How to Detect the Fundamental Frequency: Approach Motivated by Soft Computing and Computational Complexity

1st Eric Freudenthal
Department of Computer Science
University of Texas at El Paso
El Paso, Texas, USA
efreudenthal@utep.edu

2nd Olga Kosheleva
Department of Teacher Education
University of Texas at El Paso
El Paso, Texas, USA
olgak@utep.edu

3rd Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, Texas, USA
vladik@utep.edu

***Abstract*—Psychologists have shown that most information about the mood and attitude of a speaker is carried by the lowest (fundamental) frequency. Because of this frequency's importance, even when the corresponding Fourier component is weak, the human brain reconstruct this frequency based on higher harmonics. The problems is that many people lack this ability. To help them better understand moods and attitudes in social interaction, it is therefore desirable to come up with devices and algorithms that would reconstruct the fundamental frequency. In this paper, we show that ideas from soft computing and computational complexity can be used for this purpose.**

***Index Terms*—soft computing, fundamental frequency, social interactions, computational complexity**

I. Outline

According to psychologists, the fundamental frequency components of human speech carry the bulk of information about the mood and attitude of the speaker. Because of the importance of the fundamental frequency signal, even when the actual Fourier component corresponding to this frequency is absent, the brain automatically reconstructs this frequency.

The problem is that many people lack this automatic ability and thus, miss important speech-related social cues. In this paper, we use the ideas from soft computing and computational complexity to reconstruct the fundamental frequency component and thus, to help these people better understand the social aspects.

The structure of this paper is as follow. In Section 2, we explain what is fundamental frequency, why it is important, and why reconstructing this frequency is important. In Section 3, we explain ideas about this reconstruction motivated by soft computing and computational complexity. In Section 4, we explain that these ideas indeed help.

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes), and by the AT&T Fellowship in Information Technology.

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

II. What Is Fundamental Frequency and Why It Is Important

Speech uses different frequencies. Some people have lower voice, some higher ones. In physical terms, this difference means that the sounds are – locally – periodic with different frequencies:

- lower frequencies for lower voices,
- higher frequencies for higher voices.

Fundamental frequency and harmonics. Every periodic signal $x(t)$ with frequency f_0 can be represented as a sum of sinusoids corresponding to different frequencies $f = f_0$, $f = 2f_0, \ldots$:

$$\begin{aligned} x(t) = A_1 \cdot \sin(f_0 \cdot t + \varphi_1) + \\ A_2 \cdot \sin(2f_0 \cdot t + \varphi_2) + \\ \ldots + \\ A_n \cdot \sin(n \cdot f_0 \cdot t + \varphi_n) + \\ \ldots \end{aligned} \quad (1)$$

for some coefficients $A_i \geq 0$ and φ_i; this is known as the *Fourier series*; see, e.g., [2]–[4], [6], [13], [16], [19].

The lower frequency f_0 – corresponding to the given period – is known as the *fundamental frequency*. Frequencies $f = n \cdot f_0$ for $n > 1$ are known as *harmonics*.

The overall energy of the signal is equal to the sum of energies of all the harmonics, and the energy of each harmonic is proportional to the square A_n^2 of its amplitude. Since the overall energy of the signal is finite, in general, this means that the amplitudes of the harmonics decrease with n – otherwise, if the amplitudes did not decrease, the overall energy would be infinite.

This is not just a mathematical representation of audio signals: in the first approximation, this is how we perceive the sounds: there are, in effect, biological sensors in our ears which are attuned to different frequencies f; see, e.g., [5], [18].

Which frequencies are most useful for conveying information. We can convey information by slightly changing the amplitude of each harmonic during each cycle. Higher-frequency harmonics have more cycles and thus, can convey more information.

Because of this, most information is carried by such high-frequency harmonics; see, e.g., [17].

Which frequencies are most useful for conveying mood and attitude. In addition to information, we also want to convey our mood, our attitude to the person. This attitude is not changing that fast, so the best way to convey the attitude is to use harmonics which are changing the slowest – i.e., which have the lowest frequency.

And indeed, most such socially important information is conveyed on the lowest frequencies, especially on the lowest – fundamental – frequency; see, e.g., [11].

In most people, the brain automatically reconstructs the fundamental frequency. Because of the social importance of understanding the speaker's mood and attitude, our brain is actively looking for this information. So, even when the actual fundamental frequency is suppressed, our brain reconstructs it based on the other harmonics.

For example, while many people have very low voices, with fundamental frequencies so low that these frequencies are cut off by the usual phone systems, we clearly hear their basso voices over the phone – because in most people, the brain automatically reconstructs the fundamental frequency.

Related problem. The problem is that for many people, this reconstruction does not work well. As a result, these people do not get the mood and other social information conveyed on the fundamental frequency – a big disadvantage in social interactions.

