Harmonic Considerations in Musical Theory Regarding the Phenomenon of the Wolf Fifth

Floris T. van Vugt
University College Utrecht, The Netherlands

Abstract

In the old day's the organ and the clavecimble (an ancestor of what is nowadays called the piano or grand piano) were tuned in the way that would be harmonically desirable in the sense that the most of the smaller intervals were tuned precisely the way they are defined. However, it turned out that this yields a discrepancy in the fifth circle, which is achieved by transposing harmonic scales. In order to account for this discrepancy, clavecimble tuners would make one fifth in each octave out of tune. The organ players had to avoid playing it and composers were taught not to include it in their music.
In this paper the question whether this Wolf Fifth is a physical phenomenon or a result of our theoretical definitions of the musical theory will be addressed. First some of the concepts used will be clarified in a section on the theory of sound. Successively, this knowledge will be applied to discuss the classical temperature as it has been common since Pythagoras. Finally, the innovations of the new temperature will be discussed and the main question will be answered using the material presented so far.

1. Introduction

In this paper, an attempt will be made to unveil the fundamental nature of the phenomenon that causes the wolf fifth. The question whether it is a physical phenomenon or is directly caused by the definitions in the theoretical musical system will be addressed.

In order to do so, first of all the boundary conditions that result from conventional western musical theory will be discussed, and additionally, a link will be established between, on the one hand, the physical properties of the sounds that are involved in music, as well as their properties that are of interest to the solution of the problem presented here, and, on the other hand, the musical intentions of the creative genius and the musical theoretical framework that was meant to provide the theoretical framework enabling the composer to systematically communicate his musical intentions from person to person and from era to era.

A critical reader might argue that the approach that has been chosen in this particular paper places too great an emphasis on pinpointing the definitions of the terms that are at stake. However, it will be held here that precisely stating the definitions of terms is (1) exactly what is missing in many standard texts in this field of subject and (2) this approach will more accurately enable the reader to spot the underlying cause of the discrepancy that manifests itself in the Wolf Fifth.

The research question that is to be identified at the roots of this paper are relevant and can, even though the particular approach of this paper is more mathematically oriented, be linked to general fields of physics. For instance, the subject matter is an example of how the physical theory of sound can be called in to assist the musical mind.

Also, the answer to the main question can help in deciding whether our musical theoretical system suffices in this sense. Furthermore, in practice, instrument builders will also be able to let themselves be guided by musical theoretical considerations in their everyday labour. Also, as will be evident in the end, piano and organ tuners are required to take the considerations that will be presented here at heart in order to be able to live up to the musicians' expectations to the largest possible extent.

2. Methodological Notes and Terminological Examination

First of all, let us start out by noting that music starts with music. As this might seem obvious upon first examination, a more critical approach will reveal that many, musicians and scientists alike, have considered a physical description of the phenomenon of music as more fundamental than the artistic creations of some who used the medium for expressive purposes. However, in this paper it will be assumed that this approach fails to account for those aspects of musical theory that potentially are the ones which should be identified as constituting the basis of the problem that is at stake here. Therefore the method that will be used to address the question here takes as its starting point the creative intentions of the musical mind and will try to explain the related physics.

Upon first inspection it will be evident that the unit of music is the sound. Music without sound, as will be held here, is a *contradictio in terminis*. In early times the scale of sounds that were used in music was rather limited with respect to its current extension, before including a limited range of tones produced on a limited number of instruments, it nowadays is in fact not limited but by the musical creator it serves. In this discussion all the individual thinkable sounds that are being used in music, whether in any particular musical period or era, by a specific composer, or in all the music that has been so far created, will be referred to as *tones*. In this definition, every tone is a sound, but not vice versa. Similarly, the whole of all thinkable tones will be referred to as *tonal system*. Historically this tonal system is a dynamic unity to which tones are added or disappear continuously.

An important aspect of the tone is the pitch. Essentially not every tone (and similarly also not every sound) consists in a base pitch with which it can be identified; nevertheless a tone can be represented as a potentially and theoretically unlimited number of different vibrations.

Similarly, a difference in nature between tones, irrespective of a possible variation in pitch, can be distinguished by the human ear, with differing degrees of success. This nature of the tone that is inherent in each of them, in order to be classified as a sound, will be called 'timbre' here. It is this quality that distinguishes the sound of a note played on the piano from a note of an equal pitch played on any other instrument. However, on a single instrument, capable musicians can create more subtle variations in timbre.

Also, the ear is capable of distinguishing between several notes of equal timbre, by their loudness. Therefore volume must be identified as a property of the tone, separate from the timbre.

Furthermore, there can also be a difference between tones, which the ear recognises as not corresponding to a difference in the timbre or in the volume between tones, but due to some other physical property. This difference then must be a difference in pitch. The two notes, when sounding together, are called a *chord* and the difference in pitch between them is called an *interval*.

The components of tones can thus be identified as being the pitch, volume and the timbre. However, as the physical examination that is to follow will reveal, the three are not as independent as they might occur at first glance.

References to tones also occur in a more abstract form both in the mind of the musician and in the musical notation. Precisely this abstract reference will be called *note*. Like a tone, also a note pertains to a tone of a certain pitch, abstractly assumed to be played on a specific type of instrument.

3. Music Theoretical Background

A recurring element in many tonal systems that exist nowadays in the world, is the octave. An octave is used to denote the interval between a pair of tones, which is such that individual tones, equal by approximation in timbre and volume, occur as more similar than different. Therefore, notes that differ by one octave only are referred to with the same name, i.e. c, but usually it is added either by notational primes or in words that it originates from another octave.

The very backbone of our tonal system is the tonal scale.

This scale is

(1) a collection of a theoretically unlimited number of notes, differing only in pitch,
(2) which can potentially be played on any instrument for as far as its particular pitch range allows for it (since the scale is in principle unlimited, but the pitch range of the instrument necessarily limited, the restrictions are imposed by the instrument rather than the scale), and
(3) of an arbitrary base pitch.

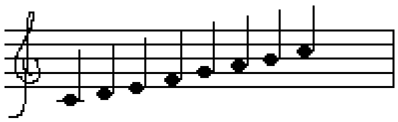
This is to say, that (1) the scale is defined by the intervals between the notes rather than the pitches of the individual notes, and that (2) for each thinkable note, a scale is thinkable that starts out from its pitch. This first assertion is exemplified by the fact that the standard pitch frequency of the middle “a” in concert settings has varied through the ages between numbers deviating at least as much as 435 Hz and 450 Hz.

Additionally, it should be noted that hardly any scale, although each individually unlimited in range, contains all thinkable notes. For instance, the minor scale contains notes that the major scale does not.

Since scales involve a theoretically unlimited number of pitches, but the tonal system is periodic, it follows that the scale can be reduced to a limited, or even very small, number of notes that recur over the octaves.

In the time that the first theoreticians of music started their demanding labour, the most fundamental scale, called the major scale, had already attained to a large extent its current form.

Identifying a particular but arbitrary note as the base, the scale for a particular octave can be drawn as follows (the interval between the lowest and highest note is an octave):



As notation is also arbitrary in this matter, there will be no explanation or elaboration of its details in this paper.

This sequence of tones is referred to as the *base tones*.

The scale, being notated in this manner, is misleading in the sense that it portrays the scale as if the intervals between the notes are equal. However, there is not merely a theoretical of physical difference between the notes; the human ear will experience little difficulty in distinguishing the interval between the first two notes, for instance, from the interval between the third and the fourth.

This discrepancy is also revealed when considering the fact that there exist derived tones for some of these tones, but not for others. These tones are historically referred to as 'coloured' versions of the tones from which they stem, which in practice means that they are either lower or higher than the original pitch. These coloured notes are therefore referred to as "chromatic" notes and do not necessarily appear in the standard major scale starting from the c. These therefore give rise to the chromatic scale, which includes generally all thinkable notes.

In the examination here, *chromatically derived* tones are called *sharp* if they are obtained from increasing the pitch of the original note, and *flat* if they are lower than the original note. In this sense chromatically derived tones can be contrasted with the base tones.

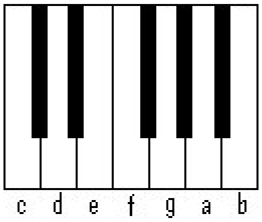
The interval between a coloured note and the original is called a *semi-tone*.

Now several essential aspects of our tonal system will have to be discussed:

(1) In our musical system, it is impossible to distinguish between a sharp note on the one hand, and the flat note derived from the higher note on the other hand. Therefore they are called *enharmonically* equivalent.

