Interpreting sound signal formulae

Matias Dahl, Ville Turunen

19.3.2004

In our examples, a sound signal is represented as a function

$$g : \mathbb{R} \to [-1, 1],$$

where the real variable is time (measured in seconds), and the value $g(t_0)$ refers to a non-scaled deviance from the surrounding air pressure at time t_0.

The standard air pressure is 1013 mBar (millibars). For human beings, the threshold of pain at audible frequencies corresponds to the air pressure variation of less than 1 mBar. Consequently, we may think that $g(t)$ corresponds to the air pressure

$$(1013 + g(t)) \text{ mBar}.$$

The threshold of hearing at audible frequencies is roughly 10^{-6} mBar (i.e. 0.000001 mBar, a millionth of 1 mBar), so that a vibration g with the amplitude lower than 10^{-6} is inaudible as far as we can tell.

The *frequency* of a sound signal within a time period is the momentary number of oscillations per second, measured in Hertz, abbreviated Hz. This is readily understood for sinusoidal vibrations like

$$g(t) = \sin(kt) :$$

here the frequency is $k/(2\pi)$ Hz, where $\pi \approx 3.14159$. For humans, audible frequencies range from 20 to 20000 Hz. Now let the signal g have the form

$$g(t) = \sin(w(t))$$

for some smooth-enough function $w : \mathbb{R} \to \mathbb{R}$. If w is increasing or decreasing within the time period $[t_0, t_0 + h]$, then there are about $|w(t_0 + h) - w(t_0)|/(2\pi)$ vibrations during this time interval of h seconds. Thus, within this time interval the average frequency is

$$\frac{|w(t_0 + h) - w(t_0)|}{2\pi h},$$

and if we let the variable h tend to 0, we get the frequency of the signal g at the time instant t_0:

$$\lim_{h\to 0} \frac{|w(t_0+h)-w(t_0)|}{2\pi h} = \frac{|w'(t_0)|}{2\pi};$$

the instantaneous frequency is the absolute value of the derivative of w divided by 2π.

Sound signals

The sampling rate is $1/\delta = 44100$ Hz and the unit for the variable t is seconds.

White noise (4a)

The characteristic for the white noise is that the samples of the signal are uncorrelated random numbers distributed uniformly over $[-1,1]$. Thus this noise does not vary in time, and all the possible frequencies are present with equal magnitude; the closest everyday instance is the sound of an untuned TV or radio channel. The noises encountered in nature are typically coloured, i.e. some frequency ranges are prevalent: for instance, a hum of a wind, sizzling of boiling water, etc.

$\mathbb{S}^1$ (2c) and $\mathbb{S}^{1000000}$ (1d)

Of the nine sounds, only two were instrumental-like. These are the sounds of $\mathbb{S}^1$ and $\mathbb{S}^{1000000}$ ringing. To distinguish these two from each other, a first observation is that one signal sounded like a plucked string and the other like a bell. We know that plucked string instruments have harmonic overtones. It also holds that bells (together with gongs) do not have harmonic overtones. Suppose Δ is the Laplace operator on $\mathbb{S}^n$ with respect of the metric induced from $\mathbb{R}^{n+1}$. We assume known that the eigenvalues of $-\Delta$ on $\mathbb{S}^1$ are $1, 2, 3, \ldots$. This implies that the sound of $\mathbb{S}^1$ has harmonic overtones. From this it follows that $\mathbb{S}^1$ has the sound of a plucked string. Therefore, by exclusion, $\mathbb{S}^{1000000}$ must sound like a bell.

Generally, the eigenvalues of $-\Delta$ on $\mathbb{S}^n$ are $\lambda_k = \sqrt{k(k+n-1)}$ for $k = 0, 1, 2, \ldots$, and the sound of $\mathbb{S}^n$ was defined as

$$g(t) = e^{-t} \sum_{k=1}^{\infty} 2^{-k} \sin\big(2\pi 220 \frac{\lambda_k}{\lambda_1} t\big).$$

Here, we have normalized the eigenvalues, such that the fundamental frequency of the signal is always 220 Hz. For $n = 1$, we obtain $\frac{\lambda_k}{\lambda_1} = 1, 2, 3, \ldots$. On the other

