# A solution to the mystery of the sub-harmonic series and to the combination tone via a linear mathematical model of the cochlea

Ugo Boscain[1], Xiangyu Ma[1], Dario Prandi[2], Giuseppina Turco[3]

[1]Laboratoire Jacques-Louis Lions, Sorbonne Université, Université de Paris, CNRS, Inria, Paris, France.

[2] Laboratoire des Signaux et Systemes, Universite Paris-Saclay, CentraleSupélec, CNRS, Gif-sur-Yvette, France.

[3]Laboratoire de Linguistique Formelle, Université Paris Cité, CNRS, Paris, France.

*The authors contributed equally to this work.*

October 1, 2025

**Abstract**

In this paper, we study a simple linear model of the cochlea as a set of vibrating strings. We make hypothesis that the information sent to the auditory cortex is the energy stored in the strings and consider all oscillation modes of the strings. We show the emergence of the sub-harmonic series whose existence was hypothesized in the XVI century to explain the consonance of the minor chord. We additionally show how the nonlinearity of the energy can be used to study the emergence of the combination tone (Tartini's third sound) shedding new light on this long debated subject.

**Keywords**: string equation, harmonic and sub-harmonic series, combination-tone, psychoacoustics

## 1 Introduction

One of the most intriguing concepts in music theory is the so-called *sub-harmonic series*, also referred to the undertone series or hypotonic series, which is defined in a way that mirrors the well-known *harmonic series* (also called overtones series).

To fix ideas, we define a **genuine sound** as a *periodic* variation in air pressure. If $\tau$ is its period, we denote by $\mathcal{F} = 1/\tau$ its *fundamental frequency*. When the variation in air pressure is not strictly periodic, we speak more generally of a *sound*.

The *harmonic series* is well understood from a mathematical, physical, and psychoacoustic perspective. A genuine sound of frequency $\mathcal{F}$ can be decomposed into sinusoidal components at frequencies that are integer[1] multiples of $\mathcal{F}$ (see Figure 1), possibly with phase differences. This decomposition is well established physically (e.g., a piano string vibrates not only in its fundamental mode but also in higher modes), mathematically (Fourier series), and perceptually: the presence or absence of the various harmonic components determines the timbre of the original genuine sound. The $n$-th harmonic refers to the Fourier component at frequency $n\mathcal{F}$.

For instance, if the fundamental frequency of a genuine sound[2] corresponds to $C_4$ ($\mathcal{F} = 262$Hz), the harmonic series will produce the notes $C_5$ (the second harmonic of frequency $2\mathcal{F}$), $G_5$ (the third harmonic of frequency $3\mathcal{F}$), $C_6$ (the fourth harmonic of frequency $4\mathcal{F}$), $E_6$ (the fifth harmonic of frequency $5\mathcal{F}$), and so on—these are exactly the notes playable on a piccolo trumpet in C without engaging any valve.[3]

Since the birth of music theory, one of the most challenging problems has been to understand why certain combinations of genuine sounds are perceived as being consonant or dissonant, and whether this perception is shaped by cultural factors or by underlying physical phenomena. The presence of certain intervals (such as the

[1]Throughout this paper, the set of integers is understood not to include zero.

[2]In Scientific Pitch Notation (SPN), see [You39].

[3]It is important to note that knowledge of the intensity of the harmonics of a genuine sound is not sufficient to reconstruct the original sound. In fact, such reconstruction also requires knowledge of the relative phases. Interestingly, the human auditory system is generally insensitive to these relative phases for genuine sounds (although they can be perceived for more complex sounds) [Moo12, ZF99].

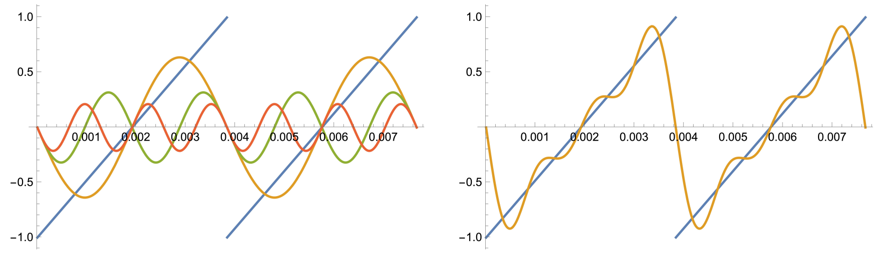


Figure 1: Two periods of a sawtooth signal with a frequency of 262Hz with its first 3 Fourier components (left) and with their sum (right).

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ... | $A\flat_1$ | $C_2$ | $F_2$ | $C_3$ | $C_4$ | $C_5$ | $G_5$ | $C_6$ | $E_6$ | $G_6$ | $B\flat_6$ | $C_7$ | $D_7$ | ... |
| | $\mathcal{F}/5$ | $\mathcal{F}/4$ | $\mathcal{F}/3$ | $\mathcal{F}/2$ | $\mathcal{F} = 262\text{Hz}$ | $2\mathcal{F}$ | $3\mathcal{F}$ | $4\mathcal{F}$ | $5\mathcal{F}$ | $6\mathcal{F}$ | $7\mathcal{F}$ | $8\mathcal{F}$ | $9\mathcal{F}$ | |

Figure 2: The harmonic and sub-harmonic series corresponding to the $C_4$

octave, the fifth, etc.) across many very different cultures has led to the idea that at least some basic intervals are perceived as consonant for physical reasons. This point of view was adopted by Zarlino (1517-1590) [Zar58], Descartes (1596-1650) [Des50], Rameau (1683-1764) [Ram22], Lagrange (1736- 1813) [Lag59, Lag62], Helmholtz (1821-1894) [Hel63], (see also [Pam89, Ste12, Cas15] for a historical perspective).

One of the most relevant chords, whose consonance has been justified using this idea, is the major triad. Actually, in this triad all constituent notes are present in the harmonic series of the root. For example, the C major triad (C–E–G) is derived from the first five harmonics. In this sense, the major triad is perceived as consonant because the auditory cortex is accustomed to hearing it within every genuine sound that is sufficiently rich in harmonics.

The question of consonance of the minor triad is more subtle. A possible explanation was already put forward by Zarlino in [Zar58]. He hypothesized the existence of the *sub-harmonic series*: the mirror image of the harmonic series, composed of sub-harmonics obtained by dividing the fundamental frequency by 2, 3, 4, etc. If we take again $C_4$ ($\mathcal{F} = 262\text{Hz}$) as the reference, the sub-harmonic series includes the notes $C_3$ (sub-harmonic 2 of frequency $\frac{1}{2}\mathcal{F}$), $F_2$ (sub-harmonic 3 of frequency $\frac{1}{3}\mathcal{F}$), $C_2$ (sub-harmonic 4 of frequency $\frac{1}{4}\mathcal{F}$), $A\flat_1$ (sub-harmonic 5 of frequency $\frac{1}{5}\mathcal{F}$), etc. (see Figure 2). The crucial point is that *the first five sub-harmonics generate an inversion of a minor triad* (in case of the $C_4$, an F minor chord: $A\flat_1$–$C_2$–$F_2$).[4] is hard to perceive this triad due to the strong presence of lower frequency components C and E.

This interpretation remained a matter of controversy for several centuries (see Rameau [Ram22], Hauptmann [Hau53], Oettingen [vO66], Riemann [Rie93], Lewin [Lew82], Levy [Lev85]), including the question of whether subharmonics are actually perceived, and the search of sub-harmonics in brass and string instruments when they are played in a nonconventional way [Abb64, Abb65, DP73]. See also [Pam89, Wie59, DP22].