(2) Furthermore, the coloured versions of some notes are notes that do appear in the major scale. This happens when lowering the f or the c, or increasing the pitch of the e or the b.

This then gives rise to the following picture of one octave as it appears on the piano (where the white keys are the base tones of the major scale starting at c, and the black keys the chromatically derived tones):



This tonal system is called the *diatonical tonal system* and it divides the octave in twelve comparable intervals, i.e. semi-tones. This might again mislead into thinking that semi-tone intervals are equal, however, they are not entirely, but rather *almost*. For the following discussion it will be assumed that they are equal, as this will turn out to be a fundamental assumption of our musical theoretical system.

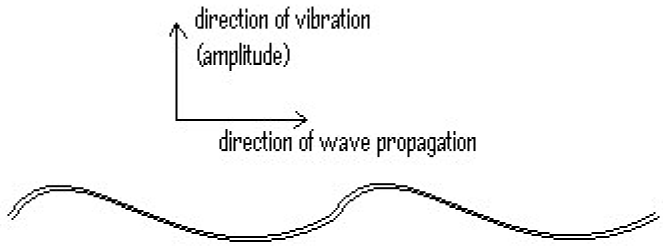
As the scale is defined only as a collection of intervals, we can start a scale from any note. Furthermore, since it is desirable that all scales sound identical (one would like all the intervals in corresponding scales to be equal), and since it is also desirable that all notes in the scales are one of the twelve tones that are initially used in our tonal system, names are given to specific intervals. The interval of *c-g* and likewise every interval of 7 semitones, is called a *fifth*. The interval of *c-e* and likewise every interval of 4 semitones is called a *major third*, the *minor third* involving 3 semitones. The series of interval *c-e-g-c* is the major triad and generally consists in a note, the third, fifth and octave intervals and this is called the *major triad*.

4. Physical Examination

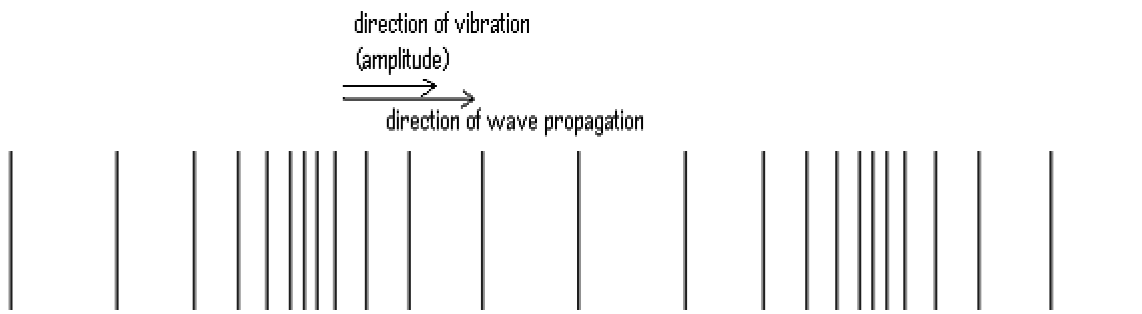
The starting point of the music theoretical framework is, understandably, the sound. The sound necessarily involves, for physical reasons, a vibration of some sort. Musical instruments are devices meant to generate notes by starting a vibration in a column of air, on a string or by any other means.

Physically speaking there is a fundamental difference between the two examples mentioned above, in the sense that a vibration in a column of air is typically a longitudinal vibration, and a vibration on a string a transverse vibration. The essential difference is that in a longitudinal vibration, the direction of the vibration is parallel to the direction of the wave propagation, whereas in a transverse vibration, the direction of the vibration is perpendicular to the direction of the wave propagation.

Transversal



Longitudinal



Since sound takes the air as its medium, the air particles are the ones vibrating and therefore the direction of the vibration is the direction of the vibrating movement of the individual particles. The concept of vibration implies that even though a wave may seem to move, the individual particles that constitute it remain in place.

Furthermore, sound waves in the air are longitudinal vibrations and therefore these cause differences in the density of the air. If we would then plot the density of the air particles in a certain part of the air, this would yield a wave function.

Therefore, every vibration can be represented through a more or less complex periodic function.

