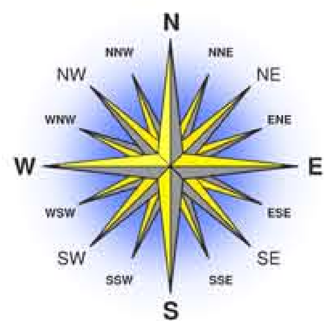
TUNING FREQUENCY ESTIMATION USING CIRCULAR STATISTICS

Karin Dressler
Fraunhofer IDMT
dresslkn@idmt.fraunhofer.de

Sebastian Streich
UPF - Music Technology Group
sstreich@beat.yamaha.co.jp



Circular statistics (also called directional statistics) is a subdiscipline of statistics that is for example applied when directions or periodic time measurements have to be evaluated. Such data are often best handled not as scalars, but as (unit) vectors in the complex plane. Typical examples for such data include the departure directions of birds from points of release, the orientation of fracture planes, the directional movement of animals in response to stimuli, directions of wind and water currents, and biorhythms (e.g. day time, days of a week, months of a year).

For the estimation of the tuning frequency, only the cent deviation of the estimated pitch frequency from the discrete 100 cent semi tone grid is of importance. Each cent value is treated as a unit vector $\hat{u}$ whose angle ϕ is the appropriate fraction of a full circle:

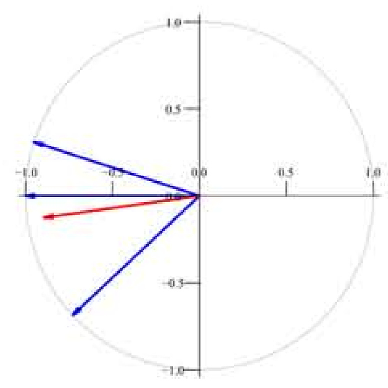
$$\hat{u} = 1 \cdot e^{j\phi} \qquad \text{with} \quad \phi = \frac{2\pi}{100} \cdot c.$$

The sum of all cent vectors is computed and then divided by the number of values N to get a mean $\bar{z}$ of the circular quantities:

$$\mathrm{Re}(\bar{z}) = \frac{\sum_i \cos(\phi_i)}{N}$$
$$\mathrm{Im}(\bar{z}) = \frac{\sum_i \sin(\phi_i)}{N}.$$

The magnitude $|\bar{z}|$ is a confidence measure for the tuning frequency estimate. The phase angle $\bar{\phi}$ of $\bar{z}$ is converted back into cents to obtain the deviation from the standard tuning frequency of 440 Hz in cents:

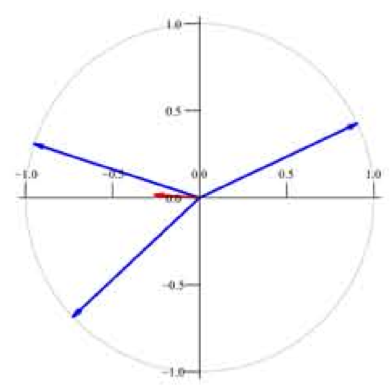
$$\Delta c = \frac{100}{2\pi}\arg(\bar{z}).$$



Example 1:
cent deviation: 45 cent, ± 50 cent, -38 cent

$$\hat{u_1} = \cos(\tfrac{45 \cdot 2\pi}{100}) + i\sin(\tfrac{45 \cdot 2\pi}{100}) = -0.951 + 0.309i$$
$$\hat{u_2} = -1.0$$
$$\hat{u_3} = -0.729 - 0.685i$$

$$\bar{z} = (\hat{u_1} + \hat{u_2} + \hat{u_3})/3 = -0.893 - 0.125i$$
$$|\bar{z}| = \sqrt{(-0.893)^2 + (-0.125)^2} = 0.902$$
$$\Delta c = \tfrac{100\,\mathrm{cent}}{2\pi} \cdot \arg(-0.893 - 0.125i) = -47.8\,\mathrm{cent}$$



Example 2:
cent deviation: 7 cent, 45 cent, -38 cent

$$\hat{u_1} = \cos(\tfrac{7 \cdot 2\pi}{100}) + i\sin(\tfrac{7 \cdot 2\pi}{100}) = 0.905 + 0.426i$$
$$\hat{u_2} = -0.951 + 0.309i$$
$$\hat{u_3} = -0.729 - 0.685i$$

$$\bar{z} = (\hat{u_1} + \hat{u_2} + \hat{u_3})/3 = -0.258 + 0.017i$$
$$|\bar{z}| = \sqrt{-0.258^2 + 0.017^2} = 0.259$$
$$\Delta c = \tfrac{100\,\mathrm{cent}}{2\pi} \cdot \arg(-0.258 + 0.017i) = 49\,\mathrm{cent}$$

Overall Estimate - Weighted Sum

In order to determine the tuning frequency of the entire audio piece the sum of weighted cent vectors is computed and then divided by the sum of all weights r_i:

$$\mathrm{Re}(\bar{z}) = \frac{\sum_i r_i \cos(\phi_i)}{\sum_i r_i}$$
$$\mathrm{Im}(\bar{z}) = \frac{\sum_i r_i \sin(\phi_i)}{\sum_i r_i}$$

Three different approaches were examined:

Spectral peaks: noisy peaks and overtones impair the result, low confidence values

Melody pitch: frequency variations of the human singing voice (eg. glissandi and vibrato) impair the result, low confidence values

Stable melody pitch: highest confidence values, but still no satisfactory results for human singing voice