Even though Zarlino and contemporaries hypothesize the existence of the sub-harmonic series, they could not know – due to limited physical knowledge at that time – that the sub-harmonic series is not physically present in the original genuine sound: mathematically, no Fourier components exist below the fundamental frequency. This, however, does not mean that Zarlino and contemporaries were focibly wrong, as the sub-harmonic series could be generated by the intricate structure of the inner ear or by the auditory cortex itself. The question of whether the perception of the sub-harmonic series is real or not remains an open question in psychoacoustics [Pam89].

It is worth noting that the sub-harmonic series is often associated with the perception of Tartini's third tone (or combination tone), which refers to the perceived frequency $|\mathcal{F}_1 - \mathcal{F}_2|$ (or more in general $|n_1\mathcal{F}_1 - n_2\mathcal{F}_2|$ with $n_1$ and $n_2$ integer[5]) when two genuine sounds of frequencies $\mathcal{F}_1$ and $\mathcal{F}_2$ are played simultaneously. When $\mathcal{F}_1$ is very close to $\mathcal{F}_2$, Tartini's third sound is simply a *beat*. Similarly to the sub-harmonic series, Tartini's third

[4] To retrieve the root of the F minor triad, the sixth sub-harmonic must be added. Adding further sub-harmonics blurs the root again. Note that minor triad notes also appear in the overtone series, though much later and mixed with other harmonics unrelated to the minor chord. For instance, the 6th, 7th, and 9th harmonics of $C_4$ correspond to a G minor chord ($G_6$–$B\flat_6$–$D_7$). Beside the fact that the $B\flat_6$ is too flat (with respect to the natural or tempered scale), it

[5] $|n_1\mathcal{F}_1 - n_2\mathcal{F}_2|$ is actually the combination tone between the $n_1$ harmonic of $\mathcal{F}_1$ and and the $n_2$ harmonic of $\mathcal{F}_2$. It is generated only if these harmonics are present in the original sound.

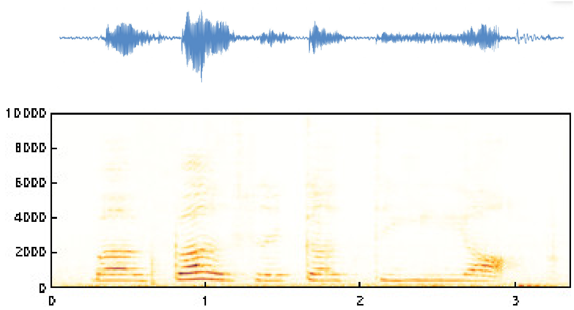


Figure 3: The sound (top) and its spectrogram (the "sound image", bottom) .

tone is not physically present (see [Moo12, CMC18] and reference therin) and starting from Helmholtz [Hel63] it was long thought to originate in the inner ear due to nonlinear phenomena. More recent research suggests that such a tone may be as well generated in deeper brain regions, as it is still perceived when the two tones are delivered separately to each ear via headphones [BBWB20].

As already noted by Helmholtz, the explanation of phenomena such as Tartini's third sound appears to require nonlinearities. [6]

In this paper, we show that both the sub-harmonic series and Tartini's third tone can in fact be explained within a linear cochlear framework based on vibrating strings, provided that the transmitted information from the cochlea to the auditory cortex is taken to be the energy stored in the strings. The nonlinearity required to explain Tartini's third tone is then inherent to the quadratic nature of the energy. We will return later to the relationship between Tartini's third tone and the sub-harmonic series.

**The cochlea.** The cochlea is a remarkable organ responsible for decomposing sound into its fundamental frequencies, thereby constructing the so-called spectrogram (see Figure 3).

Roughly speaking, signals transmitted from the cochlea to the primary auditory cortex A1 via the auditory nerve encode the amplitudes of various Fourier components (but not their relative phases). Due to its similarity with the visual cortex V1, the auditory cortex A1 has led to the term *sound image* for the spectrogram, reinforcing the idea that we do not hear sounds per se, but rather we see their sound images (see for instance [Moo12, BPST21]). The cochlea is a rather complex structure filled with a fluid called *endolymph* that is set in motion by incoming sound via the outer and middle ear (pinna, tympanic membrane, and auditory ossicles). (See Figure 4.)

These vibrations reach the *basilar membrane*, which can be modeled as a collection of tensioned strings of varying lengths linear mass and tension. We consider here the simple possible model in which the strings are **noninteracting** and are modeled by a simple **linear** damped **string equation**.

Each string has its own resonance frequency (see below for the precise definition), and the oscillations are damped due to various mechanisms, the most immediate being immersion in a viscous fluid. Strings closer to the entrance point of the cochlea (the *oval window*) have the highest resonance frequencies (around 20KHz), strings closer to the deepest point in the cochlea (the apex) have the lowest resonance frequencies (around 20Hz).

When a genuine sound is perceived, the various strings of the basilar membrane are forced into oscillation. These oscillations are maximal when the resonance frequency of a string is close to the frequency of the input sound.

The vibrations of different regions of the basilar membrane are detected by *internal hair cells*, which send signals roughly corresponding to the amplitude of the vibrations. Additional hair cells, known as *external hair cells*, serve to amplify vibrations but will not be discussed here.[7]

Let us parametrize the different strings by $x \in [0, L]$ where $0$ is the base and $L$ the apex. Roughly speaking, when the input is a genuine sound, internal hair cells transmit information on the amplitude of oscillations for

[6] It should be mentioned that there are several theories on the combination tone: Tartini's theory [Tar54, Tar67], Lagrange's theory [Lag59, Lag62] and – the most widely accepted – Helmholtz's theory [Hel63]. See, for instance, [Moo12, CMC18, Abb65, Abb72] for an historical perspective. Here, we just mention that in contrast to Helmholtz's theory which predicts $|\mathcal{F}_1 - \mathcal{F}_2|$ as main combination tone, Lagrange's theory predicts, assuming $\mathcal{F}_1/\mathcal{F}_2$ rational, the greatest common divisor between $\mathcal{F}_1$ and $\mathcal{F}_2$. See Section 6.1 for the definition of the greatest common divisor between two numbers that are not integers.

[7] In this paper we consider a passive model of the cochlea. For models including the active action of external hair cells, see for instance [MH99, ES12] and references therein.

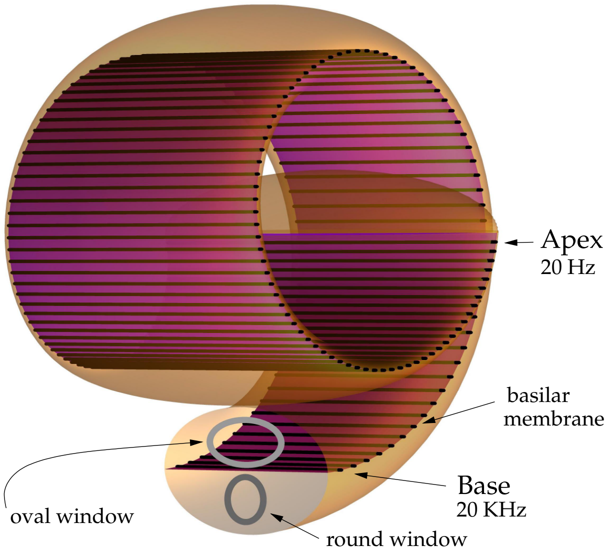


Figure 4: The Cochlea.

each string to the auditory cortex A1. Such information is encoded in a positive function $E(x,t)$ of the string in position $x$ at time $t$.