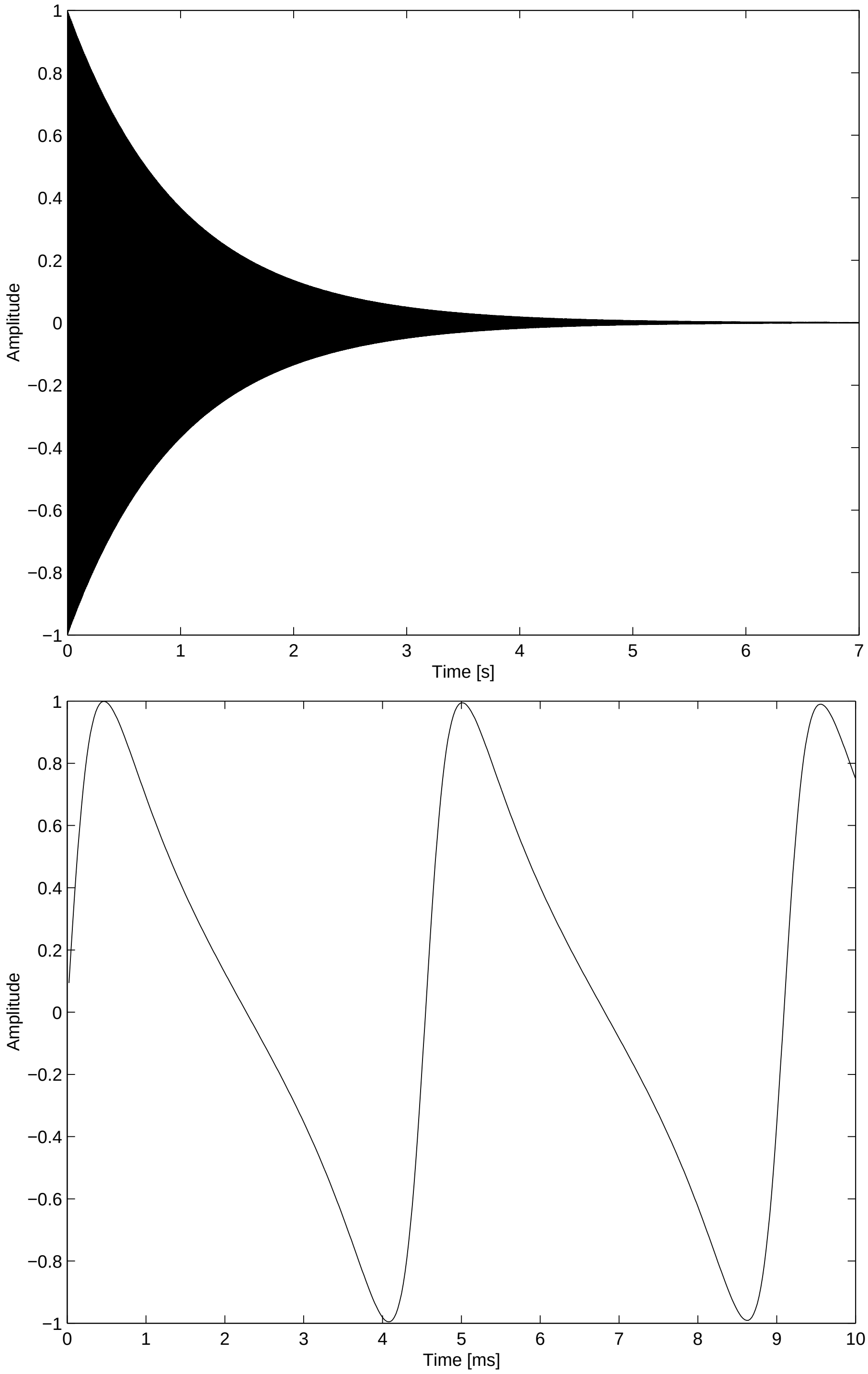

Figure 1: Exponential decay of the $\mathbb{S}^1$ signal(2c) and the first 10 ms of the signal.

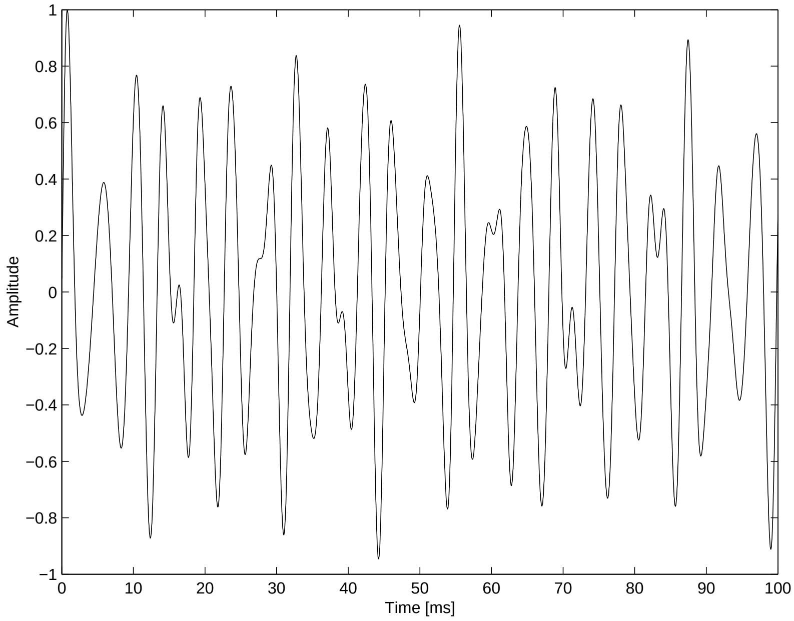

Figure 2: First $100\,\mathrm{ms}$ of the $\mathbb{S}^{1000000}$ vibrations (1d). In this signal, there is no period.

hand, when n is very large, we have

$$\begin{aligned} \frac{\lambda_k}{\lambda_1} &= \sqrt{\frac{k(k+n-1)}{n}} \\ &= \sqrt{\frac{k^2-k}{n}+k} \\ &\approx \sqrt{k}, \end{aligned}$$

so $\mathbb{S}^{1000000}$ does not have harmonic overtones.

Amplitude modulation (6e)

$$g(t) = \frac{\sin\big(40\sin(2\pi 10t)\big)}{1+\sin\sin t}.$$

This signal was the only signal with a clear variation in volume. This can be seen from the relatively slowly oscillating denumerator, which varies between $1-\sin 1 \approx 0.16$ and $1+\sin 1 \approx 1.84$. See Figure 3.

