The author introduces some problems of music perception, but in a very cursory way. The basic question Longuet-Higgins asks is, How does a person use cognitive processes to sort out the elements of music concerning rhythm and tonal relationships? Unfortunately, the author admits to having an expertise only in Western tonal music. The author doesn’t really define this well, saying that “tonality” refers to a dependence on musical intervals—but all music depends on intervals, in a way. Non-Western/tonal music must be surmounted at some point. The reasoning that many authors have as to why we should first focus on Western tonal music is that it is structured in a recognizable and reproducible way. However, this is a potential confound for asking what "perception" is. Longuet-Higgins seems to think that perception is independent of interpretation or emotional response. A computer could easily be "musically competent" and able to perceive music, based on the author’s assumptions.

The author provides us with a system for breaking down rhythm into trees. The syncopation example (Figure 6) was questionable in accuracy, or at least was not clearly defined. While an attempt to clarify and simplify rhythmic structure is lauded, there are some fallacies. For instance, the author mentions that an organ performer has "no control over the relative loudness of successive notes." Since we are in the realm of perception, we must take into account the fact that organists use legato versus staccato articulation as a perceptual tool of designating accents, which imply a sort of loudness or softness in the listener’s mind.

In discussing interval distinctions, there is a minor issue. The author says that the major third can be tuned “to the satisfaction of every other musician,” but this is not as simple as it seems. I think there would be a discrepancy in many ears as to whether the major third sounds sharp, since it is in fact sharp in the equal-tempered system.

Longuet-Higgins’ thought experiments on the naming of pitches leave something to be desired. The listeners in question, who are familiar with notation systems, may in fact just be naming the pitches in terms of voice leading conventions. For instance, sharped notes tend to go up, while flatted notes tend
to go down. So, if the example given ends on an upward motion, the listener would be more inclined to
name the penultimate note by its sharp name, rather than the enharmonic flat.

The concept of tonal space was straightforward, and the idea of remoteness was fine, except the
linear model of remoteness does not account for why twelve half-steps away would sound much closer than
11 half-steps away. The octave is important to consider, but it is glossed over in this model.