

I N T O
| | | |
S O U N D

On the practical analysis and processing of sound

Ton Wempe

Copyright © 2018 by Ton Wempe. All rights reserved.

Printed in The Netherlands

CONTENTS

Preface	v
Acknowledgements	vii
Introduction	2
The program "Praat"	3
How to run scripts in Praat	4
PART A. UNDERSTANDING BASIC SIGNAL ANALYSIS	6
1. The nature of sound	6
2. Decibel	10
3. Waveform, frequency, spectrum	12
4. Fourier series	16
5. Sine wave basics	22
6. Fourier transform	26
7. Resonator	33
8. Filtering	35
9. Continuous spectra	47
10. Multiplying signals	52
11. 'Negative frequencies'?	59
12. Convolution in the frequency domain	65
13. Windows	70
14. Convolution in the time domain	80
15. Noise	85
16. Correlation	89
17. Sampling of sounds	98
18. Digital filters	113

PART B. PRACTICAL CONSIDERATIONS	125
19. The principle of speech production	125
20. Formant measurements	133
21. Unvoiced speech	150
22. Prosodic features	152
22.1. Pitch	153
22.2. Intensity	156
22.3. Speech rate	159
23. Overall properties of speech signals	161
24. Sound data compression	166
25. Noise suppression	180
26. Musical sounds	183
27. Miscellaneous practical subjects	193
27.1. Praat's Spectrum	193
27.2. Frequency range	200
27.3. Signal clipping	204
27.4. Signal to noise ratio	210
27.5. Distortion	214
27.6. Microphones, acoustics	219
Appendix I: Fourier series	233
Appendix II: Complex representation of sinusoids	236
II.1. Complex plane	236
II.2. Complex exponentials	240
II.3. Complex Fourier transform	244
Appendix III: Laplace transform	250
Appendix IV: Z transform	256
References	262
INDEX	264

Preface

As an electronics engineer having worked at the language department of the University of Amsterdam for many years, I assisted researchers and students in many aspects of processing of sound, such as various signal analysis methods and automation of analysis and manipulation handling procedures. A great deal of their needs for assistance fell into the category spectral analysis, a field which most language researchers experienced as ‘difficult’. Generally, people with a background in linguistics have not had much training in mathematical and technical subjects. Sometimes, in the planning of a research, the focus might be shifted so as to avoid signal analysis techniques. Often, those analyses that could not be avoided will be carried out applying ‘blindly’ the default parameter settings of a suitable program like “Praat” [2]. These may work perfectly in many cases but will need to be adapted when specific sound material is analyzed. Also, adaptations need to be made depending on the aim of the analysis. This lack of technical background might also cause serious errors in the interpretation of the results of the analyses. Language researchers from other countries than my own (The Netherlands) have told me similar stories, so it seems that this problem might be quite widespread.

There are numerous books about signal analysis, varying from simple introductions to highly advanced mathematical works. Why then this book when there are so many already? It seems that the books available are either highly theoretical and technical and thus only suited for people skilled in mathematics, or oversimplified so that the background is not explained clearly and cannot be understood sufficiently. This applies especially to researchers in the humanities who have not had an extensive training in mathematics, which limits a person’s ability to understand most books on signal analysis. In addition, many books concentrate solely on digital signal processing and microprocessor applications, which would be a too limited area for the audience for which this book is intended.

In this book, therefore, I will try to explain the principles of manipulations, analyses and productions of sound that researchers who work in this area generally come across, rather than to try to cover all aspects of signal processing. The tools that will be used are first of all, the reader’s own common sense, and secondly, the versatile free program “Praat” (see the **Introduction** for accessing and using Praat). The mathematics used here hardly exceed secondary school levels. Possibly, this kind of simplification might be at odds with the mathematical correctness now and then. However, because the purpose of this book is getting insight in the signal processing mechanisms, that is what should have the highest priority. As long as the mathematical incorrectness does not influence the general truth of the underlying principles, the complications of a strictly mathematical approach are avoided in favor of clarity.

The book focuses on speech research. However, the basic principles of analysis used are valid for all kinds of sound analyses, or more generally, 'low frequency' waves analysis. Therefore, sounds ranging from animal sounds, musical instruments, singing voices, machine noise and heart beats to earth quakes, can be analyzed using the same principles. This is provided that the microphone or transducer used, and the sound input of the computer and other electronics used are able to process the specific signals adequately, and that the above-mentioned parameter settings are adapted to the specific properties of these sounds.

Some properties of speech sounds and musical sounds are dealt with in some detail in special sections.

Acknowledgements

A great number of people have had some influence, directly or indirectly, on the contents of this book. During the many years I worked at the Institute of Phonetics of the University of Amsterdam I had the great opportunity to learn from all of my colleagues there. In particular, the discussions with Paul Boersma (who, being a great scientist, gave always highly valuable answers), David Weenink, Louis Pols and, in embryonic times, Toni Rietveld, have formed my interests greatly about the subject of sound signal analysis. I owe special thanks to Dirk Jan Vet who made a great job in correcting my mistakes in the previous versions of the book, all from his solid knowledge and ability to think very logically. Naturally, the remaining mistakes (which cannot be eliminated completely, due to the exponential behavior of its number function) are completely on my account.

The readers' group, led by Mirjam Ernestus of the University of Nijmegen, had the patience to point out a lot of passages in my explanations which were too cryptic formulated for 'normal people' like linguists and thus guided me to make many clarifying modifications.

In addition, the questions of all kind from the students of the Institute gave me some insight into the subjects which should get special attention.

The program Praat, created by Paul Boersma and David Weenink, provided me with an ideal tool for generation of sound examples, all signal processing and *all graphic pictures* (except two) I needed for the book.

Finally, I would express many thanks to Marije van Wieringen who corrected the many mistakes in my, Dutch biased, English writing.

Introduction

This book consists of two main parts:

Part A deals with the basic principles of the analysis and manipulation of sound. The focus is on explanation through logical reasoning. Why do signals behave as they do, and what are the consequences of the various methods of analysis and manipulation? This part of the book tries to answer your fundamental questions about many aspects of signal processing. This is the theoretical part of the book, but all of it is embedded in a practical context.

Generally, for most studies in *exact sciences* a next step in learning a subject can only be taken when the earlier steps are understood: the new step is built on those that came before. Of course, it would have been possible to organize the book in such a way that the reader might choose a topic at random. However, the book would then have ended up containing a huge number of references to other parts. Instead, I chose the first option: new sections are based on the knowledge of the earlier ones. Nevertheless, many references to an earlier section are included.

