Fundamental Frequency Detection

Detecting the fundamental frequency can be done either in the time domain by estimating the period length of the fundamental or in the frequency domain by finding the frequency of the fundamental. Because we are working with discrete signals both domains have a maximum accuracy determined by the distance between two samples which results from the sampling frequency.

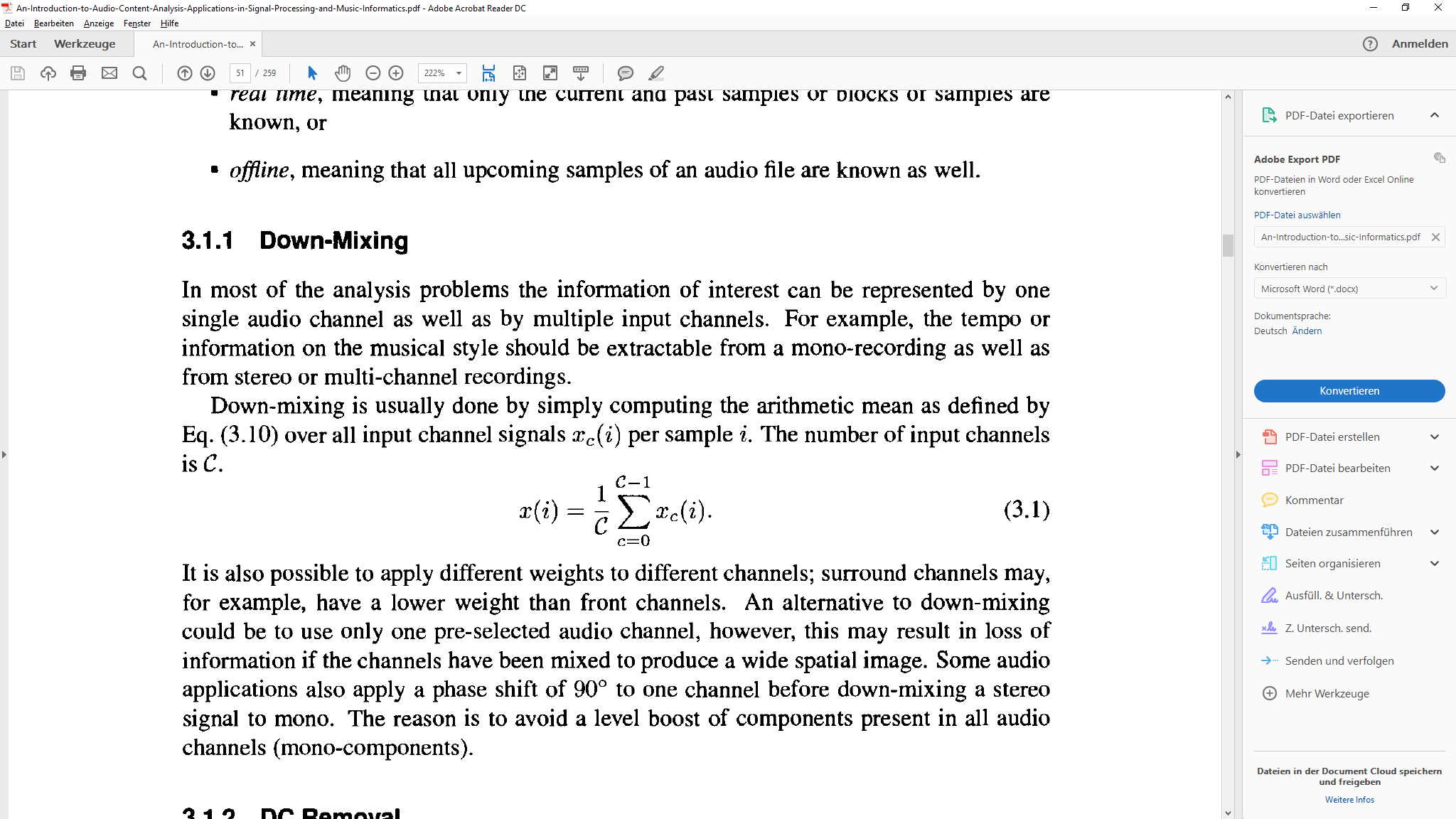
# Pre-Processing

In order to reduce the amount of audio data to be analyzed, the raw audio data is being pre-processed before computing the frequency from it. During this process unnecessary information is omitted to reduce computing overhead or to minimize the impact of unwanted information. Therefore, this chapter addresses the most important pre-processing techniques for computing the fundamental frequency of a drum [page 33 book].

## Down-Mixing

Audio data can be represented by one single audio channel (mono) as well as by multiple input channels (stereo, dolby, …). However, in many audio analysis problems including pitch estimation a single audio channel is sufficient. Therefore, the size of the data to be analyzed can be reduced drastically by using Down-Mixing.

Down-mixing is usually done by computing the arithmetic mean over all input channel signals per sample.



An alternative to this process could be the use of only one pre-selected audio channel by omitting the data from the other audio channels. While this seems like a simple workaround it is not recommended because it may result in loss of information if the channels have been mixed to produce a wide spatial image [page 33 book].

Although most microphones work in mono, down-mixing should still be considered in the implementation because the underlying software (Android, Java) may still generate a stereo or multi-channel signal from the recording.

## Normalization

Audio data can be recorded and represented at different levels. Recording hardware, microphone gain, analog-to-digital converters, the recording environment (room) and many other factors can have an impact on the audio level and therefore it’s amplitude. To compensate for this effect normalization is used. In this process the input signal is scaled in such a way that it has a pre-defined maximum. This helps us to extract features like fundamental frequency independently of the amplitude scaling of the input signal [page 34 book].

Normalizing is done by detecting the overall maximum of the signal and scaling the signal so that this maximum’s value is mapped to 1:

