International Journal of Industrial Ergonomics 53 (2016) 219–228

Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

journal homepage: www.elsevier.com/locate/ergon

Assessment of the subjective perception of music quality using psychoacoustic approach

Kuo Hao Tang [a, *], Shih Ting Chen [a], Yu Ting Tsai [b]

[a] Department of Industrial Engineering and System Management, Feng Chia University, Taichung 40724, Taiwan, ROC
[b] Bachelor's Program in Precision System Design, Feng Chia University, Taichung 40724, Taiwan, ROC

ARTICLE INFO

Article history:
Received 18 August 2014
Received in revised form
22 November 2015
Accepted 17 January 2016
Available online 15 February 2016

Keywords:
Psychoacoustics
Roughness
Music quality
Subjective perception
Prediction model

ABSTRACT

It is generally agreed that sound quality is one of the most difficult to measure characteristics of an electroacoustic products such as an earphone or a loudspeaker. A conventional approach used to measure people's subjective perception of these sound reproduction products is to conduct a jury test on a group of experiment participants; however, jury tests require considerable costs, including those of effort and time. As development speed and cost become strategic competitive dimensions, electroacoustic industry needs a more efficient approach to assess their newly developed products for subjective sound quality. This study developed and validated a quantitative model, the tonal harmony level (THL), that can effectively predict people's subjective perceptions of music quality. Participants' subjective perception and preference was measured for four music genres by listening to short music excerpts (8 s) in both ordinal and interval scales. The purpose of using two scales is to examine the consistency between subjective perceptions and to determine the robustness of the subjective measurements. The experimental results were very stable over the two assessment procedures, and the objective THL measure is highly correlated to subjective preference. The analysis suggests that the construction of subjective music quality prediction models should also consider music genre. Among four types of music, musical solos consisted of human vocals accompanied by a few instruments has a distinct pattern from the other three types. Thus, while R^2 value of the overall regression model is 0.707, the R^2 values are 0.955 and 0.901 when four music genres are categorized into two groups according to their patterns. When efficiency and accuracy were simultaneously considered, according to the results of this study, the approach of two-group categorization can be adopted.

Relevance to Industry

The result suggests that using psychoacoustic parameters can accurately assess the subjective perception of music quality. This information is particularly useful for electroacoustic industry. The prediction models developed in this paper can expedite the development speed of sound reproduction products by avoiding excess testing time and efforts.

1. Research background and motivation

Although music quality is influenced by numerous factors such as the recording and audio compression technology (Fazekas and Sandler, 2007; Kowalgin and Gamage, 2002), music quality is primarily determined by subjective feelings (Usher, 2006) and some research use music samples as an emotion measurement scale (Lu and Petiot, 2014). Because people's tastes in music differ, during the process of developing electroacoustic products such as earphones or loudspeakers, a conventional approach used to measure people's subjective perception of the newly developed products is to conduct a jury test on a group of experiment participants for data collection; however, jury tests require considerable costs, including those of effort and time (Kahana et al., 1997). One of the current trends in the electroacoustic industry is to increase the product specifications to meet the needs of consumers while reducing the development time and costs. A quick response time becomes an important competitive advantage for electroacoustic industry, like many other industries. Therefore, in the electroacoustic industry, establishing a model that can rapidly and accurately assess users' preference for the sound quality of music is crucial. Based on the concept of auditory roughness proposed by

* Corresponding author.
E-mail address: khtang@fcu.edu.tw (K.H. Tang).

http://dx.doi.org/10.1016/j.ergon.2016.01.004
0169-8141/

Vassilakis (2005), this study developed and validated a quantitative model, the tonal harmony level (THL), that can effectively predict people's subjective perceptions of music quality. Experiment using both ordinal and interval scales were adopted for verification during in the model construction process to enhance the robustness and validity of the measurement results (Rossi et al., 2005; Poirson et al., 2010).

The physical vibration of sound waves stimulates the human auditory system, resulting in the sense of hearing, which closely relates to individual physiological conditions and psychological statuses. Because the same sound exerts distinct psychological effects on people, the study of psychoacoustics has been developed (Fastl and Zwicker, 2007). By describing the ear structure and applying auditory physiology, researchers have constructed models that can depict people's auditory perceptions to quantify their subjective feelings of sounds and reflect the differences in subjective hearing perceptions. Common parameters used in psychoacoustics are loudness, sharpness, roughness, and fluctuation strength (Fastl and Zwicker, 2007).