The section on *discrete time signal processing* (aspects of sampling signals for computer processing) is postponed to the end of part A. Although most books on signal processing deal with sampled signals from the beginning, I will discuss the properties and analysis of the analog signals prior to the discretization, in order to avoid confusion of the sampling effects with the inherent signal's properties.

Part B contains practical notes that are important for knowing how to make good sound recordings, avoiding signal distortions or background noise effects, choice of equipment, avoiding mistakes, etc. An important ingredient of this part B is the explanation of the effects of parameter settings in programs for the most commonly used signal transforms, like “Praat”. Using the wrong settings for analyses can sometimes lead to inaccurate results! An especially tricky part is that the graphs may look accurate even when the output could still be misleading. It is important to know how to measure correctly.

Part B relies heavily on the information given in Part A and I would advise you to read Part A first. If you know enough already about basic signal processing principles you could skip this part, but, why miss all the fun?

Finally, in the Appendices a slightly more mathematical approach of some signal transforms and representations is presented. Here, the mathematics used is not very complicated but of course there are a few formulas there! For people with allergic

reactions to formulas with e.g. integrals there is no need to suffer: they can skip them. Part A without appendices offers sufficient information to understand the basic principles. For other people the appendices may serve as a step between the popularizing part of this book and other more mathematical books on the subject. These appendices may even offer you some more insights in the 'nature' of this sort of mathematics.

The program "Praat"

I would strongly advise you to download, install and use the free program "Praat" [2] from the site [www.fon.hum.uva.nl/praat] with which you can see and hear the effects described in the book. To run the demos in this book I made scripts for this program. The scripts are small programs that contain commands for Praat, and programming features, which make it possible to show effects of signal processing mechanisms and to play signals automatically.

The scripts can be found in the self-extracting file "scriptsIntoSound.exe" on the internet pages [www.fon.hum.uva.nl/wempe] or [www.audion.nl]. It is highly probable that the demos in the book will work as well with Praat versions that are newer than the one I used for testing, because the program is always kept downward-compatible. Instead of downloading the newest version of the program from the Praat site mentioned in the reference you could download the version from one of my pages to ensure that the demos and descriptions in the book apply to the same Praat version that I used to test all scripts.

The installation of Praat on your computer is very simple. The file "Praatxxx_winny.zip" (x for version number and y for 32 or 64 bit) produces only one executable file (Praat.exe). You can put it on your desktop or anywhere else on your hard disk.

Praat is a versatile program for signal analysis. Originally designed as a basic speech analysis tool, it has been developed to a universal signal research platform where all kinds of analyses and manipulations can be made on signals, with the flexibility to adjust all kinds of parameters. The possibility to produce high quality graphs and the intuitive script language make this program tailor-made for researchers in this field.

There is no need to *learn* using Praat for running the demos in this book. It is used only as a tool together with the demo scripts. The basics for getting the scripts ready to run will be explained, however.

There are Praat versions for Windows, Mac and Linux but the demo scripts for this book are only tested in the Windows version. Possibly, for the other platforms some minor modifications of the scripts may be needed, and if you need help with that, there is a Praat user group (<http://uk.groups.yahoo.com/group/praat-users/>) where you can

ask questions about scripting in general. There are some beginner's manuals downloadable from Praat's home page (<http://www.fon.hum.uva.nl/praat>).

You do not need to understand the contents of the demo scripts. For people who are interested in how the scripts work: although the structures are quite simple and are clarified by explanatory comments, the scripts sometimes contain lines that will not be clear during the stage the reader has reached at the moment the demo is being used. That will not cause any problems as the purpose of the demos is the demonstration of the effects, not an explanation of the way the scripts work. (Nevertheless, it can be interesting to make your own scripts, of course!)

For listening to the demo sounds you should connect loudspeakers or headphones to your computer's sound output. For laptops or tablets I would advise to use headphones/earphones or external speakers instead of the built-in speakers because of the highly limited frequency range of these tiny built-in speakers. Make sure that the sound mixer (Windows Volume Control) or your special sound software is adjusted for playing wave (.wav) sounds. Adjust the (Windows) Volume Control and Wave sliders to proper levels to avoid distortion, and be sure that eventual corresponding mute boxes are unmarked.

For using the demos, it is convenient to alter some of Praat's default settings, to avoid confusion by displaying more than what you are working on at that moment. This is taken care of in the special script for this purpose, called INIT.script. These new settings are preserved by Praat so that there is no need to adjust the settings after the next time you start the program. When the settings happen to be accidentally overruled, you could run the INIT script again so that the preferred settings are restored. The script also ensures the proper sound playing settings for all demo scripts (apart from the Windows settings mentioned above: they depend on the hardware used).

How to run scripts in Praat

If you have not downloaded and unpacked the scripts yet, please do it now and put them either in a new directory, or in the directory where Praat can be found. It is better not to use the desktop for all these scripts to avoid cluttering your screen with icons. When you have done that, run the Praat program. You will see two windows. The left window is called 'Praat Objects'; the right window is called 'Praat Picture'. In the Object window under 'Praat' select 'Open Praat script...'. browse to the directory where you put all the demo scripts, select the demo script mentioned in the book and then click 'Open'. In the window that appears you will see the text of the script. Under its menu 'Run' select 'Run Ctrl-R'.

This is the way to run all demo scripts. When in the book the text DEMOx.x appears (x.x stands for an index number) you should open the Praat script in the way described

above and run it. Sometimes a window appears wherein you can enter some parameter values mentioned in the text or continue after the script has interrupted itself: many scripts produce a window where the user can select options or give data. When you no longer need a certain script, you can close its window, but you can leave it alone as well: the program does not limit the number of scripts in memory. When the script has generated one or more windows you can close them in the normal way (by clicking the X in their upper right corner). However, you should not close the window named "Praat picture": if you want to get rid of the pictures drawn earlier you can erase them by selecting 'Erase all' under the 'Edit' menu of the 'Praat picture' window.

(The *sound editor* window in Praat is displayed slightly different from its default settings. The INIT script will automatically adjust the settings needed for our purposes. It has to be run only once.)

The scripts will produce certain objects in the Object window. When you quit Praat, the program will ask whether or not you want to save them. As the scripts produce the necessary objects anew when you run them there is no need to save them.

All this sounds more complicated than it is: during your actual Praat sessions you will find that it is all very straightforward.