It is therefore desirable to come up with some signal transformation that would help these people detect the fundamental frequency.

III. Ideas Motivated by Soft Computing and Computational Complexity

Where can we get ideas: a general comment. We would like to use computers to help people reconstruct the fundamental frequency. In other words, we need to find an appropriate connection between people's brains and computers.

In general, if we want to find a connection between two topics A and B, a natural idea is:

- either to start with topic A and try to move towards topic B,
- or to start with topic B and try to move towards topic A.

Let us try both these approaches.

Let us start with the brain: ideas motivated by soft computing. Let us first analyze what we can get if we start with the brain side, i.e., with the what-we-want side. From this viewpoint, what we want is to somehow enhance the signal $x(t)$ so that the enhanced signal will provide the listener with a better understanding of the speaker's mood and attitude.

Since we started from the brain side, not from the computer side, naturally the above goal is imprecise, it is not formulated in terms of computer-understandable precise terms, it is formulated by using imprecise ("fuzzy") words from natural language. Our ultimate objective is to design a computer-based gadget that would pursue this goal. Thus, we need to translate this goal into computer-understandable form.

The need for such a translation has been well known since the 1960s, when it was realized that a significant part of expert knowledge is formulated in vague natural-language terms. To translate this knowledge into precise terms, Lotfi Zadeh came up with an idea of fuzzy logic, in which we describe each imprecise property like "small" by assigning:

- to each possible value of the corresponding quantity,
- a degree to which this value satisfies the property of interest – e.g., is small;

see, e.g., [1], [10], [12], [14], [15], [20]. Fuzzy techniques are a particular case of what is known as *soft computing* – intelligent techniques motivated by not-fully-precise ideas like neural networks or evolutionary computations.

If our case, the desired property is "being informative";

- when the signal is 0, we gain no information;
- the stronger the signal, the more information it conveys.

The simplest way to capture this idea is to have a degree d which is proportional to the signal's strength $x(t)$:

$$d = c \cdot x(t)$$

for some constant c.

We want to enhance this information. In terms of fuzzy techniques, what does enhancing mean? In general, it means going:

- from tall to very tall,
- from strong to very strong, etc.

In fuzzy approach, the easiest way to describe "very" is to square the degree. For example:

- if the degree to which some value is small is 0.7,
- then the degree to which this value is very small is estimated as $0.7^2 = 0.49$.

So, whenever the original degree was proportional to $x(t)$ – corresponding to the signal $x(t)$ – the enhanced degree is proportional to $(x(t))^2$ – which corresponds to the signal $(x(t))^2$. Thus, from the viewpoint of soft computing, it makes sense to consider the square $(x(t))^2$ of the original signal.

An important point is what to do with the sign, since the signal $x(t)$ can be both positive and negative while the degree is always non-negative. So, we have two choices.

The first choice is to take the formula $(x(t))^2$ literally, and thus to consider only non-negative values for the enhanced signal.

The second choice is to preserve the sign of the original signal, i.e.:

- to take $(x(t))^2$ when $x(t) \geq 0$ and
- to take $-(x(t))^2$ when $x(t) \leq 0$.

In the following text, we will consider both these options.

What if this does not work. What if we need an additional enhancement? From the viewpoint of fuzzy techniques, if adding one "very" does not help, a natural idea is to use something like "very very" – i.e., to apply the squaring twice or even more time.

As a result, we get the idea of using $(x(t))^n$ for some $n > 2$.

What if we start with a computer: ideas motivated by computational complexity. Because the desired signal enhancement is very important for many people in all their communications, we want the signal enhancement to be easily computable – not just available when the user has access to a fast computer.

In a computer, the fastest – hardware-supported – operations are the basic arithmetic operations:

- addition and subtraction are the fastest,
- multiplication is next fastest, and
- division is the slowest.

So, a natural idea is to use as few arithmetic operations as possible – just one if possible – and to select the fastest arithmetic operations.

Using only addition or subtraction does not help – these are linear operations, and a linear transformation does not change the frequency. So, we need a non-linear transformation, and the fastest non-linear operation is multiplication. At each moment of time t, all we have is the signal value $x(t)$, so the only thing we can do is multiply this signal value by itself – thus getting $(x(t))^2$.

Similarly to the soft computing-motivated cases, we can deal with the sign in the same two different ways, it does not change the computation time.

And similarly to the soft computing case, if a single multiplication does not help, we can multiply again and again – thus getting $(x(t))^n$ for some $n > 2$.

Discussion. The fact that two different approaches lead to the exact same formulas makes us confident that these formulas will work. So let us analyze what happens when we try them.