	1.) spectral peaks			2.) melody pitch			3.) stable melody pitch		
file	$\lvert\bar{z}\rvert$	Δc	f_{ref}	$\lvert\bar{z}\rvert$	Δc	f_{ref}	$\lvert\bar{z}\rvert$	Δc	f_{ref}
daisy1	0.54	1.7	440.4	0.77	0.8	440.2	0.83	0.7	440.2
daisy4	0.70	0.2	440.1	0.87	0.5	440.1	0.95	0.5	440.1
jazz1	0.43	-2.1	439.5	0.67	-2.2	439.5	0.71	-1.1	439.7
jazz4	0.37	-0.5	439.9	0.83	0.3	440.1	0.84	1.4	440.4
midi3	0.82	-4.4	438.9	0.96	-4.7	438.8	0.99	-4.7	438.8
midi4	0.39	-2.9	439.3	0.93	-0.7	439.8	0.94	-0.8	439.8
opera4	0.28	8.5	442.3	0.12	12.8	443.3	0.30	-6.9	438.3
opera5	0.13	15.4	443.9	0.11	10.9	442.8	0.48	25.8	446.6
pop2	0.31	-4.4	438.9	0.35	-10.5	437.3	0.45	-13.1	436.7
pop3	0.21	-2.4	439.4	0.41	3.2	440.8	0.50	3.2	440.8

Table 1. Overall tuning frequency estimation of the ISMIR2004 data set

Exponential Moving Average

The exponential moving average (EMA) is a technique that applies weighting factors which decrease exponentially in time. If the reference frequency might change over time, the current reference frequency may be estimated using the following recursive formula:

$$z'_i = (1-\alpha) \cdot z'_{i-1} + \alpha \cdot r_i\, e^{j\phi_i} \quad \text{with} \quad z'_0 = 0,$$

where α determines the exponential decay.
The sum z'_i is normalized, so that the magnitude of $\overline{z'_i}$ lies in the interval $[0, 1]$:

$$\overline{z'_i} = \frac{z'_i}{r'_i} \qquad \text{with} \qquad r'_i = (1-\alpha) \cdot r'_{i-1} + \alpha \cdot r_i$$

The estimation of the adaptive reference frequency can be used to study the evolution of the tuning over time, which may be useful to monitor the tuning of instruments or singers/choirs singing a capella. The EMA approach also proved useful for the quantization of a sung melody into a note representation.

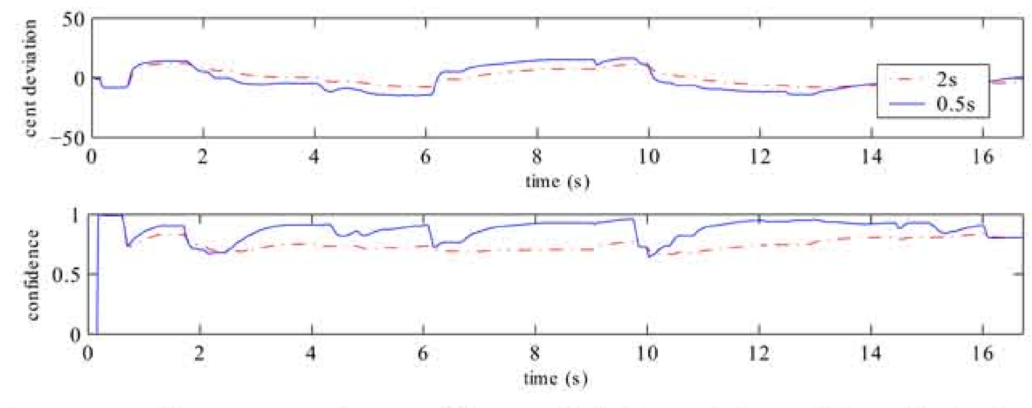


Figure 1. Short-term and long-term estimates of the cent deviation and the confidence for the jazz1 recording

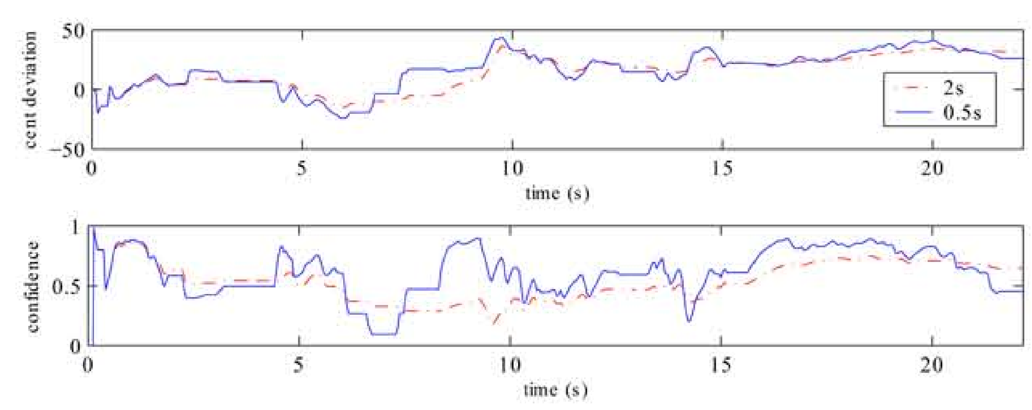


Figure 2. Short-term and long-term estimates of the cent deviation and the confidence for the opera5 recording

Fraunhofer Institut Digitale Medientechnologie
IDMT
UPF Music Technology Group