Part A. Understanding basic signal analysis

1. The nature of sound

Most readers of this book probably know that sound in air consists of quickly varying local alterations of air pressure. All sound sources, i.e. all things or beings that make sound, do that by vibrating mechanical parts that move the air back and forth locally (or up and down or left and right or a combination of directions). These air movements cause local air pressure alterations which push and pull their neighboring areas and they push and pull the next areas in turn. In this way, waves of locally vibrating air travel in all directions, away from the sound source. There is a small delay before neighboring areas start to move: the propagation of sound in air takes time. At room temperature this speed is 340 m/s. Therefore, the sound wave propagates through air but the *air molecules themselves* remain where they are. Think about the waves that occur when you throw a stone in a pond: the waves travel along the surface of the water away from the position where the stone hit the water (the *source*) but the duck nearby moves only up and down. It remains at the same distance from the source. Of course, in air the waves travel in all three dimensions instead of two, as is the case with water waves.

Apart from air, sound can be propagated in all kinds of media: gases, liquids and solid matter. The propagation velocity depends on both the compressibility and the density of the medium. The higher the compressibility ("elasticity") the lower the propagation velocity will be. Also, the higher the density, the lower the propagation velocity will be. For helium gas, for example, the compressibility is quite the same as for air but the density is lower so that the propagation velocity is higher. For metal, the density is much higher than that of air but the compressibility is very, very much lower than that of air so that, as a result, the propagation velocity of sound in metal is about 5000 m/s which is 15 times the velocity in air. The propagation *velocity* is practically *not* dependent on wind; for sound propagation in air, the air is needed as a medium; the *type* of medium is not changed by wind, and the moving of the medium is much slower than the velocity of sound.

The spreading of the sound in all directions causes the sound **intensity**, which is expressed as power per unit of surface area, to decrease as the distance from the source increases: the **spreading loss**. If we think of a sphere with a certain radius r_1 around the sound source, all sound power is divided equally over the entire surface of the sphere so that everywhere on the sphere's surface the sound intensity has the same value.

The surface of the sphere with a radius r_1 is equal to $4\pi r_1^2$. If we alter the distance to r_2 the total surface is $4\pi r_2^2$. The intensity anywhere on the surface will have changed inversely proportional to the ratio of the *squares* of the radii. Consequently, if we define

one distance from the sound source as d_1 and one other distance as d_2 , and the intensity found at distance d_1 as I_1 and the intensity at distance d_2 as I_2 , we can write it as the following formula:

$$\frac{I_2}{I_1} = \frac{d_1^2}{d_2^2} \quad (1.1)$$

This formula is only valid for **point sources** which means that, theoretically, the sound source is supposed to be concentrated in one point. In practice, many sound sources can be considered as point sources if the size of the sound source is small compared to the measuring distances.

Another restriction on the validity of the formula is that the sound waves occur in the **free field** which means that there are no reflections by walls or furniture: all sound waves travel away from the sound source unhindered¹. In a room the sound loss depends highly on the sound reflection, absorption and dimensions of the room. (For example, in a typical office room without special acoustic adaptations the intensity at 2 meters from the source can be about 2 or 3 times as high as the intensity of the same source in the free field at the same distance, due to the reflections in the room. When the microphone is placed near a wall this factor must roughly be doubled, caused by ‘standing wave’ effects, which will be discussed later.)

Naturally, when the sound waves travel away from the source there is some loss of energy during the takeover of the movements by adjacent parts of air: the **atmospheric absorption**. This type of sound loss must be distinguished from the spreading loss as described above. This atmospheric absorption is dependent on the *viscosity* (toughness) of the molecular structure of the air. Generally, this energy loss increases when the sound contains higher tones (or frequencies, which is explained in section 3.). This type of energy loss is only important in cases of high distances: higher than about 1 km or so. In cases of shorter distances, the formula 1.1 will suffice.

The intensity being a measure of power per unit of area, this is not what our microphones register. Inside the microphone its diaphragm is activated by the **sound pressure** (p). The sound pressure is commonly called **SPL** (Sound Pressure Level) and must be seen as a force acting on a surface. It is expressed in pascal. 1 pascal is defined as 1 Newton per square meter. This SPL is what our microphones transform into proportional electrical voltages. Do not mix up the sound pressure with the intensity: *pressure* is force per unit of surface area, measured in Newtons per square meter, and *intensity* is power per unit of surface area, measured in Watts per square meter. The intensity is the acoustic power that the sound source emits through the unit of area; the pressure is the *effect* of this emission. This means that the intensity is proportional to

¹ The term *free field* refers to a theoretical space without sound reflections. In practice, even out-of-doors, there always will be some sound reflections.

the SPL *squared*. In the box called **OHM's LAW** this relation is explained in a bit more detail.

Now our formula 1.1 can be rewritten for SPL values instead of intensity values:

$$\frac{p_2^2}{p_1^2} = \frac{d_1^2}{d_2^2} \text{ or } \frac{p_2}{p_1} = \frac{d_1}{d_2} \quad (1.2)$$

If there is more than one sound source active at the same time, the air pressure at a particular position will be the result of all wave interactions. Therefore, at each particular point in time the voltage level of the microphone is proportional to the resulting air pressure level of all sound wave interactions at that point in time.

Our ear drums are put in motion in exactly the same way. Our brains are able to distinguish different sounds quite well from the resulting movements of the eardrums caused by different sound waves. Using both ears we can even hear the directions where the sounds come from, based on the little time delay differences of the sound wave

OHM'S LAW

The relation between pressure and intensity can be explained by using an analogy in electricity. You have *voltage* (the 230 volts of the mains outlet, for instance), and *current* (the amperes that flow through, say, a light bulb). If no lamp is connected with the outlet there is no current but there is a voltage. When a lamp is connected, the amount of current depends on the lamp type: a 100 Watt lamp causes a higher current than a 40 Watt lamp. The difference between the lamp types is determined by their *resistance*. Ohm's law defines the simple relation: the current is the voltage divided by the resistance ($I = V/R$). The resistance is a property of the lamp and has nothing to do with the outlet voltage. Consequently, $R = V/I$ and is a constant for a particular lamp which implies that, for example, doubling the voltage causes also doubling of the current.

Now, the amount of light that the lamp produces is proportional to the *power* (P , in Watts), which is the voltage multiplied with the current ($P = V \times I$). Substituting I by V/R gives: $P = V^2/R$. The power is proportional to the *square* of the voltage!

Accordingly, this can be applied to the subject of sound. Just as, for a certain lamp, the voltage causes a proportional current, the acoustical pressure level causes a proportional particle velocity, for a certain medium. 'Ohm's law' then is: $v = p/z$, where v is the particle velocity, p the pressure level (SPL) and z the acoustic **impedance** of the medium (in analogy with current, voltage and resistance, respectively). Here, the power is called *intensity* (in Watts/m²) and can be expressed as: $Int = vp = p^2/z$. The intensity is proportional to the square of the pressure!