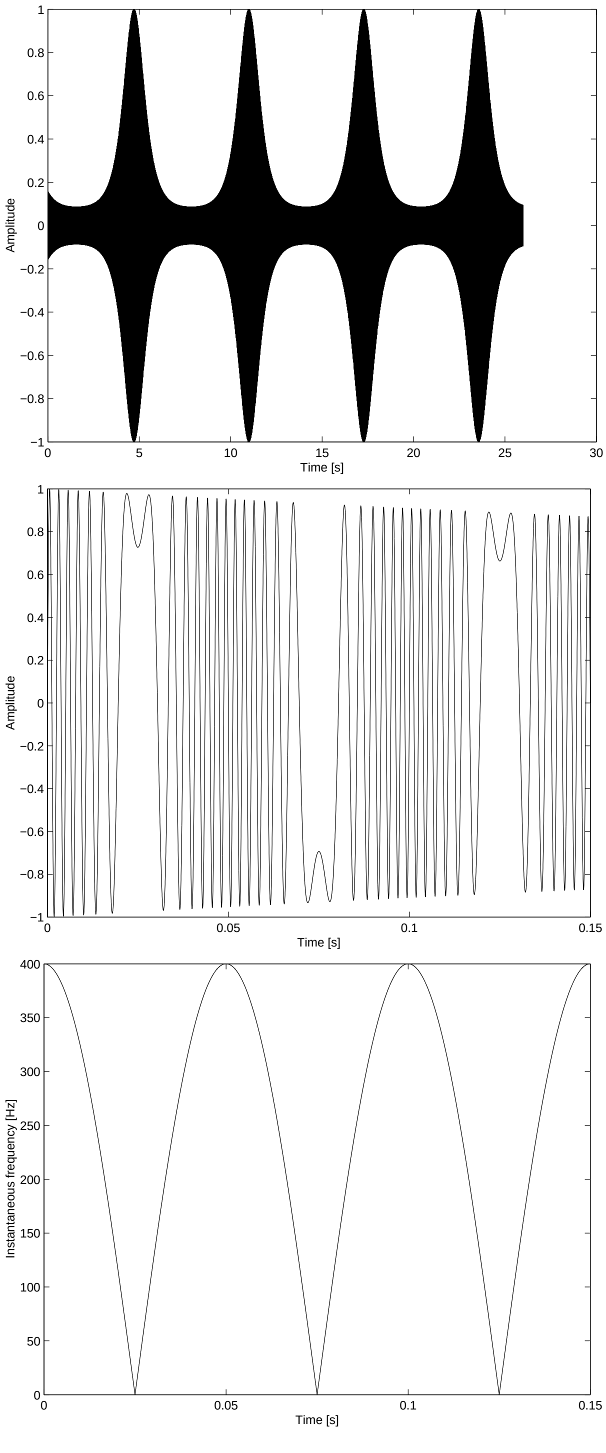

Figure 3: The first figure shows the amplitude modulation (6e), the second shows the signal for the first $100\,\mathrm{ms}$, and the last shows the instantaneous frequency over the first $0.15\,\mathrm{s}$.

Topologist's sinusoid (7b)

$$g(t) = \sin\big(2\pi\, 50\frac{1}{t}\big).$$

At $t = 0$, the signal has infinite instantaneous frequency, which smoothly drops below the audible range in less than 2 seconds. This results into a familiar sci-fi sound effect. See Figure 4.

Karplus–Strong (5i)

$$g(t) = \begin{cases} \sin\frac{9\pi t}{2\Delta}, & \text{when } t < \Delta, \\ \frac{1}{2}\Big(g(t-\Delta) + g(t+\delta-\Delta)\cdot(S\circ g)(t-\Delta)\Big), & \text{when } t \geq \Delta. \end{cases}$$

Here $\Delta = 10\,\text{ms}$ and $S(t) = \operatorname{sign}\sin\sinh(100\,t)$.

This function is an implementation of the Karplus-Strong algorithm, which is a standard algorithm for producing plucked string and drum sounds. The algorithm starts with an initial signal (here $\sin\frac{9\pi t}{2\Delta}$ for $t \in (0,\Delta)$), which is inductively mutated using the lower branch in the definition. With good accuracy, S randomly takes values $+1$ and -1. Thus in the copy process, the signal is randomly passed through a lowpass filter (when $S = 1$), or a highpass filter (when $S = -1$). In effect, the sound becomes noisier over time while its amplitude decays. See Figure 5.

Recurring pulse with increasing frequency (3f)

$$g(t) = \sin\left(2\pi 100 \sin^{50}\left(2\sinh\frac{t}{4}\right)\right).$$

Since $\sin^{50}$ is almost zero for most of the time, this signal contains pulses, and since $\sinh$ grows exponentially, these pulses become more frequent over time. Also, from the expression for the instantaneous frequency, one can see that the frequency of the signal grows exponentially. This effect can be seen in Figure 6.