One of the crucial hypotheses we make in this paper is to quantify this concept. We will assume that:

**(P)** the signal sent to the auditory cortex by the cochlea is, for each string $x$, the energy $E(x,t)$ stored in it.

We now excite our model with a genuine sound and study the function $E(x,t)$ in the long-time limit, after the transient dynamics have decayed. Details are given in Sections 2.2.3 and 4. The function $E(x,t)$ has some important features:

- Roughly, the function $E(x,t)$ is almost constant in time and reaches its maximum at the value of $x$ corresponding to the string whose resonance frequency is close to the frequency of the incoming sound. This value corresponds to the main *peak* of the function. However, secondary peaks—though generally smaller—are also present. (See Figure 5.) Let us examine why.

  **Peaks near to the harmonics.** Assume that a genuine sound corresponding to $C_4$ ($\mathcal{F} = 262$Hz) is perceived. If this sound is not a perfect sinusoid, it can be decomposed into several harmonic components. These physically present harmonics induce oscillatory peaks in those strings whose natural frequencies are $2\mathcal{F}$, $3\mathcal{F}$, $4\mathcal{F}$, etc., enabling the perception of the timbre corresponding to the original genuine sound.

  **Peaks near to the sub-harmonics.** A vibrating string has different oscillation modes. If the fundamental mode corresponds to a frequency $\mathcal{F}_0$, the next modes correspond to frequencies $2\mathcal{F}_0$, $3\mathcal{F}_0$ etc. Hence, if a genuine sound corresponding to the $C_4$ ($\mathcal{F} = 262$Hz) is perceived, this will activate the first mode of the strings whose frequency of resonance is close to $\mathcal{F}$ but also the second mode of the strings whose frequency of resonance is close to $\frac{1}{2}\mathcal{F}$, the third mode of the strings whose frequency of resonance is close to $\frac{1}{3}\mathcal{F}$, etc... Notice that all these strings will vibrate at the frequency of the forcing signal ($\mathcal{F} = 262$Hz) but the information that is transmitted to the auditory cortex (being the energy stored in the strings only) is the presence of oscillations at the frequencies of $\frac{1}{2}\mathcal{F}$, $\frac{1}{3}\mathcal{F}$ etc. These frequencies correspond exactly to the frequency of the sub-harmonics and this idea explains their perception.[8]

  As we shall see, in the simplest (more natural though) model, **only the odd sub-harmonics** appear. These are the most important ones since they contain the minor triad (the even sub-harmonics would produce

[8] The same reasoning also applies to all the harmonics of the incoming sound, producing sub-harmonics of the different harmonics. See Section 5.

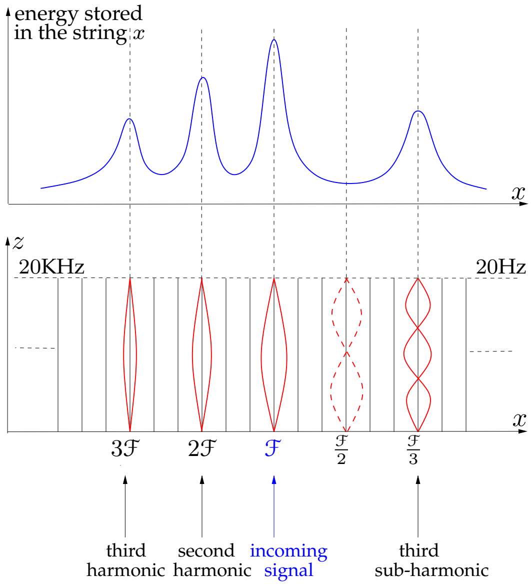


Figure 5: The basilar membrane (bottom) is modeled as a set of vibrating strings responding to an incoming genuine sound of frequency $\mathcal{F}$. The signal, which contains several harmonics (e.g., the sawtooth of Figure 1), excites the strings whose resonance frequencies are $\mathcal{F}$ (fundamental), $2\mathcal{F}$ (second harmonic), $3\mathcal{F}$ (third harmonic), and so on, as well as $\frac{\mathcal{F}}{3}$ (third sub-harmonic), $\frac{\mathcal{F}}{5}$ (fifth sub-harmonic), etc. As explained in Section 2.2, the sub-harmonics $\frac{\mathcal{F}}{2}$, $\frac{\mathcal{F}}{4}$, etc. are absent (the case $\frac{\mathcal{F}}{2}$ is shown with a dashed line). The upper panel is a qualitative illustration of the energy stored in the strings. Further details are given in Sections 4 and 5, in which we show that residual temporal oscillations occur and that sub-harmonics of the harmonics are also activated.

only lower octaves of the original sound). This fact is essentially due to certain symmetries of the model (see Section 2.2).

Notice that, contrary to harmonics, sub-harmonics are present even if the original sound is a pure sine wave. In fact, in the case of a pure sinusoid, they become even more evident, as there are no higher harmonics to mask their perception. This is one reason why pure sine waves – or, for example, the sound of a clarinet, which lacks all even-numbered harmonics – are perceived as particularly "deep" or dark in timbre.

- $E(x,t)$ increases with oscillation amplitude but is quadratic rather than linear in the input, which – as we will see – gives rise to interference phenomena. This fact permits to explain Tartini's third sound (which originates from nonlinearities) still working with a linear model.

  More precisely, as explained in Section 6, for a combination of two sinusoidal sounds of frequencies $\mathcal{F}_1 > \mathcal{F}_2$, the nonlinearity of the energy stored in each string $E(x,t)$ permits as well to see the appearence of residual oscillation in $E(x,t)$ of frequency $\frac{\mathcal{F}_1 - \mathcal{F}_2}{q}$, under the hypothesis that $\frac{\mathcal{F}_1+\mathcal{F}_2}{\mathcal{F}_1-\mathcal{F}_2} = p/q$ with $p$ and $q$ integer and $p/q$ an irreducible fraction.[9]

  For the most relevant combination tones, as for instance when $\mathcal{F}_1 = \frac{3}{2}\mathcal{F}_2$, we have $q = 1$ and hence the oscillations of $E(x)$ have frequency $\mathcal{F}_1 - \mathcal{F}_2$. Such oscillations in the energy of the strings are transmitted to the endolymph and returned back to the strings, producing the perception of the combination tone (notice, however, that such feedback is not described in our linear model). Our explanation of Tartini's third sound is different from those ones by Helmholtz ([Hel63], which predicts $\mathcal{F}_1 - \mathcal{F}_2$) and Lagrange ([Lag59, Lag62], which predicts the greatest common divisor of $\mathcal{F}_1$ and $\mathcal{F}_2$). The three theories provide the same combination tone for the most relevant pairs of input frequencies. But the presence of the factor

[9] This is due to the fact that the energy is, roughly, a quadratic function of the incoming signal. Hence, if the incoming signal is $\sin(\mathcal{F}_1 2\pi t) + \sin(\mathcal{F}_2 2\pi t)$, then the energy contains terms of the type $(\sin(\mathcal{F}_1 2\pi t) + \sin(\mathcal{F}_2 2\pi t))^2 = \sin^2(\mathcal{F}_1 2\pi t) + \sin^2(\mathcal{F}_2 2\pi t) + 2\sin(\mathcal{F}_1 2\pi t)\sin(\mathcal{F}_2 2\pi t)$. The last term can be written as $2\cos((\mathcal{F}_1 - \mathcal{F}_2)2\pi t) - 2\cos((\mathcal{F}_1 + \mathcal{F}_2)2\pi t)$ and the greatest common divisor between $\mathcal{F}_1 + \mathcal{F}_2$ and $\mathcal{F}_1 - \mathcal{F}_2$ under the hypothesis that $\frac{\mathcal{F}_1+\mathcal{F}_2}{\mathcal{F}_1-\mathcal{F}_2} = p/q$, with $p$ and $q$ coprime, is $\frac{\mathcal{F}_1-\mathcal{F}_2}{q}$.