For z the more general term *impedance* is used instead of resistance because of the possible frequency dependent phase shift between p and v . But that's another story!

paths between source and different ears, the *phase difference*. Likewise, the diaphragms of two microphones placed in similar positions as our ears, produce the complete information that our ears receive. Listening through earphones to recordings made in this manner gives the realistic sensation as if you are present in the room in which the sound was produced (*artificial head stereophony*). Even the awareness of sounds coming from behind or above is preserved. The reason why I am emphasizing this subject is that there are people who think that for more realistic spatial sound properties

you need more microphones, which obviously is a mistake. On the contrary, simply mixing more than two microphone signals into the stereo signal creates the opposite effect. Sometimes certain small signal delays are applied to the extra channels to get some artificial spatial sensation ('surround sound') but the real direction information is then disturbed.

In the practical part of the book we will go into the *acoustic* implications of the sound recording in some more detail, when we deal with the strategy to use microphones in rooms.

2. Decibel

The impression of sound intensity ('volume') is not linear: when the intensity of a sound, for example, doubles in magnitude, the impression is a certain increase of the volume. When the intensity doubles again, the *same increasing step* is perceived, and so on. Each doubling of the intensity we perceive as the same increase. Thus, the impression of the intensity behaves as if our hearing works logarithmically. This phenomenon should not be very surprising. Generally, our perception behaves not in an absolute but in a relative way. If you lift an object that weighs 1 kg, put it down and then lift an object of 1.5 kg, you will have some impression of the difference. If you do the same with objects of, say, 4 kg and 4.5 kg, the impression of the difference is much smaller. Only when the second set of objects weigh 4 kg and 6 kg respectively, the impression of the difference is the same compared with the first set. Apparently, for equal impressions of differences, the *percentage of increase* has to be constant. This general property of our perception is expressed by **Weber's law**:

$$\frac{\Delta I}{I} = C \quad (2.1)$$

where I is the physical value, ΔI the change of value which causes a certain impression and C a constant. Of course, the formula does not hold for very low (and very high) ΔI values. If the difference of the values is so small that it has reached the discrimination threshold or the *Just Notable Difference* (JND), the constant C becomes C_W : the *Weber's fraction*. Obviously, this JND too is a constant percentage of the absolute physical values.

For the representation of sound intensity, the logarithmic measuring unit **decibel** (dB) is used. In principle, it is nothing more than an exponential measuring unit for a *ratio* of two quantities. Its name implies that it is derived from the *bel* (named after Alexander Graham Bell) and, indeed, the decibel is 0.1 bel. One bel simply means a factor 10^1 . Two bels mean 10^2 and so on. Likewise: -1 bel = $10^{-1} = 0.1$. Then an increase by 1 decibel = $10^{0.1}$, which is about a factor 1.26 (or 26 % increase). The other way goes accordingly: for example, a factor 1000 means 10^3 ; its log is 3 which is 30 decibels. A factor 1/1000 means 10^{-3} which has a log of -3 which is -30 decibels. In general:

$$\#dB = 10 \log (INRatio) \quad (2.2)$$

$$INRatio = 10^{\#dB/10} \quad (2.3)$$

Now, this decibel is NOT the dB that is used for the expression of sound pressure! It is the norm to use the dB for the acoustical *intensity* which is the SPL *squared*. The acoustic intensity is the power per unit of area, and the SPL is the force per unit of area (see the box OHM'S LAW). As an example, when the SPL is raised by a factor 3, the intensity is raised by a factor 9. That implies that we have to *multiply* the number of dBs *by two* in formula 2.2 when dealing with SPLs because the dBs *refer to intensity*! In this way, the general formulas for SPL dBs become:

$$\#dB = 20\log (SPLratio) \quad (2.4)$$

$$SPLratio = 10^{\#dB/20} \quad (2.5)$$

For example, when the SPL ratio doubles, the number of dBs is increased by $20 \log (2)$ which is about 6 dB, as $\log (2) = 0.30103$.

So, when we want to express one *intensity* value with respect to a second *intensity* value, the formulas 2.2 and 2.3 should be used. When we want to express one sound *pressure* value with respect to a second sound *pressure* value, the formulas 2.4 and 2.5 should be used. This fact is emphasized because it seems that many people are unsure when to use the factor 10 and when the factor 20. It stems from the (historical) decision to interpret sound pressure in terms of the effect it has on the intensity. (The same applies to electrical voltages: the voltage ratio is also expressed in terms of the effect on the electrical power.)

Now what, for example, is meant by “the sound level is 80 dB”? As the dBs imply a *ratio* of SPL levels why are they used to define an *absolute* sound level? In other words, what level is 0 dB? The answer is that the average lowest hearing threshold is used as a reference. The weakest sound that is just noticeable for the average person is 20 μ Pa (micropascal: a millionth of a pascal) and defined as 0 dB. So, according to formula 2.5, 80 dB means a factor $10^{(80/20)} = 10000$. The SPL then is $10000 \times 0.00002 \text{ Pa} = 0.2 \text{ Pa}$. For indication of this reference, dB_{SPL} is sometimes used.

The strongest SPL that people can bear is about 120 dB which corresponds to 20 Pa. Levels like that cause pain and can damage the nerve cells of the inner ear. We may conclude from this that our ears can manage a pressure range ratio of $10^6 = 1$ million! As you may know, the atmospheric air pressure is about 10^5 Pa . So, the atmospheric air pressure is $10^5/20 = 5000$ times as high as the highest sound pressure we can bear! Luckily this atmospheric pressure is present at both sides of our eardrums...

Another measuring unit of sound levels is quite common: the **dB_A**. Our ear’s sensitivity is dependent on frequency. An attempt to take this into account is the application of a generalized frequency dependent function: the ‘**A-weighting**’ of sound. The dB_A values are adapted to this standardized frequency dependence. In Part B we will learn more about this method (and its limitations) to compensate for our ear’s properties.

A consequence of logarithmic measuring is that if there is no sound at all (zero *SPL*) this *cannot* be expressed on a log scale. In that case the *SPLratio* is $0/20 \mu\text{Pa}$ which is 0. And $\log (0)$ amounts to minus infinity. Therefore, please do not make the mistake that zero sound pressure corresponds with 0 dB! In practice, in our atmosphere there is always sound with *some* level, no matter how small, so that it always can be expressed in dB.

