An Assessment of Sound Annoyance as a Function of Consonance

Song Hui Chon
CCRMA / Electrical Engineering
Stanford University
Stanford, CA 94305
shchon@ccrma.stanford.edu

Abstract

In this paper I study the annoyance generated by synthetic auditory stimuli as a function of consonance. The concept of consonance is one of the building blocks for Western music theory, and recently tonal consonance (to distinguish from musical consonance) is getting attention from the psychoacoustics community as a perception parameter. This paper presents an analysis of the experimental result from a previous study, using tonal consonance as a factor. The result shows that there is a direct correlation between annoyance and consonance and that perceptual annoyance, for the given manner of stimulus presentation, increases as a stimulus becomes more consonant at a given loudness level.

1 Introduction

1.1 Addressing Annoyance

Annoyance is one of the most common feelings people experience everyday in this modern society. It is a highly subjective sensation; however this paper presumes that there is a commonality behind each person's annoyance perception.

It is generally understood that one aspect of annoyance is a direct correlation with spectral power [1], although the degree of correlation is under investigation [2][3]. Previous studies have assumed that noise level is a strict gauge for determining annoyance [4], but it is reasonable to suggest that the influence of sound level might be more subtle; annoyance is, after all, a subjective and highly personal characteristic. One study reveals that subjects perceive annoyance differently based on their ability to influence it [5]. Another demonstrates that age plays a significant role in annoyance judgments [6], though conflicting results indicate more complex issues. In [7], a number of hypotheses were examined with respect to annoyance, but it failed to obtain a conclusive evidence for any.

1.2 Consonance

The Miriam-Webster dictionary defines "consonance" as "harmony or agreement among components." Consonance is a measure of the perceived pleasantness of complex tones based on their harmonic structure. Often it is used to describe the perception of music, since Western classical music theory is based on the harmonic consonance. Historically, there have been different approaches to define consonance. Pythagoras used a mathematical approach, saying two tones with a ratio of small whole numbers are more consonant than those with a higher ratio [8]. Helmholtz argued that the ear is a frequency analyzer, and therefore the more harmonics (or partials) are in coincidence, the more consonant those tones are [8]. Terhardt [9] as well as Gerson and Goldstein [10] used the concept of tonal fusion and pattern matching. From the audiological perspective, Patterson [11] tried to explain it in terms of neural-firing coincidence.

Terhardt defined consonance as "a link between music and psychoacoustics" in [12]. On the same line, there were attempts to employ consonance as a perceptual parameter recently [13] [14]. This paper follows the Quantified Total Consonance (QTC) algorithm proposed by Chon, et al. in [14] for analysis of the experiment data from our previous study in [3].

2 Perceptual Experiment

This section is a summary of the experiment we presented in an earlier paper. For details, refer to [3].

2.1 Stimuli

The experiment made use of six types of stimuli generated in MATLAB at four intensity levels: 50, 60, 70 and 80 dB SPL, over a frequency range of 500 to 5000 Hz. The six stimulus groups are:

1. Pink Noise (PN)
2. Two Sinusoids (TS) at 1000 and 1023 Hz with equal amplitudes
3. Narrowband Noise (NBN), generated by filtering Gaussian-white noise with a bandpass filter with the passband of 1000 and 1500 Hz
4. Broadband noise (BBN), generated similarly, with the difference of the bandpass filter passing the frequencies between 500 and 5000 Hz
5. Narrowband tone complex (NTC) consisting of ten linearly distributed sinusoids of random amplitudes between 1000 and 1500 Hz, and
6. Broadband tone complex (BTC), consisting of 100 sinusoids of equal amplitudes whose frequencies are logarithmically distributed (so that there are an equal number of tones per octave) between 500 and 5000 Hz.

Figure 1 shows the spectrum of six stimuli at 50 dB. In each subfigure, the x-axis is in Hz and the y-axis in dB.

