**Consonance and dissonance perception**

LoPresto, 2009:

Consonance involves musical intervals with smaller number frequency ratios (adjacent or lower harmonics), while dissonance involves larger number frequency ratios (higher or non-adjacent harmonics).

The western musical scale uses the octave, which is divided into 12 equal semitones.

When a human ear hears two pure-tone (single-harmonic) musical pitches at the same time, the eardrum vibrates simultaneously at both frequencies. The brain perceives this as a pitch that is the fundamental of the harmonic series to which both of the sounded pitches belong, whether or not the fundamental is really there. The quality, or, as musicians call it, the timbre, that is heard is caused by the form of the combined wave, which is a result of the harmonic spectrum. In this process, known as fundamental tracking, the ear is performing a Fourier transform.

When the frequencies of two pitches are too close to eachother, an unpleasant “beating” sensation occurs. Complex musical notes sound dissonant if there are beats or roughness between their higher harmonics. If their fundamentals are separated by smaller whole number intervals, their higher harmonics will coincide, meaning they’ll sound consonant. Roughness occurs in adjacent harmonics of a ration smaller than 6/5 (a minor third).

Plomp and Levelt have showed that if two pure tone pitches fall within a critical band for which the ear cannot distinguish different tones, they sound dissonant. The critical bandwidth is 1/3 of an octave (a minor third). Maximum dissonance occurs at about ¼ of the critical bandwidth.

Crucially despite consonance being considered musically “good” the same cannot be said about dissonance. It is not necessarily “bad” musically as it can evoke moods or emotions.

Stefano et al., 2022:

Big ole critical review V useful

LoPresto, 2015:

The musical intervals of just intonation are called major intervals except the fourth, fifth and eighth (octave) which are called perfect. These intervals make up the major scale. They are not of equal step between them some being semitone and some tone differences.

When two complex tones are sounded simultaneously and the frequencies of adjacent harmonics are within less than a ratio of 6/5, an interval of a minor third, a “roughness” is perceived that is caused by beats between the two frequencies. This is the cause of the sensation of musical dissonance, an unpleasantness perceived by the human ear. The opposite sensation, musical consonance, is generally considered to be the absence of dissonance.

Dissonance and consonance can be judged based on smoothness, purity, blending and fusion. Intervals with less beating are considered smooth. Purity is the number of frequencies present that are harmonics of the fundamentals of both tones. Blending is a more subjective sense of whether the notes “belong” together. Fusion is if the notes combine to sound as one.

A more quantitative measure can be the sum of the numbers in an intervals ideal ratio. The ideal ratio is the fraction that has the closest decimal equivalent to the frequency ratio of the interval. The sum dissonance metric for an interval is the numerator plus the denominator. The physical basis for the sum dissonance metric is that intervals with small whole number ratios, having a smaller sum, will have more coinciding upper harmonics and be less dissonant.

Trainor & Heinmiller, 1998:

Di Stefano, Vuust & Brattico, 2022:

Consonance is usually perceived as smooth, harmonious and with positive valence, while dissonance is rough, unstable, and of negative valence.

According to different fields and schools of thought, arithmetic accounts of consonance believe it results from the combination of sounds that can be expressed by simple integer ratios while more complex ratios create dissonance. According to psychoacoustics, consonance emerges from the combination of several factors like roughness, fusion, tension, and harmonicity. According to neurophysiology, consonance and dissonance are differently processed by the auditory system and the central nervous system, with discrimination within the subcortical and cortical levels. Another belief is that consonance and dissonance are nothing more than cultural shaping.

Psychoacoustics: Roughness is the sensation that arises from complex tones whose adjacent partials interfere with each other, “beating at frequencies between 20 and 300 Hz” is what has been identified as causing roughness. Roughness is, in turn, related to beating, the sensation that is evoked when harmonics of two tones sounding simultaneously are not sufficiently spaced far apart and therefore mutually interfere, giving rise to a modified waveform with a rhythmic oscillation in the vibration pattern or amplitude. Although the role of beating and its resolution by critical bandwidths in consonance and dissonance perception has been supported, evidence shows that it only plays a secondary role as shown by studies on dichotic listening.

Fusion is the sensory correlate of the merging of harmonic complex tones into a unique percept. It results from low level processes of sensory perception, which triggers a sense of unity in the listener. It can also be defined as the inability to discriminate between constituent tones of a sonority. However, fusion can be conceived as an active perceptual process that leads to the formation of a coherent whole. The most fused intervals are the unison the octave the perfect fifth and perfect fourth. Results from experiments have shown that consonant intervals are perceived as most fused, therefore are more likely confused as single tones. However, due to contradictory results its difficult to reach a conclusion about fusion.