Frequency modulation (8g)

$$g(t) = \sin\left(150t^2 + 70\sin(2\pi\, 40\sqrt{t})\right).$$

For this sound, the instantaneous frequency is

$$\left|\frac{1}{2\pi}\left(300t + 2800\pi\frac{\cos(2\pi 40\sqrt{t})}{\sqrt{t}}\right)\right|.$$

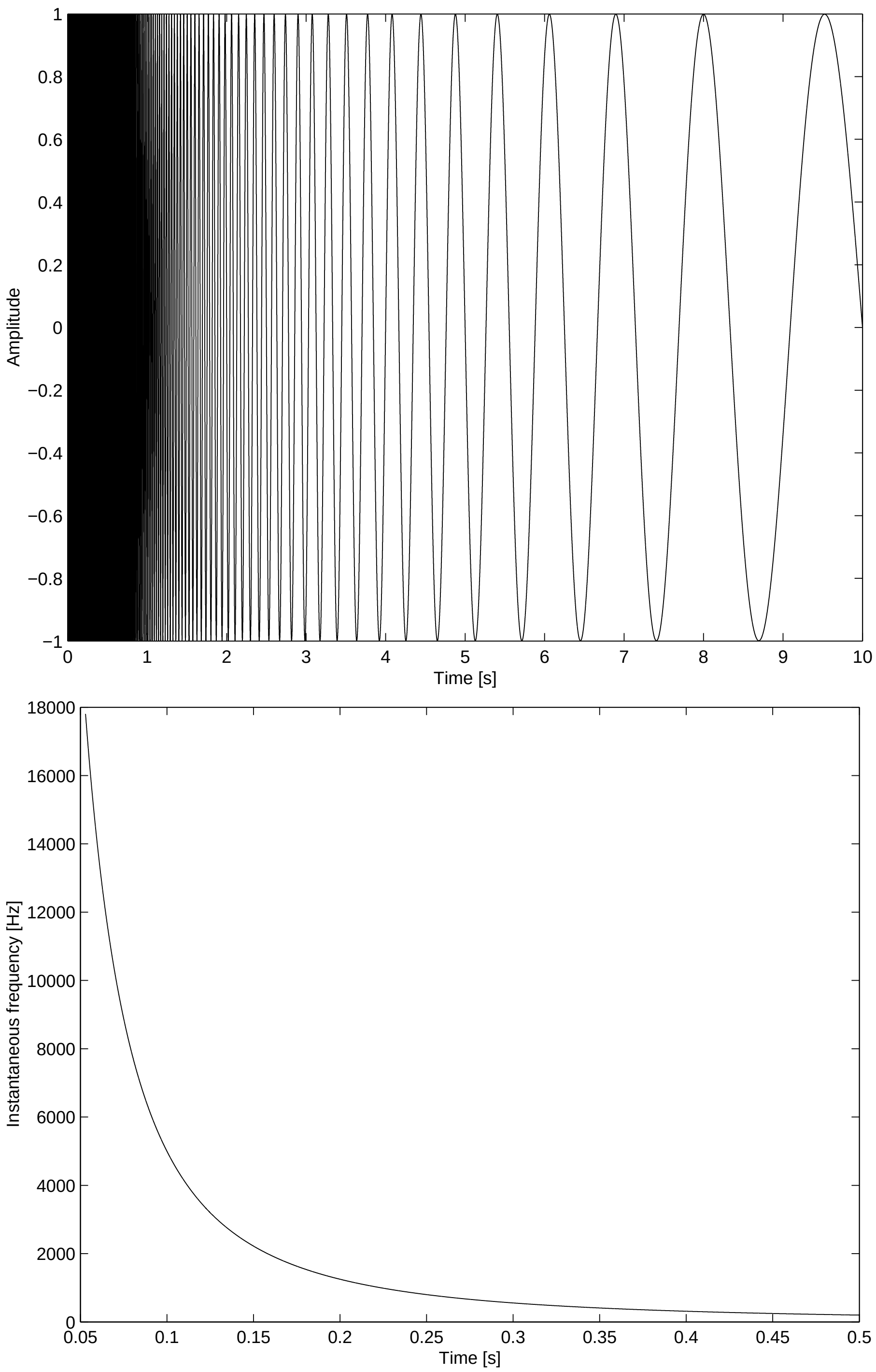

Figure 4: Topologist's sinusoid (7b): entire signal and a plot of the instantaneous frequency.