2.2 Experimental Result

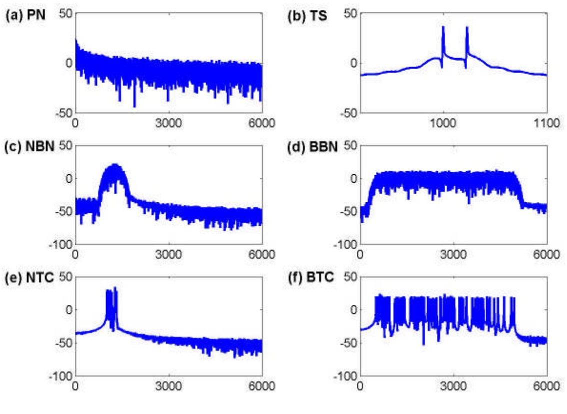

Figure 1: Frequency spectrum representations of the six stimuli at 50 dB SPL.

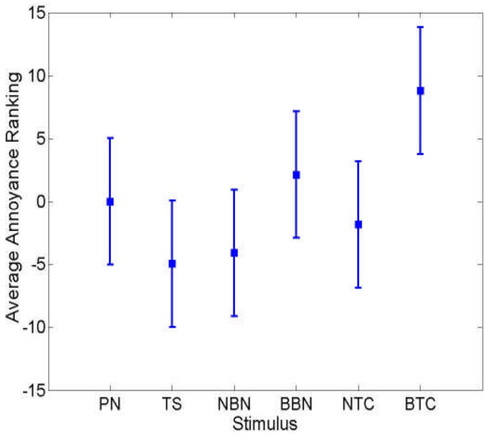

Figure 2: Absolute stimulus annoyance per stimulus type. The error bars in the graph represent the standard deviation.

As it turns out, it is possible to conclude that one type of stimulus is more annoying than another, at a given loudness level. The graph shown in Figure 2 was created by taking the average stimulus ranking over all four of the intensity levels, thus canceling the loudness effect, and then averaged over all sixteen subjects. It is clear that this deviation is rather static, and thus, conclusions can be drawn about the total average annoyance ranking per stimulus. Based on the location of the averages, it can be concluded that subjects found the two-sinusoid (TS) stimulus least annoying while they found the broadband tone cluster (BTC) the most annoying, in general. Of course, these rankings may change given sufficient loudness discrepancies. Also note that as in Figure 2, the y-axis values in this figure are irrelevant and that only relative heights of these averages are useful in determining the loudness factor.

Figure 3 provides a summary of the experimental result. Different slopes of line segments show that the same 10 dB increase affects the annoyance perception differently depending on the stimulus type. Also, we can see from the height differences of data points that there is a consistency in the experiment result between the stimulus type and the annoyance perception. This "consistency" seems to suggest a correlation between the bandwidth of the stimulus and the perceived annoyance. It is also interesting to see that stimuli with more tonal components (NTC, BTC) are perceived as more annoying than noise-derived stimuli of similar bandwidth (NBN, BBN/PN respectively). This may be related to the fact the perception of tones is different from that of noises, as frequently mentioned in masking effects [15]. From this "consistency", we can conjecture that one type of stimuli may be more annoying than others at any loudness level considered (i.e. BTC).
As pointed out in [3], the numbers on the y-axis of Figure 2 is meaningless. Only the relative height difference is of importance.

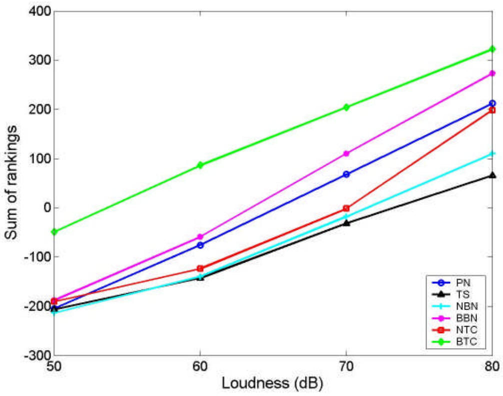

Figure 3: Total sum matrix. Results of all sixteen subjects added together.

Table 1: An example annoyance matrix

-19	-8	3	6
-17	-21	-11	3
-21	-6	0	6
-10	-7	8	19
-9	-5	9	15
6	17	19	23

2.3 Analysis Methodology

A novel method was proposed in [3] for an efficient analysis of the experiment result. Since we followed a two-interval, two-alternative forced choice (2I-2AFC) protocol, for each pair of stimuli we awarded +1 to the "more annoying" stimulus and -1 to the "less annoying" stimulus. This operation was performed for all 276 pairs and summarized in a matrix for each subject. The matrix dimension was 6-by-4, corresponding to six stimuli groups and four loudness levels. A sample "annoyance matrix" is shown in Table 1.

The rows correspond to stimuli types in the order of PN, TS, NBN, BBN, NTC and BTC and the columns to 50, 60, 70 and 80 dB SPLs in the respective order. The sum of all values in such a matrix always equals zero.

3 Quantified Total Consonance

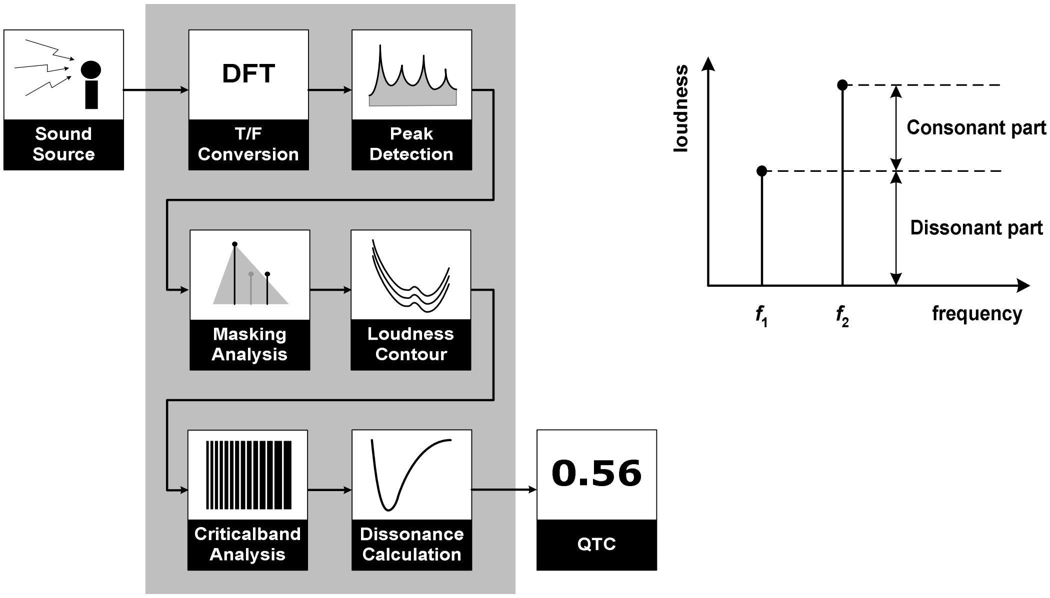

Figure 5: (a) QTC Calculation Procedure (based on [14]), (b) Decomposition of consonant and dissonant parts in two adjacent peaks (from [14])

Chon, et al. proposed in [14] an algorithm to quantify the total consonance of complex tones and demonstrated the correlation between the QTC values and the subject test results. I used a MATLAB implementation based on their algorithm. Figure 5(a) presents our QTC calculation procedure.