IV. These Ideas Indeed Help

What if we use squaring. Let us start with squaring the signal. We consider the case when the component corresponding to fundamental frequency – i.e., proportional to A_1 – is, in effect, absent, i.e., when for all practical purposes, we have $A_1 = 0$. In this case, the general formula (1) takes the form

$$x(t) = A_2 \cdot \sin(2f_0 \cdot t + \varphi_2) + A_3 \cdot \sin(3f_0 \cdot t + \varphi_3) + \ldots \quad (2)$$

Since, as we have mentioned, in general, the effect of higher harmonics – proportional to A_n – decreases with n, in the first approximation, it makes sense to only consider the largest terms in the expansion (2) – i.e., the terms corresponding to the smallest possible n.

The simplest possible approximation is to consider only the second harmonic, i.e., to take

$$x(t) \approx A_2 \cdot \sin(2f_0 \cdot t + \varphi_2).$$

However, in this case, all we have is a periodic process with frequency $2f_0$, all the knowledge about the original fundamental frequency f_0 is lost. Thus, to be able to reconstruct the fundamental frequency f_0, we need to consider at least one more terms, i.e., take

$$x(t) \approx A_2 \cdot \sin(2f_0 \cdot t + \varphi_2) + A_3 \cdot \sin(3f_0 \cdot t + \varphi_3), \quad (3)$$

in which, as we have mentioned, $A_2 > A_3$.

Since $A_2 > A_3$, the main term in $(x(t))^2$ is proportional to A_2^2. However, we cannot restrict ourselves to this term only, we need to take A_3-term into account as well. Let us therefore take into account the next-largest terms proportional to $A_2 \cdot A_3$. Thus, we get

$$(x(t))^2 = A_2^2 \cdot \sin^2(2f_0 \cdot t + \varphi_2) + 2A_2 \cdot A_3 \cdot \sin(2f_0 \cdot t + \varphi_2) \cdot \sin(3f_0 \cdot t + \varphi_3). \quad (4)$$

From trigonometry, we know that $\cos(x) = \sin(x + \pi/2)$,

$$\sin^2(a) = \frac{1}{2} \cdot (1 - \cos(2a)) = \frac{1}{2} \cdot (\sin(2a + \pi/2)),$$

and

$$\sin(a) \cdot \sin(b) = \frac{1}{2} \cdot (\cos(a-b) - \cos(a+b)) = \frac{1}{2} \cdot (\sin(a - b + \pi/2) - \sin(a + b + \pi/2)).$$

Thus, we have

$$(x(t))^2 = \frac{A^2}{2} - \frac{A^2}{2} \cdot \sin(2f_0 \cdot t + \varphi_2 + \pi/2) + A_2 \cdot A_3 \cdot \sin(f_0 \cdot t + (\varphi_3 - \varphi_2 + \pi/2)) - A_2 \cdot A_3 \cdot \sin(5f_0 \cdot t + (\varphi_2 + \varphi_3 + \pi/2)),$$

hence, in this case, $(x(t))^2$ indeed contains a component with fundamental frequency.

What if the fundamental frequency component in $(x(t))^2$ is not large enough. In this situation, both soft computing and computational complexity approaches recommended using $(x(t))^n$ for larger n. Will this work? Let us analyze this case.

To perform this analysis, let us take into account that:

- while we usually consider time to be a continuous variable,
- in reality, any sensor – be it biological sensors in our ears or electronic sensors – only measures values at several moments of time $t_1 < t_2 < \ldots < t_N$ during each cycle.

In these terms, a signal is represented by an N-dimensional tuple

$$(x(t_1), \ldots, x(t_N)).$$

At one of these moments of time, the absolute value of the signal takes the largest value during the cycle. In the N-dimensional space of all such tuples, tuples for which $|x(t_i)| = |x(t_j)|$ for some $i \neq j$ form a set of smaller dimension – thus, a set of measure 0. Thus, for almost all tuples, such an equality is not possible – and hence, the largest absolute value of the signal is only attained at one single point on this cycle.

Let us denote this point – where maximum is attained – by t_m. This means that for all other moments of time t, we have $|x(t)| < |x(t_m)|$, i.e.,

$$\frac{|x(t)|}{|x(t_m)|} < 1.$$

As n increases, the ratio

$$\frac{|x(t)|^n}{|x(t_m)|^n} = \left(\frac{|x(t)|}{|x(t_m)|}\right)^n$$

tends to 0. Thus, as n increases, the signal $(x(t))^n$ tends to the function that is equal to 0 everywhere except for one single t_m on each cycle. Such a function is known as a *delta-function* $\delta(t - t_m)$.

It is known that the Fourier transform of the periodic delta-function contains all the components – including the component corresponding to fundamental frequency – with equal amplitude. Thus, even if the square $(x(t))^2$ does not have a visible fundamental frequency component, eventually, for some n, the signal $(x(t))^n$ will have this component of size comparable to all other components – and thus, visible.