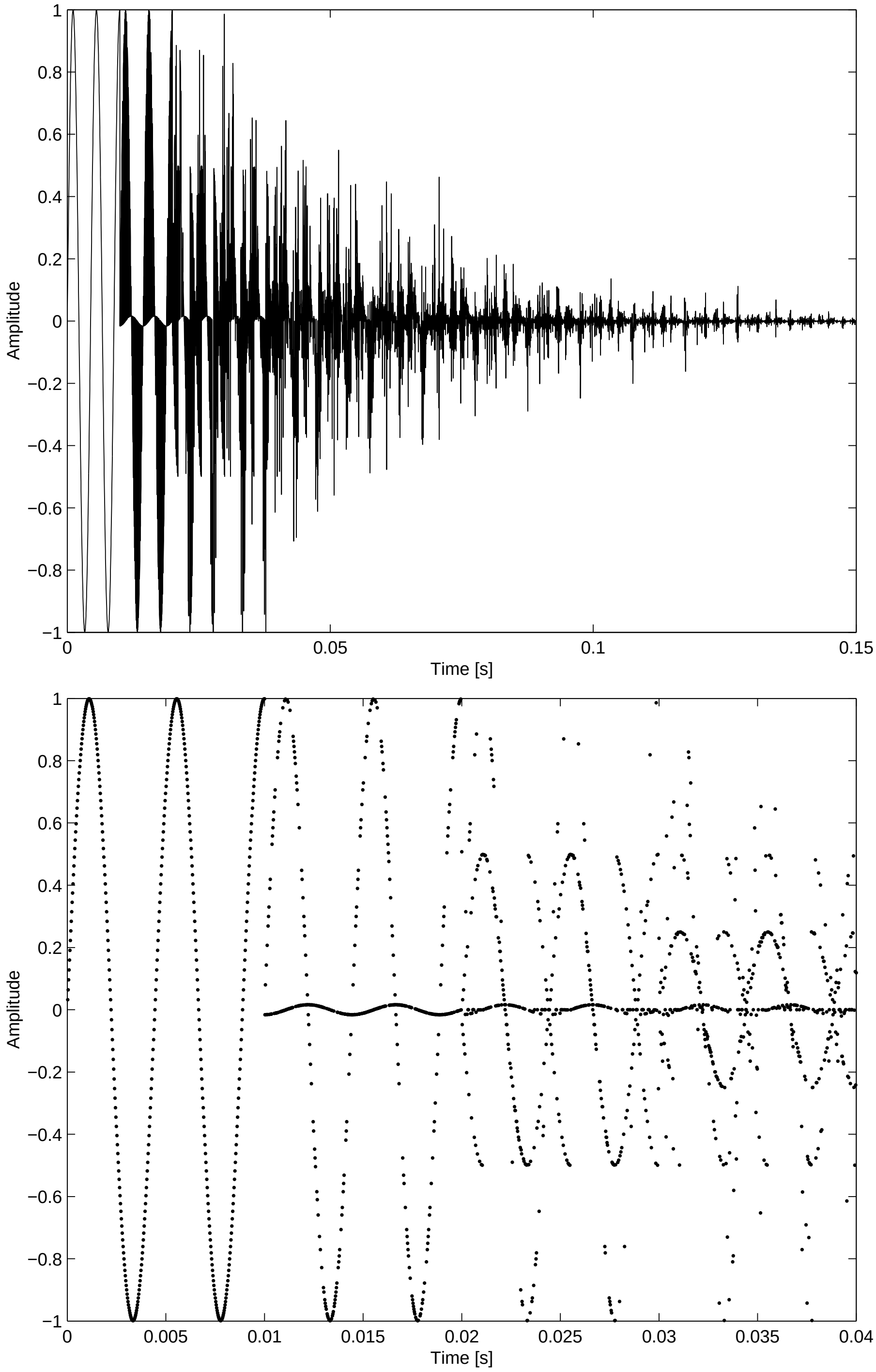

Figure 5: Karplus-Strong (drum sound) (5i).

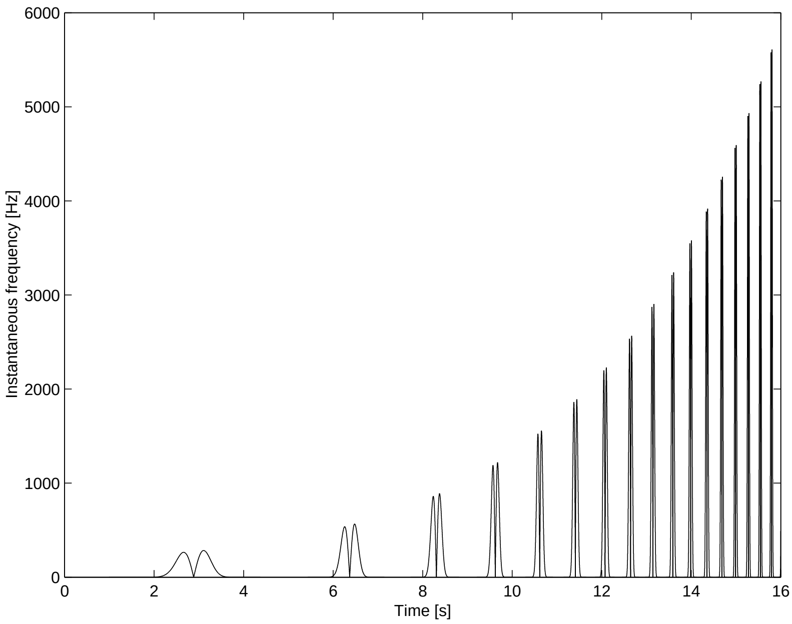

Figure 6: Instantaneous frequency for (3f).

Due to the $\frac{1}{\sqrt{t}}$-term, the signal starts with infinite frequency which rapidly drops. For large t, the $300t$-term dominates. The overall effect is that the instantaneous frequency is oscillating and the longterm average frequency is slowly rising. This can be seen in Figure 7.

Squeak sound (9h)

$$g(t) = \begin{cases} 0, & \text{when } t < 0, \\ \sin\left(4\pi t + g(t-3\delta) - \frac{t}{20} g(t-2\delta) + g(t-\delta)\right), & \text{when } t \geq 0. \end{cases}$$

Using induction, one may show that g is continuous. Moreover, g is infinitely smooth except for $t = 0, \pm\delta, \pm 2\delta, \ldots$.

Here, we have sampled the function at $t = 0, \delta, 2\delta, \ldots$, that is, at the points where g is continuous, but not necessarily smooth. This sound contains periodic clicks occurring twice a second. It can therefore be recognized from the $4\pi t$ term.