Numerous studies have developed sound quality assessment models by using psychoacoustic parameters and regression or other quantitative models, such as neural network models. The sounds involved in the assessment models included noise, traffic noise, and sounds caused by machinery appliances. For example, Raggamet et al. (2007) conducted a jury test to investigate the relationship between the subjective perception of traffic noise and the objective psychoacoustic and physiological parameters. The psychoacoustic parameters (i.e., loudness, roughness, sharpness, sound level, tonality, and fluctuation strength) and physiological parameters (i.e., heart rate and respiratory rate) in each circumstance were measured. Regression analysis results suggested that the subjective assessment was highly correlated with loudness, roughness, sharpness, and sound level; the correlation between the subjective assessment and the heart rate was also significant. Jeon and Sato (2008) assessed people's subjective perceptions of the noise produced by household refrigerators by adopting a semantic differential scale. Linear regression analysis was performed using psychoacoustic parameters and subjective perception scores. The results implied that loudness, roughness, and fluctuation strength were significantly correlated with the subjective scores, and thus, the parameters were used to construct an index to predict the subjective sound quality of household refrigerators. Yoon et al. (2012) conducted a jury test by adopting the semantic differential scale to investigate people's subjective perception of air conditioning noise in cars. The analysis involved using linear regression and neural network models, revealing that loudness, sharpness, and roughness were highly correlated with the subjective perception. Wang et al. (2013) conducted a jury test to examine the relationship between the auditory perception and roughness of car noise. In addition, Moon et al. (2015) conducted a jury test to determine the detectability of ringtones by adjusting frequency and sound level. Overall, in most of the models created to evaluate subjective sound quality, psychoacoustics has been applied to determine the subjective perception of noise rather than to measure how pleasant the music was to the ear.

According to Western music theory (Helmholtz, 1885), Vassilakis (2005) reinterpreted the notion of auditory roughness to examine the sound harmony perceived by the human ear. Fig. 1 displays the standard curve for the consonance and dissonance obtained empirically by subjective rating of pairs of sine waves (Plomp and Levelt, 1965). For example, assuming that the two monotones were at frequencies of f_1 and f_2, when $f_1 = f_2$, the frequency difference was 0 (at unison) and the consonance level was maximum at 1. When f_1 remained the same and f_2 gradually increased, suggesting the frequency difference increased, the consonance level of the two monotones initially decreased, and then increased up towards, but never quite reached the consonance of the unison.

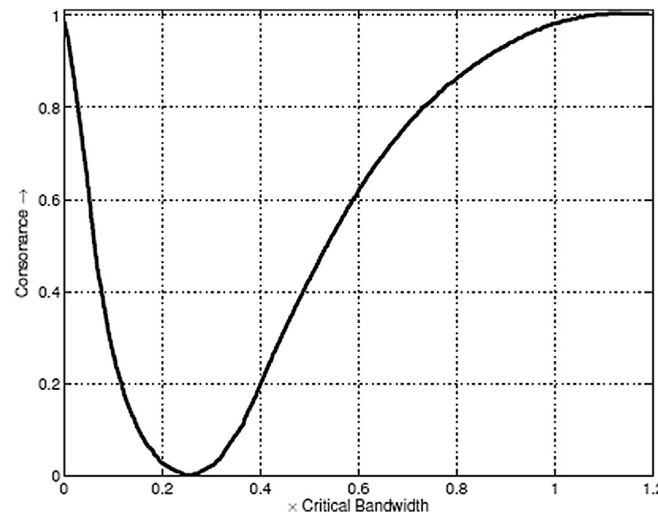

Fig. 1. The standard curve of two monotones at dissimilar frequencies; the X axis indicates the frequency differences measured by critical bandwidth (Plomp and Levelt, 1965).

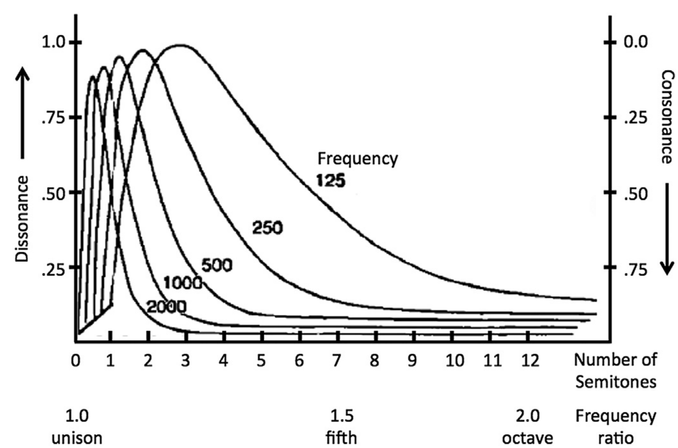

Fig. 2. The consonance and dissonance of two monotones with dissimilar basic frequencies; the X axis indicates the number of semitones, and the Y axis represents the consonance and dissonance (Sethares, 1993).

Sethares (1993) employed the Plomp-Levelt curve to investigate the consonance and dissonance of a chord of two monotones at dissimilar frequencies (i.e., f_1 and f_2). From Fig. 2, for example, when $f_1 = 125$ Hz, the chord was perceived as dissonant when the tone pair was five semitones apart. By contrast, when $f_1 = 2000$ Hz, the chord perceived as consonant when tone pair was only three semitones apart. Figs. 1 and 2 suggest that the degree of sensitivity of the human ear to dissonant chords at low and high frequencies differed. Therefore, Sethares (1993) proposed a model that could be used to calculate the level of dissonance of a chord comprising two monotones at various frequencies, as expressed Eq. (1).

$$d(x) = e^{-3.5x} - e^{-5.75x} \quad (1)$$

where d(x) is the level of dissonance of the two monotones, and x can be calculated using Eq. (2).