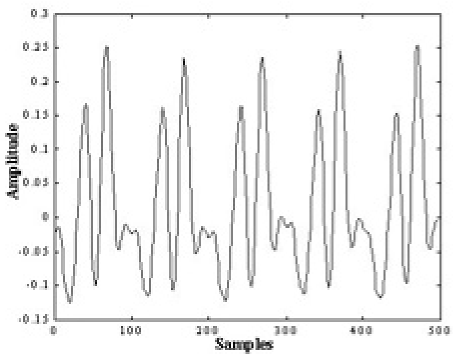
Essential properties of each wave function are (1) amplitude and (2) frequency, the first being defined as the maximum deviation from the average density or position of the particles, and the latter being defined as the number of vibrations in a certain interval of time. The standard unit of the frequency is Hz, the unit of the amplitude can be any unit corresponding to the spatial deviation or deviation in pressure, for instance meters.

$$f = \frac{vibr}{time} = \ldots s^{-1} = \ldots Hz$$

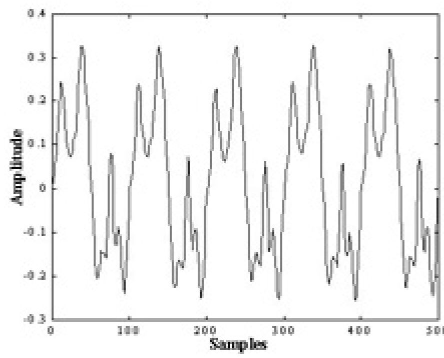
$$A = x_{\max} - \bar{x} = \Delta x_{\max} = \ldots .m$$

This representation is slightly too simplistic in the sense that any sound occurring in nature is rather an addition of several distinct vibrations of varying frequency and amplitude. This results in sum waves that physically characterise various instruments and allow the ear to distinguish between them. This is illustrated below.

Oboe



Harmonica



This physical difference between sounds generated by various instruments will be referred to as the "wave content" of sounds and enables the ear to distinguish timbres.
In this paper the attention will be confined to the harmonic aspects, that is, the intervals, which is why for convenience purposes can be chosen to only consider simple sine waves. The results will of course equally well hold for more complex wave contents.

5. Interval Definitions After Strict Interpretation of Musical Theory

Now some general connections can be established between sound experience and physical properties of the sound.
The experience of the tonal "height" of tones is determined by the physical property of the frequency. The idiosyncratic property of octaves is that the frequency of the total vibration of the one tone equals the double of the other, lower, tone, which explains why sounds, separated by this interval only, sound so much alike.
Bearing all this in mind, it might already be noticed that the restrictions that would strictly be imposed by this musical theoretical system are already defining all the intervals between the notes.

The following assumptions can be strictly ascribed to the initial formulation of the musical theoretical system as it has been introduced so far:

(1) the interval of the octave must correspond to a ratio of 1:2 in the frequencies.

(2) starting from any note, a scale can be composed, consisting in the same intervals as those in the base scale (which will here be referred to as the *transposition condition*), and

(3) each of the notes in this scale must be present in the chromatic tonal scale (which will be referred to as the *chromatic condition*).

It follows immediately from these assumptions, that the frequency ratio of the intervals between each semi-tone should be equal.

This then leads to the following equation for a semi-tone interval

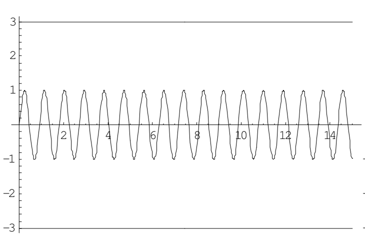
$$I_{semitone}{}^{12} = \left(\frac{f_{x\#}}{f_x} \right)^{12} = 2 \Rightarrow I_{semitone} = 2^{\frac{1}{12}} \approx 1.059463094$$

In this wave setting all intervals would be equal and therewith the most important condition, the transposition condition, will be met. Finally, an essential fact is that this value for the semitone interval is the *only* that satisfies all conditions mentioned above.

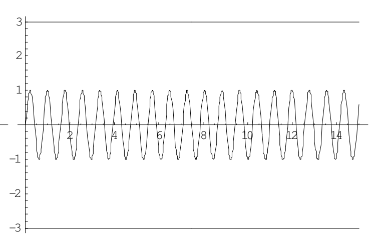
5. Interval Definitions After Musical Intuition

If two notes (Y_1 and Y_2 below) separated by a certain interval sound together, their amplitudes at different times are added, and it becomes possible to distinguish a new periodical function.