$q$, when different from 1, shed a new light on the understanding of this mysterious phenomenon.[10]

In conclusion, this paper shows that the emergence of subharmonics and Tartini's third sound can be accounted for by a linear model of the basilar membrane, providing that two additional assumptions are made: **i)** the auditory cortex receives as input the energy stored in the various strings of the basilar membrane, and **ii)** not only the fundamental mode but also higher modes of oscillation must be taken into account. Thus, phenomena hypothesized as early as the 16th century are naturally predicted by a simple cochlear model. The description of the cochlea as a continuum of linear vibrating strings dates back to Helmholtz [Hel63] (see also [RH14] and references therein); these two aspects, however, have been largely overlooked. We should mention that a wide range of sophisticated models of the cochlea have been investigated, including nonlinear and active formulations (see, for instance, [Dui11, RR01, MH99, ES12] and references therein) as well as models focusing on traveling waves [Bék47, VW60, DJ03, Hul24]. In contrast to these approaches, our framework does not rely on the description of traveling waves, which are not particularly relevant for our purposes since we are mainly concerned with periodic input signals and their steady-state response. Moreover, while nonlinearities are essential to account for phenomena such as Tartini's third sound, in our model these are incorporated directly into the quantity transmitted to the auditory cortex – namely, the energy – while the underlying mechanical system is represented by a family of linear vibrating strings. While preserving the necessary non-linear effects, this choice has the significant advantage of keeping the model explicitly computable, as the mapping from the input signal to the corresponding energy can be written in closed form.

The paper is organized as follow. In Section 2 we introduce the model and describe its unforced oscillations, its response to a pure sinusoidal signal and its response to a genuine sound. In Section 3 we present our model for the signal sent to the auditory cortex, namely the energy stored in the different strings. In Section 4 we study the energy stored in the different strings under the action of a sinusoidal signal and we show the emergence of sub-harmonics. In Section 5 we study the energy stored in the different strings under the action of a genuine sound and show the presence of harmonics and sub-hamonics at the same time. In Section 6 we study the emergence of a combination tone when two sinusoidal signals of different frequencies are applied to the system.

Throughout the study, whenever possible we used experimentally validated parameters for the numerical simulation. Unfortunately, it is very hard to obtain these parameters experimentally and their knowledge is fairly limited. We use two sets of parameters. The first set is extracted from [NVTM03] and a second set is modified by us, to obtain a more reasonable range of frequency covered by the model and to fit an assumption allowing to make certain computations more explicitly (see hypothesis (A03) in Section 2.1). These choices are explained in Section 7 and in Appendix A. The paper is, however, written in such a way that other values of the relevant parameters can be easily used. The emergence of the sub-harmonic series and of the combination tone appear to be independent from the choice of parameters, at least in a reasonable range.

In Appendix B we give the proof of a crucial result concerning Tartini's third sound.

# 2 A model of the basilar membrane as a set of noninteracting vibrating strings

## 2.1 The physics

In the following we are going to model the basilar membrane as a set of non-interacting strings. The different strings are parameterized via their longitudinal position $x \in [0, L]$, where $0$ denotes the base and $L$ the apex. Each string has its own length $\ell(x)$, linear mass $\rho(x)$, tension $T(x)$ and damping $\gamma(x)$.

Let $z \in [0, \ell(x)]$ be the coordinate along one fixed string $x$. Let $u^x(t, z)$ the displacement of the string $x$ at position $z$ at time $t$. The external force acting on the strings, representing the sound, will be indicated by $F(z, t)$. Its coupling with the different strings could in principle depend on $x$ and will be indicated with $c(x)$.

Each string is going to be described by the following damped string equation

$$\rho(x)u^x_{tt} = T(x)u^x_{zz} - \gamma(x)u^x_t + c(x)F(t, z), \quad u^x(t, 0) = u^x(t, \ell(x)) = 0. \tag{1}$$

[10]The appearance of the combination tone is not restricted to the case in which $\mathcal{F}_1$ and $\mathcal{F}_2$ are rationally dependent. However, if this is not the case, the resulting sound is not a genuine sound (i.e., it is not described by a periodic function) and the problem should be treated differently.

For most of the results in this paper, it is not necessary to know the explicit expressions of $\rho(x), T(x), \gamma(x), c(x)$. However, it will be useful sometimes to have models for these functions, mainly for numerical simulations. According to [RH14], we have that (more details are given in Section7 and Appendix A):

- $\ell(x)$ is an increasing function of $x$. However, its variation is not very large. Along the paper we will often consider that $\ell$ is constant.
- The linear mass of each string is varying as a function of $x$ with an exponential law $\rho(x) = A_\rho e^{k_\rho x}$, with $A_\rho, k_\rho > 0$. Strings closer to the apex are much heavier.
- The tension of each string is varying as a function of $x$ with a negative exponential law $T(x) = A_T e^{-k_T x}$ with $A_T, k_T > 0$. Strings closer to the apex have less tension.
- The damping of each string is varying as a function of $x$ with an exponential law $\gamma(x) = A_\gamma e^{k_\gamma x}$, with $A_\gamma > 0$ and $k_\gamma \in \mathbf{R}$. See below for a discussion on the sign of $k_\gamma$.
- The function $c(x)$ is a decreasing function of $x$ modeling the fact that the energy of the sound is already partially dissipated when the sound reaches the most internal part of the cochlea (large $x$). However, the decrease of $c(x)$ is not very relevant and we will often consider the case $c = 1$.

In the following, any time we use explicit expressions for $\rho(x), T(x), \gamma(x), c(x), \ell(x)$, we will refer to the following assumption

**(A0)** We assume that $\rho(x) = A_\rho e^{k_\rho x}$, $T(x) = A_T e^{-k_T x}$, $\gamma(x) = A_\gamma e^{k_\gamma x}$, with $A_\rho$, $A_T$, $A_\gamma$, $k_\rho$, $k_T > 0$, $k_\gamma \in \mathbf{R}$. We also assume that $c$ and $\ell$ are positive constants and we normalize $c = 1$.

In the literature one also finds the following further assumptions in addition to (A0):

**(A01)** $k_\gamma = k_\rho = k_T$. See for instance [Hul24] and references therein.

**(A02)** $k_\gamma = \frac{k_T + k_\rho}{2}$ (to have that $\gamma$ is proportional to the inverse of the undumped frequency of the string defined below in (2)). See for instance [Hul24] and references therein.

**(A03)** $k_\gamma = 0$ (same damping for all the strings). This assumption, that we often make in the following, permits to simplify certain analyses. It is justified by the fact that sometimes in the literature one finds $k_\gamma > 0$ (meaning higher damping for lower frequencies, see for instance [Hul24, SCH15]) and somtines $k_\gamma < 0$ (meaning lower damping for lower frequencies, see for instance [NVTM03, NEAT14]). The case $k_\gamma = 0$ has also be considered in [ZLP76].