To analyze this sound, let us define highpass and lowpass filters H_δ and L_δ, respec-

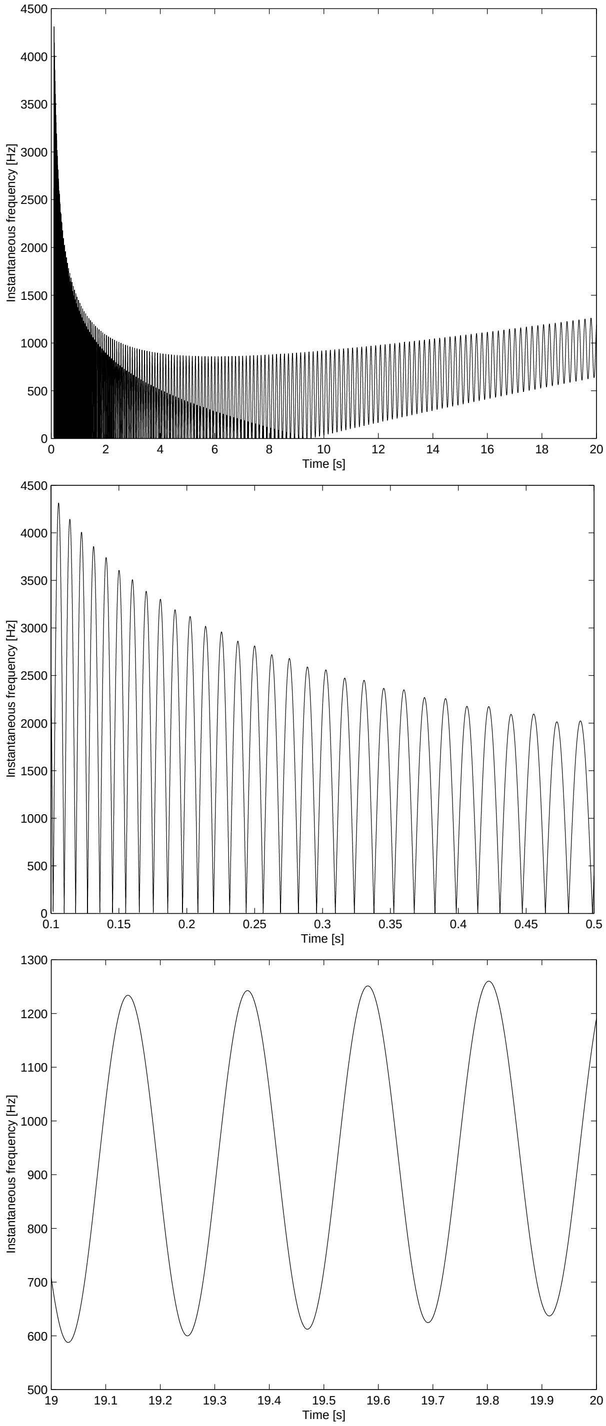

Figure 7: Instantaneous frequency for (8g).

tively. For a signal h, these filters are defined as

$$\begin{aligned} H_\delta h(t) &= \frac{h(t) - h(t-\delta)}{2}, \\ L_\delta h(t) &= \frac{h(t) + h(t-\delta)}{2}. \end{aligned}$$

Suppose $h(t) = \sin(2\pi\omega t)$. Then using the summation formulas for $\sin$ and $\cos$ one can show that

$$\begin{aligned} H_\delta h(t) &= \sin(\pi\omega\delta)\ \sin\left(2\pi\omega(t - \frac{\delta}{2}) + \frac{\pi}{2}\right) \\ &\approx \sin(\pi\omega\delta)\ h(t), \\ L_\delta h(t) &= \cos(\pi\omega\delta)\ \sin\left(2\pi\omega(t - \frac{\delta}{2})\right) \\ &\approx \cos(\pi\omega\delta)\ h(t), \end{aligned}$$

where, in the approximations, we have assumed that δ is much smaller than ω. In both cases, we see that the original signal is scaled by a function depending of the angular frequency ω of h. Such a function is called a *frequency response function*. Figure 8 shows how these filters indeed single out the high and low frequency components of h. Note that the range of audible frequencies is roughly 20 to 20000 Hz.

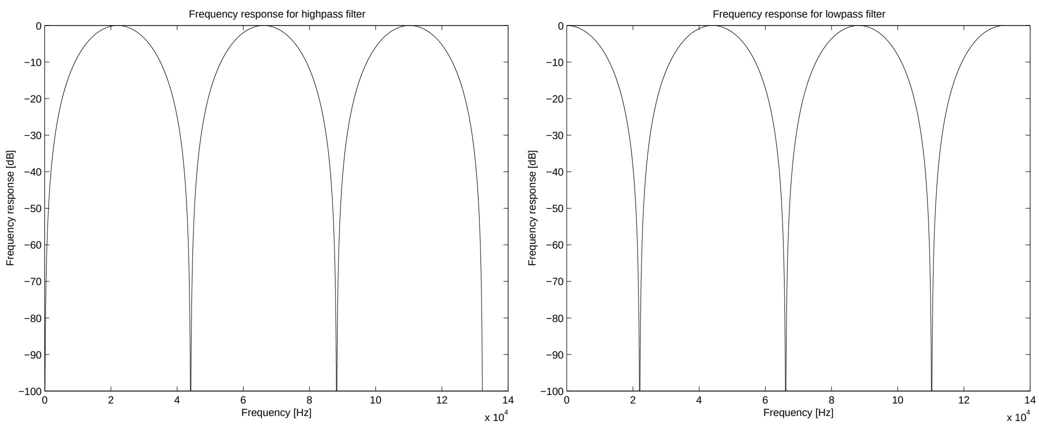

Figure 8: Lowpass and highpass frequency responses.