3. Waveform, frequency, spectrum

When you give a swing a single push from its position at rest, it moves forward and backward in a smooth, periodical way. It goes on like that, while the greatest deviations from its rest position slowly decrease until it finally comes to rest again. Were it not for the air resistance and the friction of the bending rope or the hinges, the movement would go on forever. In that case it would move exactly like a **sine wave** (provided that the deviations remain small). This type of movement forms the most basic periodic “vibration” there is. When a graph is plotted of the swing movement as a function of time, a graph like the example in fig. 3.1 will emerge, called the **waveform**. The maximum absolute value of this sine function is the **amplitude**.

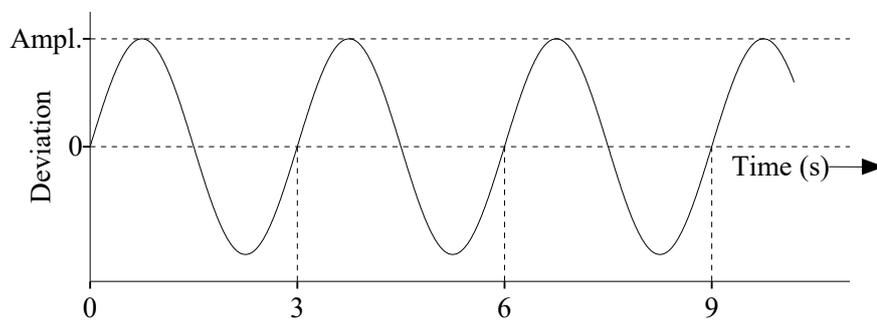


Fig. 3.1. Movement of a swing as a function of time.

If the air pressure is changing according to this sine waveform (although at a much faster pace), a so-called pure tone can be heard. Run DEMO3.1.script in the program Praat to play an example of a tone and displaying its graph. (Switch on your sound playing device or put on your headphones.)

As already mentioned, the signal which the microphone registers is the SPL of the sound. The vertical amplitude axis, therefore, refers to the SPL.

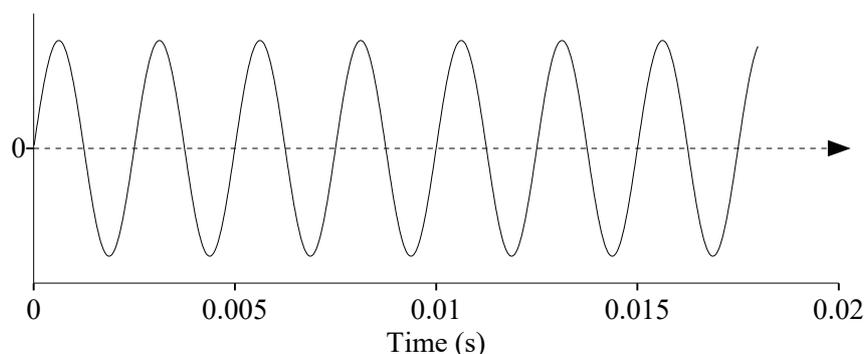


Fig. 3.2. Waveform of 18 ms of a sine wave.

In this example there are 400 complete cycles per second. Its **frequency** is 400 hertz. This can be read from the waveform as follows: the time of one complete “cycle” (= the time between two similar positions in the waveform, let's say two corresponding “zero crossings”), is 2.5 milliseconds (ms), so the number of cycles per second is $1/0.0025$ or $1000/2.5$ which is 400 hertz. (See also Fig. 3.2.)

In section 1 the propagation of sound in air was mentioned, being 340 m/s. So, one complete period of this 400 Hz tone spreads out in space over a length of $0.0025 \times 340 \text{ m} = 85 \text{ cm}$. Therefore, a tone of 400 Hz has a **wavelength** of 85 cm. Because our ears remain in the same position, we do not perceive these wavelengths, only the air pressure alterations. (In fact, if we would move towards the sound source, this would increase the speed with which the pressure alterations reach our ears and this increase of the propagation speed simulates a shorter wavelength and, therefore, a higher tone. The reverse, if we would move away from the sound source, the tone would be heard at a lower frequency. This is the well-known *doppler effect*.) The relation between wavelength, propagation speed and frequency is expressed by the formula:

$$\lambda = \frac{c}{f} \quad (3.1)$$

where λ is the wavelength, c the propagation speed and f the frequency.

It's a boring sound, isn't it? Let's make it a bit more interesting. Running DEMO3.2 will play a chord that is generated by the script but it could also have been produced by some electronic musical instrument. You see that its waveform looks quite complicated already (see also fig. 3.3). And yet it consists of only 3 pure tones added together. It is almost impossible to distinguish these individual sine waves from its waveform by looking at it. However, we can display this sound in a different way: we can create a graph that displays the amplitudes of the individual pure tones as a *function of frequency*.

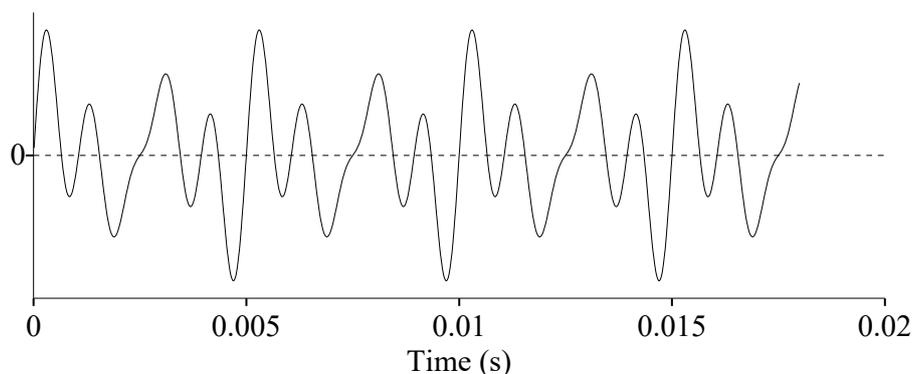


Fig. 3.3. Waveform of three added sine waves.

Fig. 3.4 shows this graph which is called **spectrum**, and it looks extremely simple. Basically, this spectrum consists of only three points, the horizontal position representing the frequency and the vertical position the amplitude. It is customary to draw vertical lines from each point to the horizontal axis: the *spectral lines*.

The vertical axis is scaled in dBs, which is used normally in spectra because the auditory impression of the intensity is logarithmic (see section 2). Another reason is that weak but audible components could be invisible if a linear scale was used. For example, the spectral line of a component that is 40 dB lower than the maximum would have a length of 1/100th of this maximum which is very near to zero on a linear scale. On a logarithmic scale with a customary range of, say, 80 dB this component reaches half of the maximum height in the graph!