Let us define the *undamped fundamental angular frequency*[11] of each string (see Section 2.2.2 for the justification of this formula)

$$\xi(x) := \sqrt{\frac{T(x)}{\rho(x)}} \frac{\pi}{\ell(x)}. \tag{2}$$

As in all realistic models, along the paper we will assume that $\xi$ is a decreasing function of $x$. This permits us to parametrize solutions of (1) by $\xi$ which is actually the only important parameter in the following.
In particular, this is true under (A0), which implies

$$\xi(x) \underset{\substack{\uparrow \\ \text{under (A0)}}}{=} \sqrt{A_T/A_\rho}\, \frac{\pi}{\ell} e^{-\frac{k_T + k_\rho}{2} x} \tag{3}$$

**Remark 1. (Range of frequencies covered by the model)** *Let us re-write* (3) *as*

$$\xi(x) = \tilde{A} e^{-\tilde{k} x}, \tag{4}$$

$$\text{where } \tilde{A} = \sqrt{A_T/A_\rho}\, \frac{\pi}{\ell} \text{ and } \tilde{k} = \frac{k_T + k_\rho}{2} \tag{5}$$

*The corresponding frequency range in Hz is* $[\xi(L)/(2\pi), \xi(0)/(2\pi)] = [\tilde{A} e^{-\tilde{k} L}/(2\pi), \tilde{A}/(2\pi)]$.

[11] In this formula $\xi(x)$ is an angular frequency because it is measured in rad/sec. The frequency in Hertz is $\frac{\xi}{2\pi}$.

### 2.2 The Mathematics

Let us re-write (1) as

$$u_{tt} = \frac{T}{\rho}u_{zz} - \frac{\gamma}{\rho}u_t + \frac{c}{\rho}F(t,z), \quad u(t,0) = u(t,\ell) = 0, \tag{6}$$

(here by simplicity of notation we have dropped the dependence on $x$ of $u,\ \rho,\ T,\ \gamma,\ c,\ \ell$).

Let us take a regular enough forcing term. To fix the ideas let us take $F(t,\cdot) \in L^2([0,\ell])$ for a.e. $t \in \mathbf{R}$ and the map $t \mapsto \|F(t,\cdot)\|$ in $L^\infty(\mathbf{R})$, that is $F \in L^\infty(\mathbf{R}, L^2([0,\ell]))$. We consider initial conditions $u(0,\cdot) \in H^1_0([0,\ell])$ and $u_t(0,\cdot) \in L^2([0,\ell])$. In this setting, the existence and uniqueness of solutions $u \in L^\infty(\mathbf{R}, H^2([0,\ell]) \cap H^1_0([0,\ell])$ is standard. See for instance [Hal88, Section 4.7].

Considering the basis $e_n(z) = \sqrt{\frac{2}{\ell}}\sin(\frac{\pi}{\ell}nz)$ for $n = 1,2,\dots$ of $L^2([0,\ell]) \supset H^2([0,\ell]) \cap H^1_0([0,\ell])$ we have

$$u(t,z) = \sum_{n=1}^{\infty} p_n(t)e_n(z), \text{ where } p_n(t) = \langle e_n, u(t,z)\rangle = \int_0^\ell u(t,z)\sqrt{\frac{2}{\ell}}\sin(\frac{\pi}{\ell}nz)dz.$$

This yields the sequence of ODEs

$$\ddot{p}_n = \frac{T}{\rho}\left(-\frac{\pi^2}{\ell^2}n^2\right)p_n - \frac{\gamma}{\rho}\dot{p}_n + \frac{c}{\rho}f_n(t), \quad n = 1,2,3,\dots \tag{7}$$

where

$$f_n(t) = \langle e_n, F(t,z)\rangle = \int_0^\ell F(t,z)\sqrt{\frac{2}{\ell}}\sin(\frac{\pi}{\ell}nz)dz.$$

Recall that the force $F(t,z)$ is applied to the string through endolymph motion driven by the auditory ossicles. Neglecting boundary effects in the vestibular canal, the endolymph motion depends only on $x$. Thus, we assume:

**(A1)** The applied force is independent of $z$, i.e., $F(t,z) = F(t)$.

As we will see this assumption will be responsible for the absence of even sub-harmonics.

Under (A1) we have that

$$f_n(t) = F(t)\int_0^\ell \sqrt{\frac{2}{\ell}}\sin(\frac{\pi}{\ell}nz)dz = \begin{cases} F(t)\sqrt{\frac{2}{\ell}}\frac{2\ell}{\pi}\frac{1}{n} = F(t)\frac{2\sqrt{2}}{\pi}\sqrt{\ell}\frac{1}{n} & \text{if } n \text{ is odd,} \\ 0 \quad \text{if } n \text{ is even.} \end{cases}$$

Then (7) becomes

$$\ddot{p}_n \underset{\substack{\uparrow \\ \text{under (A1)}}}{=} -\xi_n(x)^2 p_n - \mu(x)\dot{p}_n + \nu_n(x)F(t), \quad n = 1,2,3,\dots \tag{8}$$

where

$$\xi_n^2(x) = \frac{T(x)}{\rho(x)}\frac{\pi^2}{\ell(x)^2}n^2 \qquad \Rightarrow \qquad \xi_n(x) = \sqrt{\frac{T(x)}{\rho(x)}}\frac{\pi}{\ell(x)}n =: \xi\, n$$

$$\mu(x) = \frac{\gamma(x)}{\rho(x)} \tag{9}$$

$$\nu_n(x) = \begin{cases} \frac{c(x)}{\rho(x)}\frac{2\sqrt{2}}{\pi}\sqrt{\ell(x)}\frac{1}{n} & \text{if } n \text{ is odd} \\ 0 \quad \text{if } n \text{ is even} \end{cases} \tag{10}$$

#### 2.2.1 Parameterization in function of $\xi$

If we can invert (2), we can re-write (8) in function of $\xi$

$$\ddot{p}_n \underset{\substack{\uparrow \\ \text{under (A1)}}}{=} -\xi^2 n^2 p_n - \mu(x(\xi))\dot{p}_n + \nu_n(x(\xi))F(t), \quad n = 1,2,3,\dots \tag{11}$$

Under (A0) this can be done explicitly and we obtain $x(\xi) = \frac{2}{k_T + k_\rho}\log\left(\frac{\pi}{\ell}\sqrt{\frac{A_T}{A_\rho}}\frac{1}{\xi}\right)$. As a consequence

$$\mu(x(\xi)) \underset{\substack{\uparrow\\ \text{under (A0)}}}{=} \frac{A_\gamma}{A_\rho}\left(\sqrt{\frac{A_\rho}{A_T}}\frac{\ell}{\pi}\xi\right)^{\frac{2(k_\rho-k_\gamma)}{k_\rho+k_T}} =: B_\mu \xi^\alpha, \quad \text{where } \alpha := \frac{2(k_\rho-k_\gamma)}{k_\rho+k_T} \tag{12}$$

$$\nu_n(x(\xi)) \underset{\substack{\uparrow\\ \text{under (A0)}}}{=} \begin{cases} \frac{1}{A_\rho}\left(\sqrt{\frac{A_\rho}{A_T}}\frac{\ell}{\pi}\xi\right)^{\frac{2k_\rho}{k_\rho+k_T}} \frac{2\sqrt{2}}{\pi}\sqrt{\ell}\frac{1}{n} =: B_\nu \xi^{\frac{2k_\rho}{k_\rho+k_T}}\frac{1}{n} & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases} \tag{13}$$

Next we will also need

$$T(x(\xi)) \underset{\substack{\uparrow\\ \text{under (A0)}}}{=} A_T\left(\sqrt{\frac{A_\rho}{A_T}}\frac{\ell}{\pi}\xi\right)^{\frac{2k_T}{k_\rho+k_T}} =: B_T \xi^{\frac{2k_T}{k_\rho+k_T}} \tag{14}$$

In the following, with abuse of notation, we write $\mu(\xi)$, $\nu_n(\xi)$ and $T(\xi)$ in place of $\mu(x(\xi))$, $\nu(x(\xi))$ and $T(x(\xi))$.

Notice that

- Under (A01) we have $\alpha = 0$ and $\mu(\xi) = B_\mu,\ \ T(\xi) = B_T\xi,\ \ \nu_n(\xi) = B_\nu \frac{\xi}{n}$ (for $n$ odd and zero otherwise).
- Under (A02) we have $\alpha = \frac{k_\rho - k_T}{k_\rho + k_T} \in (-1, 1)$.
- Under (A03) we have $\alpha \in (0, 2)$ and $\mu(\xi) = B_\mu \xi^\alpha,\ T(\xi) = B_T \xi^{\frac{2k_T}{k_\rho+k_T}},\ \nu_n(\xi) = B_\nu \frac{\xi^\alpha}{n}$ (for $n$ odd and zero otherwise).

**Remark 2.** *For the undamped case (i.e., $\xi > \mu(\xi)/2$) to occur at high frequencies, it is natural to assume that $\alpha < 1$ (see Section 2.2.2). Actually, this is always the case for the set of parameters given in Section 7 and Appendix A*

### 2.2.2 Unforced case

Let us discuss briefly the solution of the equation (11) in the case in which there is no forcing term ($F(t) = 0$), i.e.

$$\ddot{p}_n = -\xi^2 n^2 p_n - \mu(\xi)\dot{p}_n \quad n = 1, 2, 3, \ldots \tag{15}$$

This can be alternatively written as a first order equation

$$\begin{pmatrix} \dot{p}_n \\ \dot{q}_n \end{pmatrix} = M_n \begin{pmatrix} p_n \\ q_n \end{pmatrix} \text{ where } M_n = \begin{pmatrix} 0 & n\xi \\ -n\xi & -\mu(\xi) \end{pmatrix}, \quad n = 1, 2, 3, \ldots \tag{16}$$

- If there is no damping $\mu(\xi) = 0$, the solutions are oscillations of angular frequency $n\xi$. This is why $\xi$ has been called the undamped fundamental angular frequency of the string.

In the presence of damping, since the eigenvalues of the matrix $M_n$ are

$$-\frac{\mu(\xi)}{2} \pm i\sqrt{n^2\xi^2 - \left(\frac{\mu(\xi)}{2}\right)^2}, \tag{17}$$

we have the following cases:

- (underdamped case) $\xi > \frac{\mu(\xi)}{2}$ then for every $n$ the solutions are oscillations of angular frequency $\omega_n = \sqrt{n^2\xi^2 - \left(\frac{\mu(\xi)}{2}\right)^2}$ and of exponentially decreasing amplitude. We call $\omega_n$ the $n$-resonance angular frequency of the string.
- (overdamped case) $\xi < \frac{\mu(\xi)}{2}$ then the above is true for $n$ large enough only. For $n \leq \frac{\mu(\xi)}{2\xi}$ the solutions go to zero without oscillating.

### 2.2.3 Asymptotic response to a sinusoidal signal

In this section, we apply the simplest possible sound (i.e., a sinusoidal function) to our string equation 11, and we study the corresponding solution for large time, once the solution has become periodic. Studying this case permits to see the emergence of the sub-harmonic series.

To this purpose, we start directly from the decomposition in Fourier modes under (A1)

$$\ddot{p}_n \underset{\substack{\uparrow \\ \text{under (A1)}}}{=} -\xi^2 n^2 p_n - \mu(\xi)\dot{p}_n + \nu_n(\xi)F(t), \quad n = 1, 2, 3 \dots \tag{18}$$

with

$$F(t) = \sin(\mathbf{k}t + \varphi) \tag{19}$$

Remarkably for every initial condition $p_n(t_0) = a$, $\dot{p}_n(t_0) = b$ the solution of (18) converges exponentially to the periodic solution (see e.g., [Hal88])

$$p_n(t) \underset{\substack{\uparrow \\ F(t) = \sin(\mathbf{k}t+\varphi)}}{=} -\nu_n(\xi)\frac{\left(\left(\mathbf{k}^2 - \xi^2 n^2\right)\sin(\mathbf{k}t + \varphi) + \mathbf{k}\mu(\xi)\cos(\mathbf{k}t + \varphi)\right)}{\mathbf{k}^4 + \mathbf{k}^2\left(\mu(\xi)^2 - 2\xi^2 n^2\right) + \xi^4 n^4}, \quad n = 1, 2, 3 \dots \tag{20}$$

In other words, this solution is globally attractive and periodic, and considering it amounts to ignoring the transient in the cochlear response.

Notice that due to (10), we have that $p_n(t) = 0$ for $n$ even. We can write (20) in the form[12]

$$p_n(t) \underset{\substack{\uparrow \\ F(t) = \sin(\mathbf{k}t+\varphi)}}{=} \mathcal{R}_n(\xi, \mathbf{k})\sin(\mathbf{k}t + \varphi + \phi_n(\xi, \mathbf{k})), \quad n = 1, 2, 3 \dots \tag{21}$$

where

$$\mathcal{R}_n(\xi, \mathbf{k}) = \nu_n(\xi)\frac{1}{\sqrt{\mathbf{k}^4 + \mathbf{k}^2\left(\mu(\xi)^2 - 2\xi^2 n^2\right) + \xi^4 n^4}} \tag{22}$$

$$\phi_n(\xi, \mathbf{k}) = \arctan\frac{\mathbf{k}\mu}{\mathbf{k}^2 - n^2\xi^2} \tag{23}$$

**Remark 3.** *Under forcing, all strings oscillate with angular frequency* $\mathbf{k}$*, even those ones for which we are in the over-damped case.*

**Remark 4.** *The function* $\mathcal{R}_n$ *is always well defined for every (positive) value of* $n, \xi, \mathbf{k}$ *and every function* $\nu(\xi)$, $\mu(\xi)$*.*

If, in addition to $F(t) = \sin(\mathbf{k}t + \varphi)$, we assume (A0), we have

$$\mathcal{R}_n(\xi, \mathbf{k}) \underset{\substack{\uparrow \\ \text{under (A0)}}}{=} \begin{cases} \frac{B_\nu \xi^{\frac{2k_\rho}{k_\rho + k_T}}}{n} \frac{1}{\sqrt{\mathbf{k}^4 + \mathbf{k}^2((B_\mu \xi^\alpha)^2 - 2\xi^2 n^2) + \xi^4 n^4}} & n = 1, 3, 5, \dots \\ 0 & n = 2, 4, 6, \dots \end{cases} \tag{24}$$

**Study of $\mathcal{R}_n(\xi, \mathbf{k})$ for $n$ odd**

- For fixed $\xi$ its maximum is at $\mathbf{k}_n^{\max} = \sqrt{n^2\xi^2 - \frac{\mu(\xi)^2}{2}}$. Notice that this is not the $n$-resonance angular frequency of the string, which is $\omega_n = \sqrt{n^2\xi^2 - \left(\frac{\mu(\xi)}{2}\right)^2}$
- For fixed $\mathbf{k}$, its maximum cannot be computed explicitly without an explicit expression of $\mu(\xi)$. However, we have the remarkable result that under (A03) its maximum is at

$$\xi_n^{\max} \underset{\substack{\uparrow \\ \text{under (A03)}}}{=} \mathbf{k}/n.$$

  If we assume (A01) we have that the maximum is a bit displaced:

$$\xi_n^{\max} \underset{\substack{\uparrow \\ \text{under (A01)}}}{=} \frac{\sqrt{\mathbf{k}}\sqrt[4]{B_\mu^2 + \mathbf{k}^2}}{n}.$$

[12]Using the fact that $\sin(\alpha)\cos(\mathbf{k}t) + \cos(\alpha)\sin(\mathbf{k}t) = \sin(\alpha + kt) \Rightarrow A\sin(\alpha)\cos(\mathbf{k}t) + A\cos(\alpha)\sin(\mathbf{k}t) = A\sin(\alpha + kt)$ and writing $A\cos(\alpha) = k^2 - \xi_n^2$ and $A\sin(\alpha) = k\mu$.

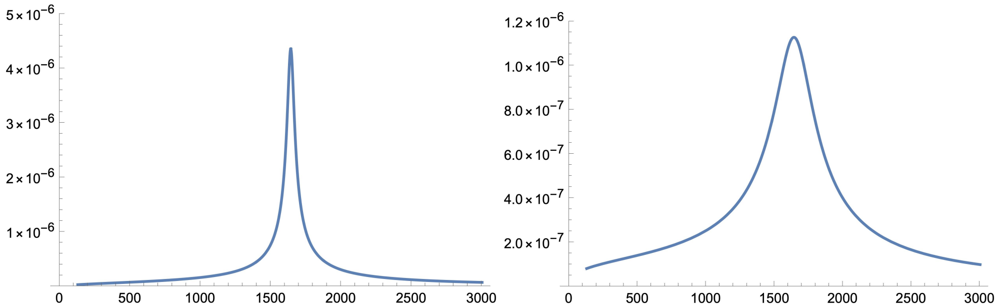


Figure 6: The function $\mathcal{R}_1$ for an input that is a sinusoidal signal corresponding to the $C_4$ for the set of parameters of Section 7.1 and of Section 7.2.

**Remark 5.** *Notice that each string has three different frequencies associated with it:*

- *Its undamped resonance angular frequency* $\xi$.
- *Its resonance angular frequency* $\omega_1 = \sqrt{\xi^2 - \left(\frac{\mu(\xi)}{2}\right)^2}$.
- *The angular frequency of an external signal provoking the largest oscillations* $\mathbf{k}_1^{\max} = \sqrt{\xi^2 - \frac{\mu(\xi)^2}{2}}$.

In Figure 6, we depict $\mathcal{R}_1(\xi, 2\pi 262)$ – i.e., we consider a sinusoidal input signal that corresponds to $C_4$ – for the set of parameters of Section 7.1 (parameters of [NVTM03]) and of Section 7.2 (modified parameters and assumption A03). Notice the similar qualitative behavior.

#### 2.2.4 Asymptotic response to a genuine sound

Consider now a periodic signal $F(t)$ with angular frequency $\mathbf{k} > 0$. Under the hypothesis given in Section 2.2 we have that that $F|_{[0,\frac{2\pi}{\mathbf{k}}]} \in L^2[0, \frac{2\pi}{\mathbf{k}}]$. Then $F(t)$ can be decomposed in its Fourier components

$$F(t) = \sum_{j=1}^{\infty} c_j \sin(j\mathbf{k}t + \varphi_j)$$

(here for simplicity we are assuming that $F(t)$ has zero average, so that $j$ starts from 1).

By linearity, the corresponding response is

$$p_n(t) = \sum_{j=1}^{\infty} c_j \mathcal{R}_n(\xi, j\mathbf{k}) \sin(j\mathbf{k}t + \varphi_j + \phi_n(\xi, j\mathbf{k})), \quad n = 1, 2, 3, \ldots \tag{25}$$

where $\mathcal{R}_n$ and $\phi_n$, are given by (22) and (23).

## 3 A model for the information sent to the auditory cortex

The observed insensitivity to the relative phases of the Fourier components (see, for instance, [Moo12, ZF99]) is a strong indication that the signal transmitted from the cochlea to the auditory system is not the full solution $u(t,z)^x$ of equation (1), but rather a smoothed version in which fast oscillations are averaged out.

For instance, in the case of the response to a sinusoidal input as discussed in Section 2.2.3, what is sent from the cochlea is not the complete solution (here expressed through its Fourier coefficients) $p_n(t)$ for $n = 1, 2, 3, \ldots$, but rather a quantity only related to the amplitude of the oscillations $\mathcal{R}_n(\xi, \mathbf{k})$.

For a signal more complex than a pure sinusoid (e.g., the sum of two sinusoids with different frequencies), it is not immediately clear how to perform such averaging of the fast oscillations.

In the following, we introduce a natural assumption that accounts for at least part of this averaging process, and that leads to the emergence of the sub-harmonic series in a very natural way:

(P) **The signal transmitted from the cochlea to the auditory cortex is the energy stored in each string.**

Referring to (1), such energy is

$$E(x,t) = \int_0^{\ell(x)} \Big(\frac{1}{2}\rho(x)(u_t^x)^2 + \frac{1}{2}T(x)(u_z^x)^2\Big)\, dz \tag{26}$$

The first term is the kinetic energy due to the movement of the strings and the second term is the potential (elastic) energy.

Decomposing in the Fourier basis used in Section 2.2.3 and writing $x$ in function of $\xi$ (with the usual abuse of notation) we obtain

$$\begin{aligned} E(\xi,t) &= \sum_{n=1}^{\infty} \frac{1}{2}\rho(\xi)\Big(\dot p_n(t)^2 + \xi_n^2 p_n(t)^2\Big) \\ &= \sum_{n=1}^{\infty} \frac{1}{2}\rho(\xi)\xi^2 n^2\Big(p_n(t)^2 + \frac{1}{\xi^2 n^2}\dot p_n(t)^2\Big) \\ &= \frac{1}{2}\frac{\pi^2}{\ell(\xi)^2}T(\xi)\sum_{n=1}^{\infty} n^2\Big(p_n(t)^2 + \frac{1}{\xi^2 n^2}\dot p_n(t)^2\Big) \\ &= \sum_{n=1}^{\infty} E_n(\xi,t), \end{aligned}$$

where

$$E_n(\xi,t) = \frac{1}{2}\frac{\pi^2}{\ell(\xi)^2}T(\xi)n^2\Big(p_n(t)^2 + \frac{1}{\xi^2 n^2}\dot p_n(t)^2\Big), \tag{27}$$

is the energy stored at time $t$ in the Fourier mode $n$ by the string whose undamped resonance angular frequency is $\xi$.