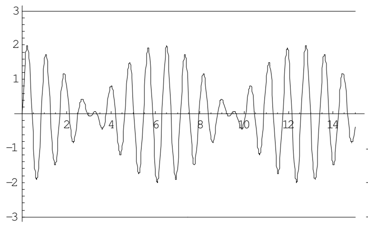
Y_1

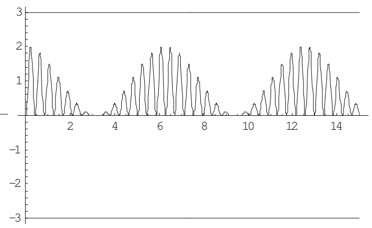


Y_2



Y_1+Y_2 **Intensity $(Y_1+Y_2)^2$**





If the first two vibrations are described as sinus waves and we assume them to start with no phase difference, we can write (using $f_x(t)$ to denote the phase of any sine graph x, that is, the number of times the graph went through its period at a certain time t, and $f_d(t)$ the phase difference between the graphs):

$$
\begin{aligned}
&y_1(t) = \sin at && f_{y_1}(t) = \frac{at}{2\pi} \\
&y_2(t) = \sin bt && f_{y_2}(t) = \frac{bt}{2\pi} \\
&y_{1+2}(t) = \sin at + \sin bt && f_d(t) = \frac{a-b}{2\pi}t
\end{aligned}
$$

Solving for $f_d = 1$, we find:

$$t_d = \frac{2\pi}{|a-b|}$$

Which denotes the time after which the phase difference is one, therefore this time interval is equal to the period of the sum graph.

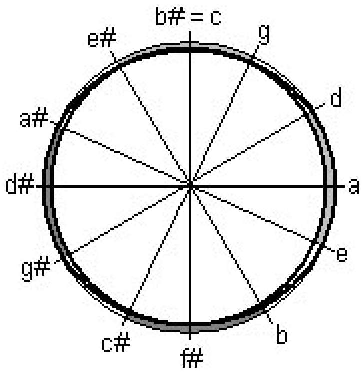
This reveals that once the frequency difference becomes small enough, it will eventually become impossible to perceive the amplitude variations (because t_d becomes very large). This explains why some intervals are experienced as comfortable whereas others are not; the first are called *pure* intervals and correspond generally to simple ratios of the different frequencies. Furthermore this immediately reduces the intervals that can be used in music to a limited number.

Pythagoras was the first to provide a systematic analysis of the scales that were conventional at his time. His enterprise started in pinning down the ratios of the string length between pure intervals, on the basis of two assumptions, of which the first can be stated in two equivalent ways:

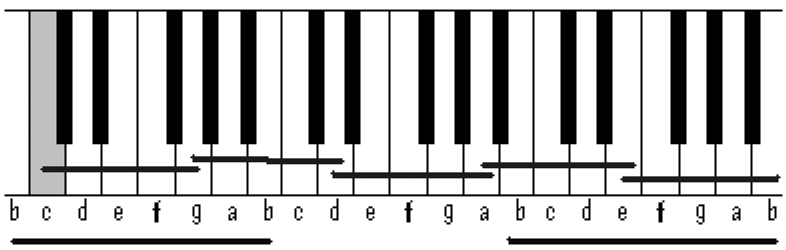
1a. The tonal system must be periodical; its period must be the octave.

1b. Every octave interval must be equal and must be 1:2.

2. Every fifth interval must be equal and must be 2:3.



It can now be verified what the semitone interval is in the Pythagorean tonal system. If one tunes up five fifths from an arbitrary base note, and subtract 3 octaves, the ratio between the newly found note and original note is a semitone.



But here follows a discrepancy with the musical theory:

$$I_{semitone} = \frac{\left(\frac{2}{1}\right)^3}{\left(\frac{3}{2}\right)^5} \approx 1.05349794 \neq 2^{\frac{1}{12}}$$

This discrepancy immediately shows that the Pythagorean tonal system cannot possibly comply with both the transposition condition *and* the chromatic condition.