1. Sound Source: The source is assumed to be clean or already de-noised, and stored in a wav file.
2. T/F Conversion: FFT (Fast Fourier Transform) is used for Time-to-Frequency conversion. A rectangular window was used.
3. Peak Detection: The *i*-th frequency line is defined to be a peak when its magnitude is above a threshold value and greater than its four neighboring lines (at *i*-1, *i*-2, *i*+1, *i*+2). The threshold as well as the consideration of four neighbors was set empirically.
4. Masking Analysis: MPEG Psychoacoustic Model 2 [15] was used for masking effect calculation. This process eliminates local peaks that are masked by nearby loud peaks.
5. Loudness Contour: This block produces Equal Loudness Contours by interpolating ISO226 specification [16]. Then the amplitudes (in dB SPL) of the peaks are converted into phons using these contours. Only the range of 20 to 12500 Hz of pink noise was considered for this step, due to the limitation in ISO226.
6. Critical Band Analysis: The critical bandwidth formula by Moore & Glasberg [17] was used to determine whether neighboring two peaks are within the same critical bandwidth, hence creating interference in the perception of those tones. I followed the concept of consonance defined by Plomp and Levelt in [18] and assumed that two peaks do not create any dissonance when they are apart by more than 1.2 times the critical bandwidth of the mid-frequency (which is the mean of the two peak frequencies).

7. Dissonance Calculation: Consider a pair of pure tones with unequal magnitudes, as given in Figure 5(b). The dissonance perceived is determined by the tone with smaller magnitude. Using all the analyses above and the formula below (from [14]), the total dissonance D of the input signal is calculated.

$$D = \frac{1}{L}\sum_{\forall i<j} \min(l_i, l_j)\cdot d_{ij}$$

   where l_i is the loudness (in phon) of i-th peak, d_{ij} the dissonance between i-th and j-th peaks and L the total loudness of all peaks.

8. QTC: The quantified total consonance C of the input signal is calculated by $1-D$.

3.1 QTC Values

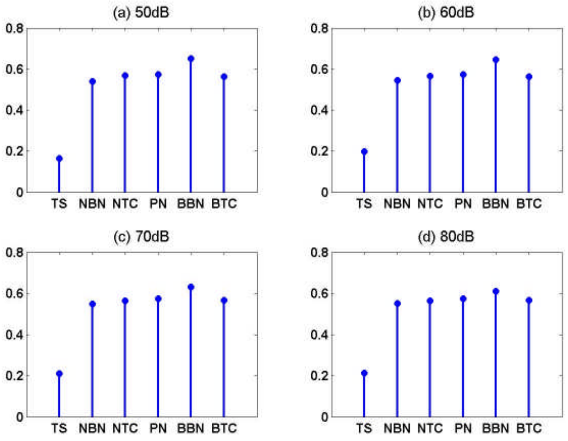

Figure 4: QTC values of the twenty-four stimuli, per loudness level (dB SPL)

Table 2: QTC Values of the Twenty-four Stimuli

	50dB	**60dB**	**70dB**	**80dB**
PN	0.5743	0.5746	0.5750	0.5754
TS	0.1642	0.1982	0.2114	0.2136
NBN	0.5409	0.5433	0.5479	0.5503
BBN	0.6508	0.6455	0.6310	0.6105
NTC	0.5678	0.5654	0.5645	0.5640
BTC	0.5621	0.5644	0.5662	0.5672

Table 2 presents the QTC values of the twenty-four stimuli, which is illustrated in Figure 4. TS exhibited the least consonance, and NBN turns out to be less consonant than BBN. The x-axis of each subfigure in Figure 4 is in increasing order of annoyance perception. We can see that, with the exception of BTC, the annoyance order is indeed a linear function of consonance (QTC values) and stimulus bandwidth. This can be seen from all four loudness levels.

Note in table 2 that the QTC values of a stimulus type change slightly depending on the loudness level. They usually increase as the loudness level increases, except for BBN and NTC. The gradual decrease of the QTC values in BBN and NTC is quite consistent, just as the other stimuli types show a consistent increase. At this point, the author does not have an explanation for what causes this discrepancy.