Tension expresses the degree of instability that characterises auditory perception. In music tension is connected to expectancy and relaxation, with tension increasing the expectation for resolution which is perceived as relaxation. Tension is affected by roughness, sensory consonance and timbre.

Sharpness describes the proportion of energy at high frequencies and is essentially influenced by the spectral content and bandwidth. Sharpness affects sensory pleasantness with sharper sounds, meaning those of higher frequency energy sounding less pleasant. Sharpness tends to increase with frequency, frequencies in the 2 to 4 kHz range sound the most annoying. One study showed that sharpness impacts perceived consonance and dissonance at a higher register, while roughness is responsible for the decrease in consonance at the lower register.

Harmonicity is a measure expressing the extent to which a sounds spectrum is similar to a harmonic series for example with a fundamental frequency f and partials at 2f, 3f etc. It has been found that although congenital amusics retain aversion to beats, they are unable to discriminate harmonic and inharmonic complex tones, suggesting that the deficit in periodicity perception causes their failure to express a preference for each consonant or dissonant chords. It has been found that roughness, harmonicity and familiarity all interact to cause consonance and dissonance perception. With some suggesting contribution of only cultural effects and being criticised for it. Taken together, these studies suggest that harmonicity affects Western listeners’ preferences independently of the tuning system. However, it remains open the question about the role of culture in the perception of harmonicity, with studies demonstrating that non-Western listeners failed to show clear preference for maximally consonant (and harmonic) chords, rather showing a preference for chords that were more familiar to them.

Culture: several studies raised issues with respect to the cross-cultural appreciation of consonance. For example, Maher [134] found that Hindustani listeners are less sensible to highly dissonant dyads that Western listeners. More recently, in addition to the already discussed study by McDermott et al. [144], subsequent studies fail to clearly demonstrate that non-Western listeners prefer consonance over dissonance ([10,122,209]). Finally, it might be worth mentioning here the existence of beat-diaphony music cultures, in which dissonance in rough intervals is promoted for aesthetic purposes.

Preference is typically measured asking participants to evaluate the attractiveness of perceived stimuli through Likert-scales. However, such methodology has limited sensitivity, reliability, and validity (see e.g., [180]), and is demonstrably affected by cultural differences in interpretation and use [49]. In fact, literature showed that respondents styles may vary with culture, with participants more prone to use extreme values of the scale while others that tend to choose average values.

Formulated by Bowling and Purves [26], the vocal similarity hypothesis (VSH) considers the attraction to the harmonic stimuli as fundamental for the biological foundation for consonance. This hypothesis is grounded in critical evidence on audio-vocal communication and on the prominent role of harmonicity in vocal communication. Crucial for VSH is the fact that most frequent intervals in the spectra of speech parallel consonance ranking, showing that the intervals we perceive as consonant are specifically emphasized in vocal spectra [203]. Together with other evidence (e.g., [80]), these findings have suggested the idea that the most plausible explanation for consonance might be an evolved attraction to the harmonic series that characterizes conspecific vocalizations based on the biological importance of social sound signals.

In contrast to VSH, the psychocultural hypothesis relies on a wider set of explanatory factors that are assumed to have equally determined consonance perception. For example, the difference between theoretical intervals and how they are typically played in performances; the further distinction between intonation in solo vs. accompanied performances (either sung or instrumental); psychological factors of interval size perception and production; and the historical evolution of musical instruments and practices. In such a view, being “cultural” means that intervals are subject to learning and exposure effects and that such effects are robust and consistent. This was clearly demonstrated by Loosen [132], who showed that violinists preferred something approaching Pythagorean tuning, pianists preferred the 12-equal division of the octave, and nonmusicians had no significant preference. Similar findings suggest that tuning preferences are primarily learned, and not primarily determined by the acoustics of the sounds themselves or by the physiology of hearing. However, such a hypothesis risks rigidly conceiving the biology vs. culture dichotomy in terms of innate vs. acquired abilities, implying that rules of any musical traditions are, at least to some extent, arbitrary (e.g., [45,47]). However, as noted above (Section 5.3), given that enculturation is a biologically constrained process and that auditory biology is shaped by cultural experience, it seems misleading to conceive of the influence of culture as ‘arbitrary’, as well as the discrimination between C/D as ‘innate’.

A third hypothesis stresses the role of sensorimotor properties of vocal production, and in particular singing, as constraining the allegedly universal features of musical scales [198]. Stressing the role of motor actions for tuning sung intervals, SH holds that the origins of tonality must be rooted in the tuning of melodic intervals when sung rather than in any theorizations or compositional practices. Based on the evidence showing that intervals’ size of sung intervals is larger than predicted on the basis of Pythagorean music theory ([172]), SH conceives intervals as imprecise “islands” of close frequencies often overlapping one another rather than as sharp integer ratios based on abstract rules.