The 0-dB reference in Praat spectra is a (one second) sound at the level of our hearing threshold (20 μ P SPL). (Although these values suggest a volume calibration, you must not rely on it because the volume of a sound played by the computer depends completely on the positions of the computer volume adjustment, the sound card amplification, the speaker or headphone properties, etc.) It is simply impossible to calibrate the level because of all these unknown variables. (See also the Praat manual about SPL calibration.) The *relative* values, however, can be read perfectly well from the graphs.

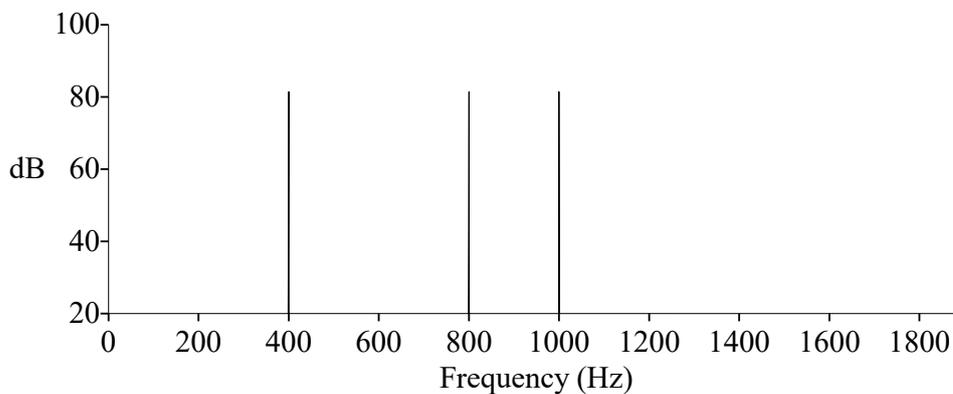


Fig. 3.4. Spectrum of three added sine waves.

Of course, we could construct this spectrum directly from our knowledge of the signal's properties: it consists of three sine waves with equal amplitudes. The Praat program, however, did not 'know' this. Nevertheless, it can accurately find the three sine components one way or another. Somehow it analyses the waveform and produces the sine wave components. The explanation of how this is made possible comprises the main part of the book.

For now, we can conclude that the *waveform* is the way the sound is represented in the **time domain**, and the *spectrum* is the way the sound is represented in the **frequency domain**, as is determined by the nature of the horizontal axes.

A commonly used name is **amplitude spectrum**, as it refers to the amplitudes of the frequency components. Strictly speaking, in the program Praat this is not correct, as this program, like many signal analysis programs, displays *density spectra* where the dB values depend on the time length. For the moment we would not bother about the difference as here it is only a matter of the way of scaling the vertical axis. In section 15 about noise the difference is explained. In Part B we will look at Praat's practical spectrum properties in some more detail.

It is also quite common to use the term **power spectrum**. Although we mean *SPLs* in the time and frequency representations of sound, the values in dB *refer to the power or intensity*, which is the amplitude *squared*, as you will remember from section 2. On a logarithmic scale, therefore, the values are doubled. Naturally, the *shapes* of the spectra remain unaltered when only the numbers along the axis are doubled.

4. Fourier series

When we look at the waveform of DEMO3.2 (fig. 3.3) we can see on the scaled time axis that a specific pattern is repeated each 5 ms during the entire signal. No matter where you define the start of this 5 ms section, the pattern is always repeated after 5 ms. In other words: this repetition period occurs 200 times per second. Therefore, the repetition has a **frequency** of 200 Hz (hertz), which is called the **fundamental** frequency. The relation between fundamental frequency and repeated period is represented by this simple formula:

$$F_0 = \frac{1}{T_0} \quad (4.1)$$

where F_0 and T_0 represent the fundamental frequency and the period, respectively. Frequency can be expressed in Hz, or *cycles per second*, which speaks for itself.

One repetition period of the waveform contains all information about this signal as all periods are the same. The complete data of all 3 sine waves are present in each single period. Now, a French mathematician and physicist, Jean Baptiste Joseph **Fourier** (1768-1830) proved that all *periodic* signals, i.e. signals that consist of repetitive parts of constant lengths and forms, can be represented by just combinations of sine waves of specific amplitudes and with frequencies that are *multiples only* of this fundamental frequency (F_0), including F_0 itself. The general form of a periodic signal is:

$$x(t) = x(t + T) \quad (4.2)$$

Here $x(t)$ stands for the varying sound amplitude as a function of time. We can conclude that any periodic signal can be represented by a set of sine waves of specific amplitudes and with frequencies of F_0 , $2F_0$, $3F_0$, $4F_0$, etc. In the example above, we have the following frequency components or **harmonics**:

1 st harmonic:	F_0 :	amplitude: 0	frequency: 200 Hz
2 nd harmonic:	$2F_0$:	amplitude: 1/3	frequency: 400 Hz
3 rd harmonic:	$3F_0$:	amplitude: 0	frequency: 600 Hz
4 th harmonic:	$4F_0$:	amplitude: 1/3	frequency: 800 Hz
5 th harmonic:	$5F_0$:	amplitude: 1/3	frequency: 1000 Hz
next multiples all have amplitude 0.			

Obviously, this is a simple example. Nevertheless, it teaches us that the amplitude of the frequency F_0 itself can be zero whereas the corresponding period (here 5 ms) is clearly present. The fundamental frequency only depends on the length of the repeated part which is determined by the *greatest common divisor* of the frequency components! Only then the frequency components will be multiples of a common fundamental frequency. In the part B section on frequency range it will also be shown that even for a sound with a clearly audible fundamental frequency this frequency component *need not necessarily be present in the spectrum!*