Limitations of squaring. Squaring the signal works well when the original signal $x(t)$ does not have a visible fundamental frequency components. However, if $x(t)$ has a strong component corresponding to fundamental frequency – e.g., if this component prevails and we have

$$x(t) \approx A_1 \cdot \sin(f_0 \cdot t + \varphi_1),$$

then squaring leads to

$$(x(t))^2 = \frac{A_1^2}{2} \cdot (1 - \sin(2f_0 \cdot t + 2\varphi_1 + \pi/2)),$$

i.e., to signal that no longer has any fundamental frequency component – indeed, it is periodic with double frequency $2f_0$.

So, in principle, we should consider *both*:

- the original signal $x(t)$ and
- its square $(x(t))^2$.

It is desirable to have a single signal instead.

One such possibility is to have sign-modified signal

$$(x(t))^2 \cdot \text{sign}(x(t)),$$

where:

- $\text{sign}(x) = 1$ for $x > 0$ and
- $\text{sign}(x) = -1$ for $x < 0$.

Indeed, in this case:

- even if the original signal is a pure sinusoid

$$x(t) \approx A_1 \cdot \sin(f_0 \cdot t + \varphi_1)$$

with frequency f_0,
- its transformation has a non-zero Fourier coefficient A'_1 – since this coefficient is proportional to

$$\int \sin(f_0 \cdot t) \cdot \sin^2(f_0 \cdot t) \cdot \text{sign}(\sin(f_0 \cdot t))\, dt =$$

$$\int |\sin(f_0 \cdot t)| \cdot \sin^2(f_0 \cdot t)\, dt > 0.$$

Preliminary confirmation. Our preliminary simulation experiments confirmed these theoretical findings; see, e.g., [7]–[9].

Acknowledgments

The authors are thankful to all the participants of the 27th Joint NMSU/UTEP Workshop on Mathematics, Computer Science, and Computational Sciences, Las Cruces, New Mexico, April 2, 2022 for valuable suggestions.

References

[1] R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.
[2] A. Brown, *Guide to the Fourier Transform and the Fast Fourier Transform*, Cambridge Paperbacks, Cambridge, UK, 2019.
[3] C. S. Burrus, M. Frigo, and G. S. Johnson, *Fast Fourier Transforms*, Samurai Media Limited, Thames Ditton, UK, 2018.
[4] Th. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, Massachusetts, 2022.
[5] S. Da Costa, W. van der Zwaag, L. M. Miller, S. Clarke, and M. Saenz, "Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex", *Journal of Neuroscience*, 2013, Vol. 33, No. 5, pp. 1858–1863.
[6] J. L. deLyra, *Fourier Transforms: Mathematical Methods for Physics and Engineering – Volume 2*, 2019. by Jorge L. deLyra
[7] Frudensing, LLC, *Freudensong: Unleashing the Human Voice*, https://www.freudensong.com/
[8] E. Freudenthal, R. Alvarez-Lopez, V. Kreinovich, B. Usevitch, and D. Roundy, "Work-in-progress: vocal intonation regeneration through heterodyne mixing of overtone series", *Abstracts of the 27th Joint NMSU/UTEP Workshop on Mathematics, Computer Science, and Computational Sciences*, Las Cruces, New Mexico, April 2, 2022.
[9] E. Freudenthal and V. Mueller, *Summary of Vocal Intonation Boosting (VIB)*, https://www.freudensong.com/vib-white-paper
[10] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
[11] N. Kuga, R. Abe, K. Takano, Y. Ikegaya, and T. Sasaki, "Prefrontal-amygdalar oscillations related to social behavior in mice", *eLife*, 2022, Vol. 11, Paper e78428.
[12] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
[13] D. F. Mix, *Fourier, Laplace, and z Transforms: Unique Insight into Continuous-Time and Discrete-Time Transforms, Their Definition and Applications*, amazon.com, 2019.
[14] H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.
[15] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
[16] B. G. Osgood, *Lectures on the Fourier Transform and Its Applications*, American Mathematical Society, Providence, Rhode Island, 2019.
[17] T. Pendlebury, *How to Improve the Speech On Your TV to Make It More Understandable*, 2022, https://www.cnet.com/tech/home-entertainment/having-a-hard-time-hearing-the-tv-your-speaker-settings-may-be-the-culprit/
[18] S. K. Reed, *Cognition: Theories and Application*, SAGE Publications, Thousand Oaks, California, 2022.
[19] J. V. Stone, *The Fourier Transform: A Tutorial Introduction*, Sebtel Press, Sheffield, UK, 2021.
[20] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.