3.2 Correlation Coefficients

Correlation coefficients were calculated to verify the relationship between QTC and annoyance perception. Table 3 shows the correlation coefficients of QTC values and loudness levels and Table 4 of QTC values and annoyance order (from the least annoying to the most). The annoyance order [TS, NBN, NTC, PN, BBN, BTC] was obtained in the earlier experiment [3]. In accordance with table 2 above, PN, TS, NBN and BTC turned out to have positive correlation coefficients in table 3, while BBN and NTC negative correlation coefficients. This means that the QTC value is an increasing function of loudness levels for PN, TS, NBN and BTC, and a decreasing function for BBN and NTC.

Table 3: Correlation coefficients of QTC values and loudness levels per stimulus type

PN	TS	NBN	BBN	NTC	BTC
0.7484	0.6853	0.7412	-0.7272	-0.7412	0.7393

The correlation coefficient ρ_{SQ_i} is calculated using the formula below.

$$\rho_{SQ_i} = \frac{\operatorname{cov}(S, Q_i)}{\sigma_S \sigma_{Q_i}} = \frac{E[(S-\mu_S)(Q_i-\mu_{Q_i})]}{\sigma_S \sigma_{Q_i}} \text{ for } i = 1,2,3,4,5,6$$

where S is the ordering of [50 60 70 80] as in dB SPL of input signal loudness, Q_i the QTC values of i-th stimulus type (in the order of [TS, NBN, NTC, PS, BBN, BTC]), μ and σ the mean and the standard variation of the sub-scripted variable respectively.

Table 4: Correlation coefficient of QTC values and annoyance order, per loudness level

50 dB	60 dB	70 dB	80 dB
0.5970	0.5987	0.5964	0.5887

The correlation coefficient ρ_{AQ_j} for QTC values and the stimuli types was calculated as below.

$$\rho_{AQ_j} = \frac{\operatorname{cov}(A, Q_j)}{\sigma_A \sigma_{Q_j}} = \frac{E[(A-\mu_A)(Q_j-\mu_{Q_j})]}{\sigma_A \sigma_{Q_j}} \text{ for } j = 1,2,3,4$$

where A is the ordering of [TS, NBN, NTC, PS, BBN, BTC] from annoyance perception, Q_j the QTC values at j-th loudness level, μ and σ the mean and the standard variation of the sub-scripted variable respectively.

The stimuli types were ordered TS, NBN, NTC, PS, BBN, BTC following the annoyance perception order. It is in the order of increasing bandwidth and noise-based stimulus type before tone-based type within similar bandwidths. In table 4, we can see that the annoyance order and the QTC values are moderately correlated with same magnitude regardless of the loudness level. It indicates that at a fixed loudness level, the annoyance order is a function of the QTC values, which in turn is a function of loudness level and stimulus type as shown in table 3. This also hints that the QTC values probably do not change with a loudness level change. A further investigation is required to verify this conjecture.

4 Discussion

From Figure 4, we saw that there is a consistent relationship between the annoyance perception order and QTC values regardless of loudness level, with the exception of BTC. This indicates that the annoyance perception is indeed a function of consonance (QTC), in general. At this stage, there is no explanation behind why BTC behaves differently from other stimuli.

The annoyance perception ordering of TS, NBN, NTC, PN, BBN, and BTC is also worth noticing. It is in the ascending order of the bandwidth of each stimulus, and within the stimuli with the same bandwidth, the noise-based stimuli (NBN and BBN) are less annoying than tone-based ones (NTC and BTC respectively). This seems to be in accordance with the masking theory [15], that human brain reacts differently to tone maskers and noise maskers. It also may explain why BTC is already very annoying at a relatively quiet level (50 dB SPL) even though its QTC values are not too different from other stimuli (with an exception of the TS).

5 Conclusion and Future Work

The experimental result of annoyance perception was analyzed using the QTC on twenty-four stimuli. The order of annoyance perception shows a positive correlation with the loudness level, as one can easily expect, for most stimuli; however, the consonance does not have a consistent correlation with all stimuli types and loudness levels. For most stimuli types, the QTC and loudness level were moderately positively correlated, but for NTC and BBN they were negatively correlated. The author could not find any literature on the matter of consonance in terms of the loudness level.