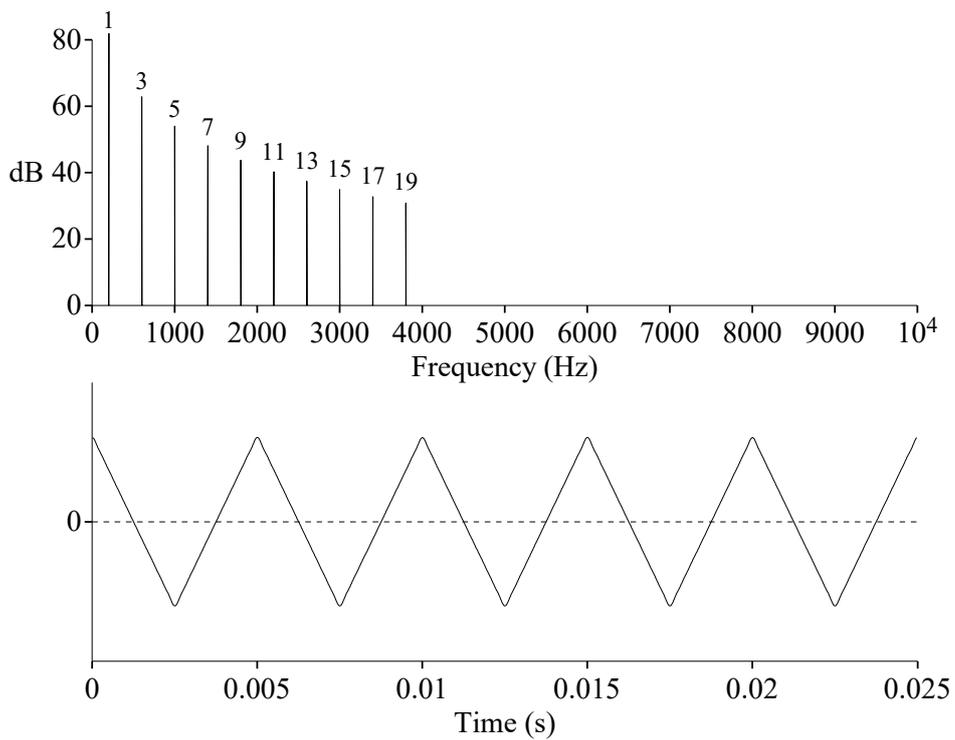


Fig. 4.1. Triangle waveform approximation with 10 odd harmonics.

Now, let's add some other sine wave components together. DEMO4.1 asks you to type the number of harmonics to generate. Try the default (already filled-in) value first.

A part of the resulting waveform is displayed and it looks like a *triangle* wave. (See also fig. 4.1.) The corresponding spectrum is displayed as well which shows the 10 harmonics. (You will see that there are only odd-numbered harmonics; the even harmonics are absent. This is always the case when the signal is symmetrical in the

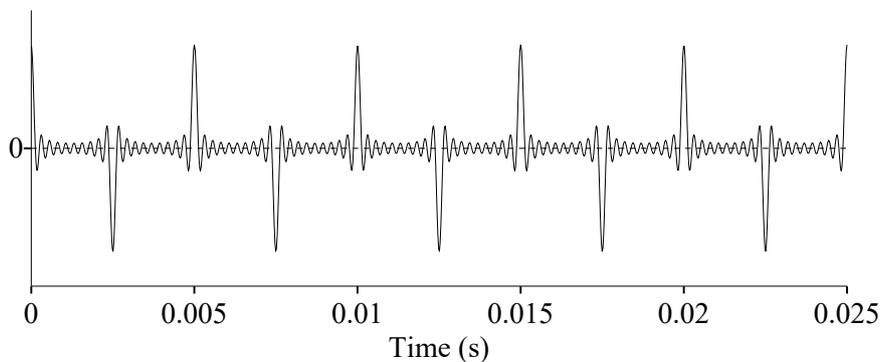


Fig. 4.2. Waveform from 10 odd harmonics with equal amplitude.

sense that the lower halves of the waveform are mirror images of the upper halves. Later, in section 6, we will learn more about symmetry properties.)

Of course, this triangular wave example is a theoretical one: in practice the waveforms are usually much smoother. This example is merely used for analytical purposes.

The spectrum shows that the demo script is designed in such a way that the amplitude of the harmonics decreases with increasing frequency. This *roll-off* is very ‘natural’: almost all waveforms of natural sounds can be synthesized with sets of harmonics that have some roll-off. As a contrast see fig. 4.2 for the waveform which arises when all 10 harmonics of the triangle example of fig. 4.1 have equal amplitude. The relatively shallow parts of the triangle have disappeared which indicates to the effect that the high frequency components are ‘emphasized’ at the cost of the low frequency components.

If you run the script again with 3 harmonics instead of 10, you will see that the triangle waveform is no longer perfect. You can try any number of harmonics. (Obviously, if you run the script with only one harmonic a perfect sine wave will emerge.)

Now run DEMO4.2 to approximate a *sawtooth* waveform (see also fig. 4.3.). You may experiment with different numbers of harmonics again. What you see is that the required number of harmonics to approximate the waveforms with some degree of

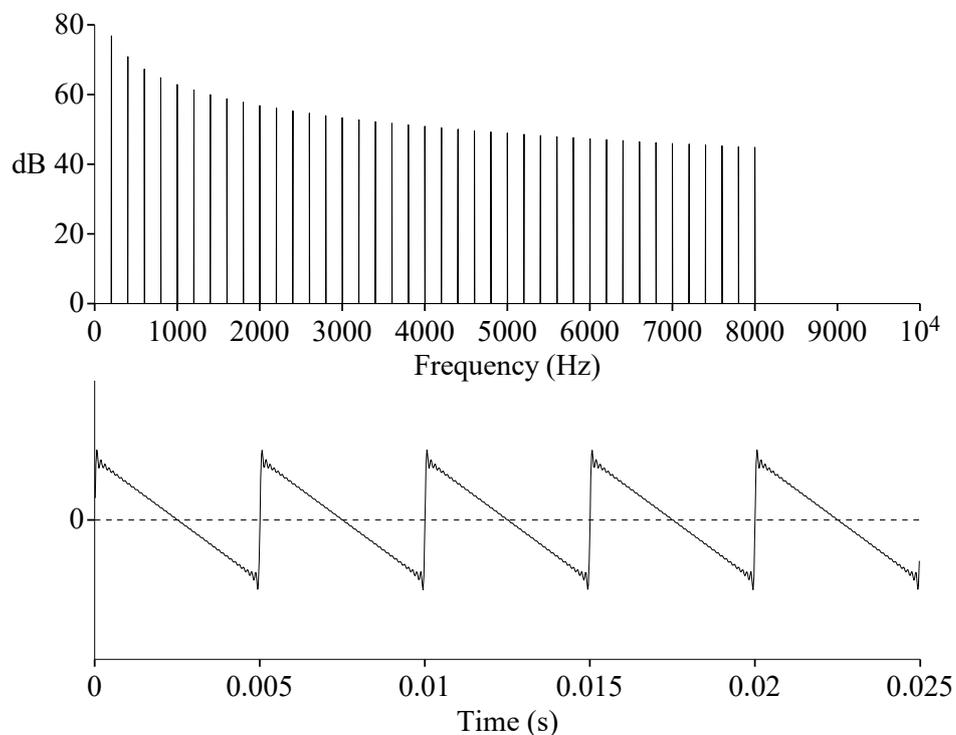


Fig. 4.3. Sawtooth waveform approximation with 40 harmonics.

accuracy is higher for the sawtooth wave than for the triangular wave (even when you include in your count the missing even harmonics of the triangular wave).