Figure 9 shows the sound filtered through the lowpass and highpass filters, respectively. Figure 10 shows the sound passed through both of these filters. (The filters commute, $L_\delta H_\delta = H_\delta L_\delta = \frac{1}{2} H_{2\delta}$, so the order is irrelevant.) In Figure 10, the clicks and squeaks are clearly seen.

Figure 11 shows zooms of some squeaks in the original signal.

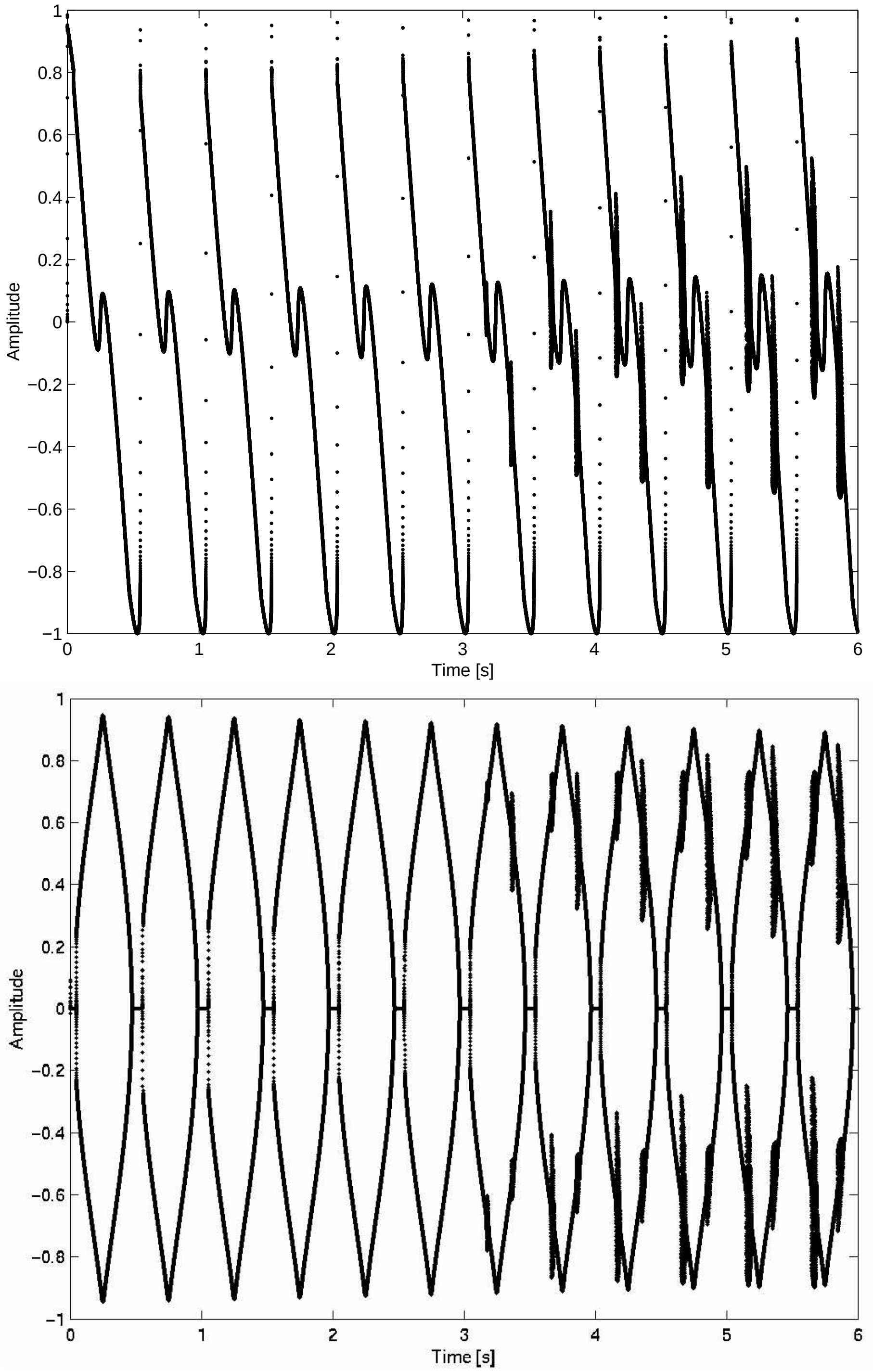

Figure 9: Lowpass and highpass filtered squeak sound (9h).