The annoyance perception order at a fixed loudness level showed a positive correlation with QTC values. The correlation coefficients for all four loudness levels were same in magnitude (with some numerical error). The implication is that QTC is a consistent and effective parameter to describe annoyance perception. The importance of this finding cannot be established at this stage. Further studies need to be pursued in the future.

The author believes that consonance may play an important role in psychoacoustics; however, the theory of consonance needs to be re-visited for that purpose. Most of research in consonance is based on the study by Plomp and Levelt in [18], which approaches the subject from a musical point of view. Consonance of noise should be pursued as well as the relationship between consonance and loudness level, for a better and more general understanding of consonance.

Acknowledgments

The author thanks Sang Bae Chon and Daniel Steele for their help and expertise at various stages in the writing of this paper.

References

[1] Rothauser, E.H., Urbaner, G.E. & Pachl, W.P. (1968) Loudness and annoyance of impulsive office noise, pp. 520 – 522.

[2] Taylor S.M. (1982) A comparison of models to predict annoyance reactions to noise from mixed sources. *Journal of Sound and Vibration* **81**(1):123 – 138.

[3] Steele, D.L. & Chon, S.H. (2007) A Perceptual Study of Sound Annoyance. *Proceedings of the 2nd Audio Mostly Conference*. pp. 19-24

[4] Miedema, H.M.E. (2004) Relationship between exposure to multiple noise sources and noise annoyance. *Journal of Acoustic Society of America* **116**(2):949 – 957.

[5] Maris, E., Stallen, P.J., Vermunt, R. & Steensma, H. (2007) Noise within the social context: Annoyance reduction through fair procedures. *Journal of Acoustic Society of America* **121**(4): 2000 – 2010.

[6] Botteldooren, D. & Verkeyn, A. (2002) Fuzzy models for Accumulation of reported community noise annoyance from combined sources, *Journal of Acoustic Society of America* **112**(4):1496 – 1508.

[7] Fields, J.M. (1993) Effect of personal and situational variables on noise annoyance in residential areas. *Journal of Acoustic Society of America* **93**(5):2753 – 2763.

[8] Helmhotz, H. (2007) On the Sensations of Tone, Courier Dover Publications.

[9] Terhardt, E. (1974) On the perception of periodic sound fluctuations (roughness). *Acoustica* **30**(4):201-213.

[10] Gerson, A. & Goldstein, J.I. (1978) Evidence for a general template in central optimal processing for pitch of complex tones. *Journal of the Acoustical Society of America* **63** 498-510.

[11] Patterson, R.D. (1986) Spiral detection of periodicity and the spiral form of musical scales. *Psychology of Music* **14**:44-61.

[12] Terhardt, E. (1984) The Concept of Musical Consonance: A Link between Music and Psychoacoustics. *Music Perception* **1**(3):276-295.

[13] Skovenborg, E. & Nielsen, S.H. (2002) Measuring Sensory Consonance By Auditory Modelling. *Proceedings of the 5th International Conference on Digital Audio Effects (DAFX-02), Hamburg, Germany.* pp. 251-256.

[14] Chon, S.B., Choi, I.Y., Lee, M. & Sung, K.M. (2006) Quantified Total Consonance as an Assessment Parameter for the Sound Quality. *121st Audio Engineering Society Convention.* San Francisco.

[15] Bosi, M. & Goldberg. R.E. (2003) Introduction to Digital Audio Coding and Standards. Kluwer Academic Publishers.

[16] International Organization for Standardization (ISO) Standard 226:2003 (1987, Revised 2003) Acoustics - Normal equal-loudness-level contours.

[17] 1. Moore, B.C.J. & Glasberg, B.P. (1983) Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns. *Journal of the Acoustical Society of America* **74** (3):750-753.

[18] Plomp, R., & Levelt, W.J.M. (1965) Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America* **38**, 548–560.