Theoretically, for signals with sudden changes or *discontinuities* in the waveform, we would need an infinite number of harmonics to acquire the perfect waveform. The sawtooth wave has a very big step in the waveform which needs strong high harmonics compared with those for the triangular wave. In practice, these waveforms do not exist in perfect form: their discontinuities always take some time which means that they are not pure mathematical discontinuities. Nevertheless, in practice it is always possible to approximate the practical waveform almost perfectly with a limited number of harmonics.

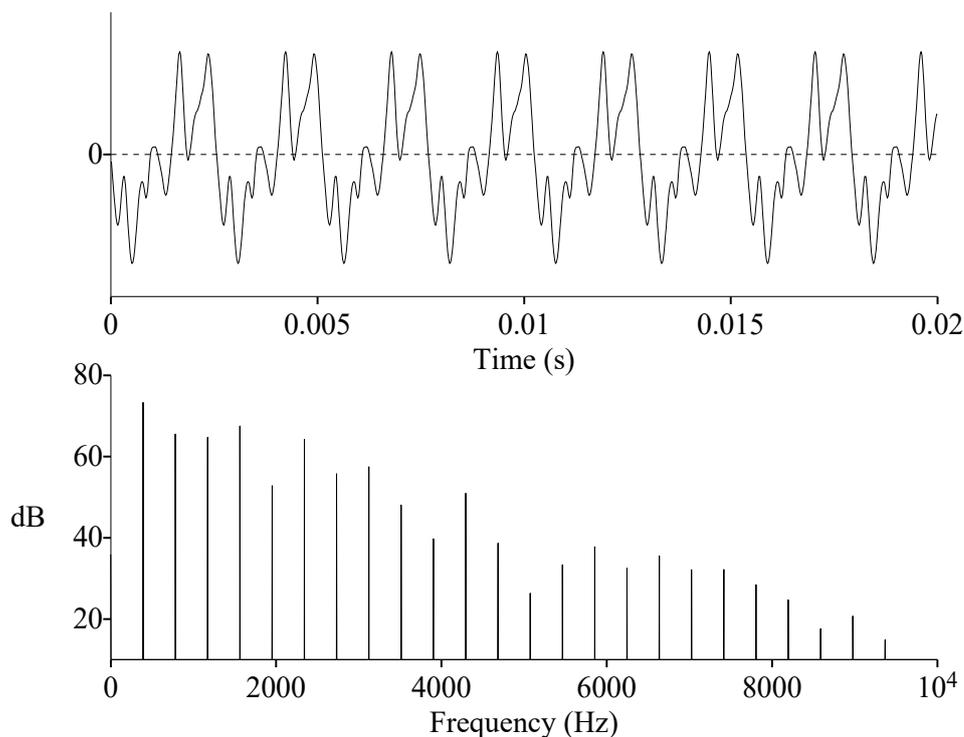


Fig. 4.4. Waveform of steady part of a violin sound, and its spectrum.

In these examples we have applied a Fourier *synthesis*. Let's do it the other way around. In DEMO4.3 we have a constant steady tone of a violin (see also fig. 4.4). The repeated part (T_0) is about 2.56 ms (milliseconds) so the F_0 is 390 Hz.

After you select *Continue* its spectrum is displayed (also in fig. 4.4). You will see that many multiples of F_0 exist and... in between there is nothing! Theoretically, all frequencies between the spectral lines are zero. (See the box **NO SILENCE ON dB SCALE** about the consequences of the logarithmic scale.)

NO SILENCE ON dB SCALE

It is not possible to properly display zero level on a logarithmic scale, like the dB scale. This is caused by the fact that the power of the logarithmic base number (say 10) must be minus infinity to get zero: $0 = 10^{-\infty}$. When generating signals with a computer program, the zero values (absent harmonics, for example), could jump to minus 300 dB or so, depending on the precision of the numbers used in your computer. Therefore, the dB range displayed in practice one mostly limits to 60 or 90 dB. In “Praat” you can adjust this range to any value. To see what is present at these extreme low levels you could run DEMO4.3 again and at “dB range” you can temporarily set it to 400 dB for example. Now you see something emerge like a “noise floor”. It is caused by the limited number precision in the computer. In practice this is not relevant: ratios of 100 dB or so are more than enough as this covers already an intensity ratio of 1 to 100000.

At the high frequency end of the spectrum the amplitudes of the harmonics decrease and gradually approach *negligible* values. The fact that the spectrum is empty between the spectral lines may seem strange and some people attribute this to the Fourier math peculiarities. However, it can easily be seen that this is a fundamental property: when the frequency of a sine component is a multiple of the fundamental frequency, its contribution to *each period* of the waveform is exactly the same. But, in the case of a sine

frequency that is, for example, $2.5 F_0$, its contribution to adjacent periods of the waveform is different: the phase of the sine component is not the same for each waveform period. (In this case it takes 2 periods of the waveform to reach the original phase; when the sine component is $2.3 F_0$, for example, it takes 10 waveform periods before the phase has returned again to its initial value.)

Therefore, the requirement for a sine component to have the *same contribution* to all periods of the waveform implies that it must have a frequency that is a multiple of the fundamental frequency.

ABOUT PERIODICITY

What if we add, say, a sine wave of 100 Hz and a sine wave of 103 Hz? Obviously, the fundamental frequency is 1 Hz (not 3 Hz!). So, strictly speaking, all multiples of 1 Hz are zero, except number 100 and number 103. Only if the ratio of the two frequencies is *exactly* 100/103 this is valid. For example, a ratio that is not exactly 100/103 but, say, 100.01/103 would produce a fundamental frequency as low as 0.01 Hz and does not make sense. In practice, we can fulfill this accuracy requirement for the Fourier synthesis very easily when we use a computer or other electronic device. It will be clear that, strictly spoken, in practice it is impossible to get this exact ratio when you add the sine waves of two *separate* sine sound sources together. Even when you use two separate computers that can generate sine waves with highly accurate frequencies, there will always exist some inaccuracy of the frequencies. Obviously, the concept of harmonics is only valid when the frequency components are *related* to each other, as is the case when they stem from one sound source only (or, one sound source *synchronizing* another one).

All these measured frequency components of a specific periodic signal together form its **Fourier analysis**. It's the opposite of Fourier synthesis. To be more precise, the