Furthermore, starting from an arbitrary base note, in this case the *c*, tuning up 12 times one fifth, makes one end up at the b#, which is enharmonically equal to the *c*, 7 octaves higher than the starting *c*. However, the ratios do not correspond:

$$\frac{\left(\frac{3}{2}\right)^{12}}{\left(\frac{2}{1}\right)^7} \approx 1.01364 \neq 1$$

In the Pythagorean tonal system, this discrepancy, known under the name *Pythagorean comma,* or *ditonic comma*, was stored in one specific fifth interval, between the g# and d#. This interval became a very unpleasant surprise to the ear and therefore many composers and organ players justly considered it unusable. Its howling sound eventually was given the name ‘wolf fifth’ or ‘organ wolf.’

Additionally, tuning up 4 times a fifth makes one end up at a third two octaves higher than the original ones. According to Pythagorean standards a third interval corresponds to a frequency ratio of 3:4. Since all octaves must be equal (assumption 1) the following ratio should be one again, but is not:

$$\frac{\left(\frac{3}{2}\right)^4}{\left(\frac{2}{1}\right)^2\left(\frac{4}{3}\right)} \approx 0.949219 \neq 1$$

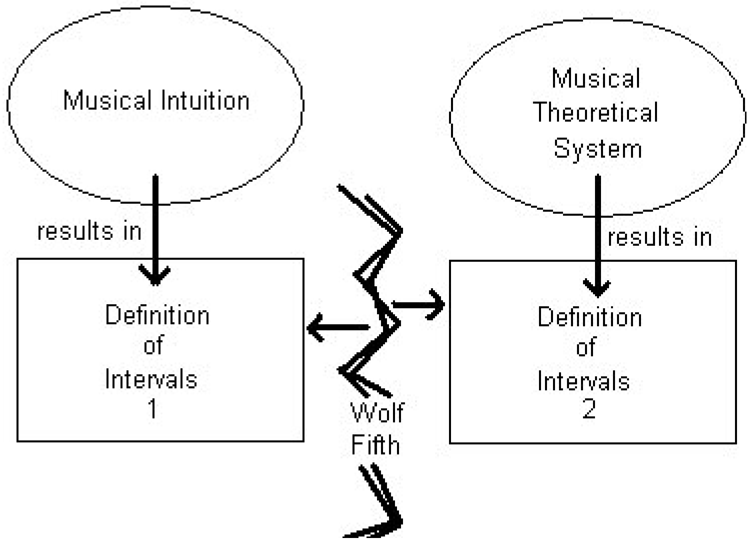
Which interval is called the Pythagorean third and sounds awful as well. The ratio is called the *syntonic comma*.

Around the 17th century, it became conventional to not store the entire discrepancy of the tonal system into one interval, but rather spread them out over all possible intervals. This resulted in piano tuners being forced to tune each interval slightly off the desirable tone, in order to avoid intervals such as the wolf fifth. The resulting interval was the mathematically more desirable $I_{semitone} = 2^{\frac{1}{12}}$ which sounds less comfortable, for obvious reasons.

6. Conclusion

In conclusion, it is evident that the western musical theoretical system is contradictory. The contradictory elements are on the one hand the intuitive notions that a limited number of intervals should sound in a particular way and not any other, and on the other hand the theoretical system that deploys 12 tones in the tonal space and assumes the equality of their intervals.

The problem can be represented as follows:



The definition of intervals that is generated by our musical intuition – those can be understood in physical terms of wave interference – is incompatible with the definition of intervals that results from our musical theoretical system – which can be understood by simple mathematics.

In reality, it would therefore be theoretically desirable to fundamentally reform the tonal system in order to gain a consistent theoretical system that would enable the correct tuning of any instrument. This, however, has been aptly considered not feasible due to the rich musical history in the old tonal setting and the fact that all our current musical instruments are based on the conventional musical theoretical system and would therefore also have to be changed.

Sources

Craats, J. van de, Takens, F. 'De juiste toon, de juiste stemming." Nieuw Wiskunde Archief (2) 2001: 136-145.

Goddijn, A. *Hoe klinkt 4• 5?* Symposium Music & Mathematics. 20 Dec. 2001

Lewis, P. *Understanding Temperaments*. 1978, 1998. http://pages.globetrotter.net/roule/temper.htm

Lürsen, M.J. *Grondslagen van de muziektheorie*. Wassenaar: Servire, 1975.

Tipler, P.A. *Physics for Scientists and Engineers*. Worth Publishing, 1998.

Willemze, T. *Algemene Muziektheorie.* Antwerpen: Spectrum, 1971.