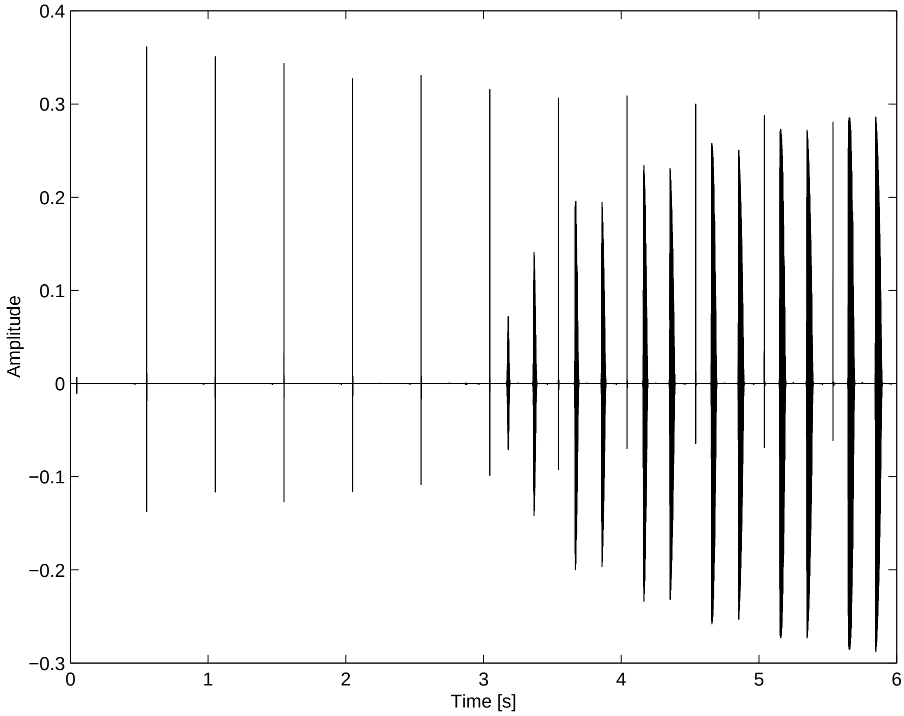

Figure 10: Squeak sound (9h), both lowpass and highpass filtered.

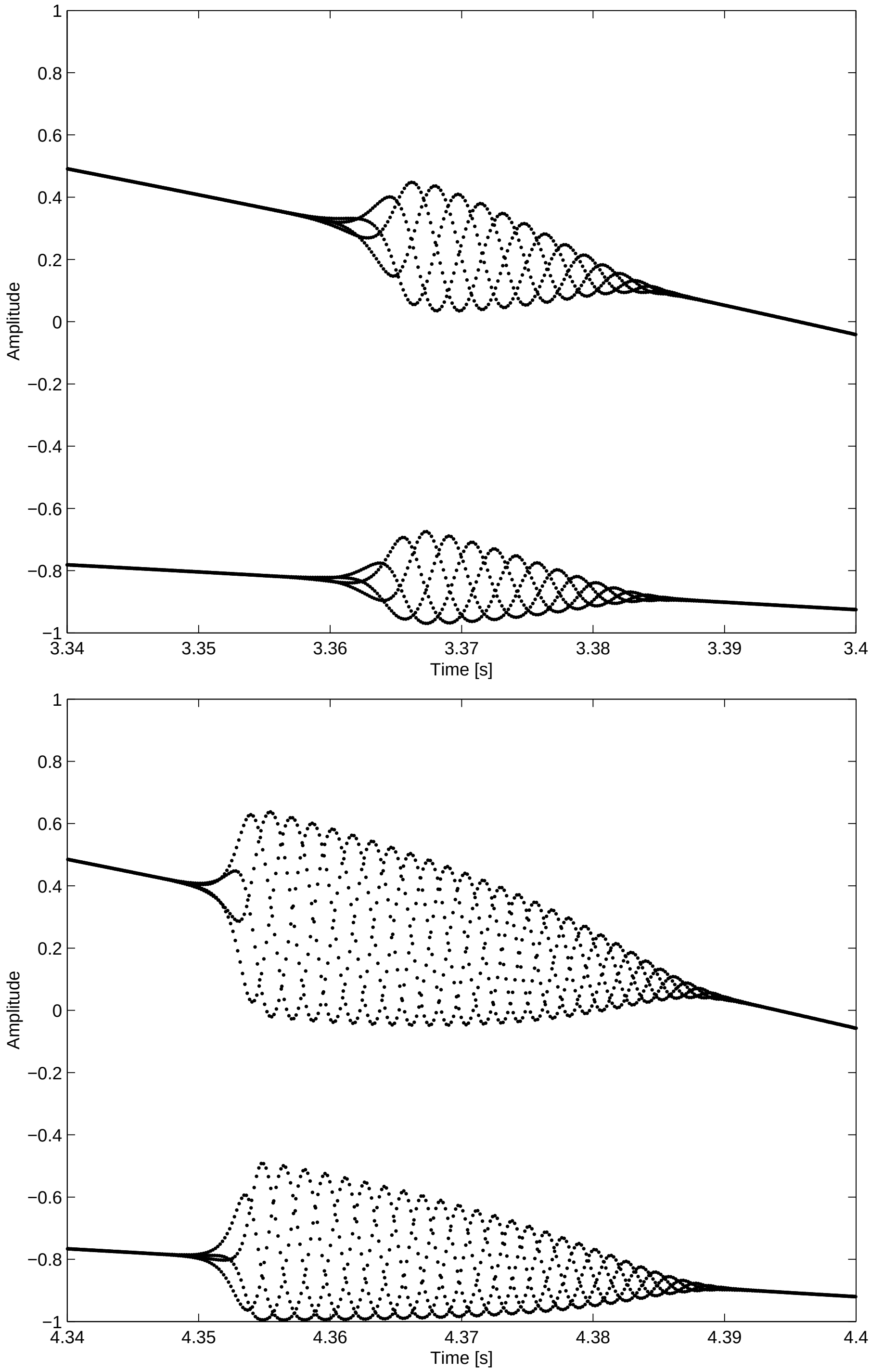

Figure 11: Zooms of the eighth and tenth period of the squeak sound (9h).