue**, A procedure for determining the prominence ratio of discrete tones in noise emissions, In *NOISE-CON 91*, pp. 475-482, 1991

7. **Bienvenue, G.R., Nobile, M.A., Corkery, M.J., and Miscedra, S.M.**, Quantifying subjective responses to discrete tones in noise from computer and business equipment, In *INTER-NOISE 89*, pp. 859-864, 1989

8. **Bienvenue, G.R., Nobile, M.A., and Barker, M.E.**, Prominence balancing experiments for noise spectra containing discrete tones, In *NOISE-CON 91*, pp. 507-514, 1991