# 4 Perception of a sinusoidal signal and emergence of the sub-harmonic series

In this section, we are in the same framework as Section 2.2.3, i.e., we assume $F(t,z) = \sin(\mathbf{k}t + \varphi)$ and, in particular, we are working under (A1)). Using (21) we get that the signal transmitted by the cochlea to the auditory cortex is in this case

$$\begin{aligned} E(\xi,t) &= \sum_{n=1,2,3,\ldots\ldots} E_n(\xi,t) \text{ where} \\ E_n(\xi,t) &\underset{F(t)\,=\,\sin(\mathbf{k}t+\varphi)}{\underset{\uparrow}{=}} \frac{1}{2}\frac{\pi^2}{\ell(\xi)^2}T(\xi)n^2\mathcal{R}_n(\xi,\mathbf{k})^2\Big(\sin^2(\mathbf{k}t+\varphi+\phi_n(\xi,\mathbf{k})) + \frac{\mathbf{k}^2}{\xi^2 n^2}\cos^2(\mathbf{k}t+\varphi+\phi_n(\xi,\mathbf{k}))\Big) \\ &= \frac{1}{2}\frac{\pi^2}{\ell(\xi)^2}T(\xi)n^2\mathcal{R}_n(\xi,\mathbf{k})^2\left(1 + \left(\frac{\mathbf{k}^2}{\xi^2 n^2} - 1\right)\cos^2(\mathbf{k}t+\varphi+\phi_n(\xi,\mathbf{k}))\right). \end{aligned} \tag{28}$$

In the following, we call $E_n(\xi,t)$ the $n$th sub-harmonic.

**Remark 6.** *Note that the series* (28) *is converging due to the term $n^4$ in formula* (22)*. Moreover, for fixed $n = 1,2,\ldots$, $\mathbf{k},\xi > 0$ , $E_n(\xi,t)$ is always positive.*

**Remark 7.** *Note that only the odd sub-harmonics are present. In addition, for $n$ odd, we have that $E_n(\xi,t)$ is independent of time when $\xi = \mathbf{k}/n$, i.e., when we are considering a string whose undamped resonance angular frequency (or one of its odd multiples) coincides with the angular frequency of the external signal. Around the same value $\xi = \mathbf{k}/n$ the function $\mathcal{R}_n(\xi,\mathbf{k})$ has a maximum. For $\xi \neq \mathbf{k}/n$ the function $E_n(\xi,t)$ has residual oscillations with angular frequency $2\mathbf{k}$. This is due to the fact that frequency of $\cos^2(\mathbf{k}t)$ is twice the frequency of $\cos(\mathbf{k}t)$. The same occurs for every odd sub-harmonics, hence the full series $E(\xi,t)$ oscillates with angular frequency $2\mathbf{k}$.*

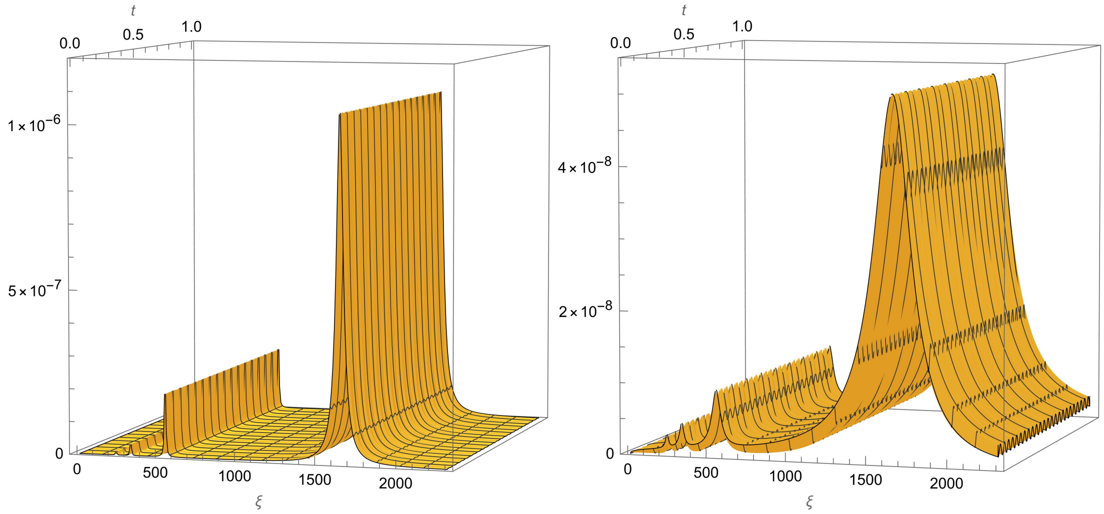


Figure 7: The energy stored in the strings $E(\xi, t)$ for an input that is a sinusoidal signal corresponding to the $C_4$ (i.e., for $\mathbf{k} = 2\pi 262$ sec$^-1$, $\varphi = 0$) for the choice of parameters reported in Section 7.1 and Section 7.2. The higher peak corresponds to the frequency of the external signals. The lower peaks correspond to the sub-harmonics.

Summing over $n$, the presence of these maxima produces several peaks, one for every sub-harmonic, as explained in more detail next.

If we assume (A0) and use (14) and (24), we get

$$E(\xi, t) = \sum_{n=1,3,5,\ldots} E_n(\xi, t) \text{ where}$$

$$\begin{aligned}
E_n(\xi, t) \underset{\substack{\uparrow \\ F(t)=\sin(\mathbf{k}t+\varphi) \text{ and under (A0)}}}{=} & \frac{1}{2}\frac{\pi^2}{\ell^2} B_T \xi^{\frac{2k_T}{k_\rho + k_T}} n^2 \left( \frac{B_\nu \xi^{\frac{2k_\rho}{k_\rho+k_T}}}{n} \frac{1}{\sqrt{\mathbf{k}^4 + \mathbf{k}^2\left((B_\mu \xi^\alpha)^2 - 2\xi^2 n^2\right) + \xi^4 n^4}} \right)^2 \times \\
& \left(1 + \left(\frac{\mathbf{k}^2}{\xi^2 n^2} - 1\right) \cos^2(\mathbf{k}t + \varphi + \phi_n(\xi, \mathbf{k}))\right) \\
= & \frac{1}{2}\frac{\pi^2}{\ell^2} B_T B_\nu^2 \xi^{2\frac{2k_\rho + k_T}{k_\rho + k_T}} \frac{1}{\mathbf{k}^4 + \mathbf{k}^2\left((B_\mu \xi^\alpha)^2 - 2\xi^2 n^2\right) + \xi^4 n^4} \times \\
& \left(1 + \left(\frac{\mathbf{k}^2}{\xi^2 n^2} - 1\right) \cos^2(\mathbf{k}t + \varphi + \phi_n(\xi, \mathbf{k}))\right). \qquad (29)
\end{aligned}$$

In Figure 7, we depict the energy stored in the strings $E(\xi, t)$ for an input that is a sinusoidal signal corresponding to the $C_4$.

For better visibility of the sub-harmonic series, in Figure 8 we draw $E(\xi, t)$ for $t = 0$ (the variations in $t$ are much smaller than the variations in $\xi$).

To show that the variations in $t$ are small, in Figure 9 we show $E(\xi, t)$ for fixed $\xi$.

**Remark 8.** *Notice that for fixed $\xi$, $E(\xi, t)$ is periodic with angular frequency $2\mathbf{k}$. In the same spirit as what is explained in Section 6, this oscillation is transmitted to the endolymph and then back to the forcing term. Although this phenomenon is not described by our linear model, it is going to produce an additional peak at $2\mathbf{k}$. This peak reinforces the perception of the second harmonic of the original sound.*

# 5 Perception of a periodic signal

When $F(t)$ is a periodic signal, the corresponding energy $E(\xi, t) = \sum_{n=1}^{\infty} E_n(\xi, t)$ is obtained by plugging (25) into (27). Notice, however, that $E(\xi, t)$ is not a linear function of $p_n(t)$. As a result, the contributions of the different Fourier components of the original sound cannot be simply added; instead, they interact and give rise

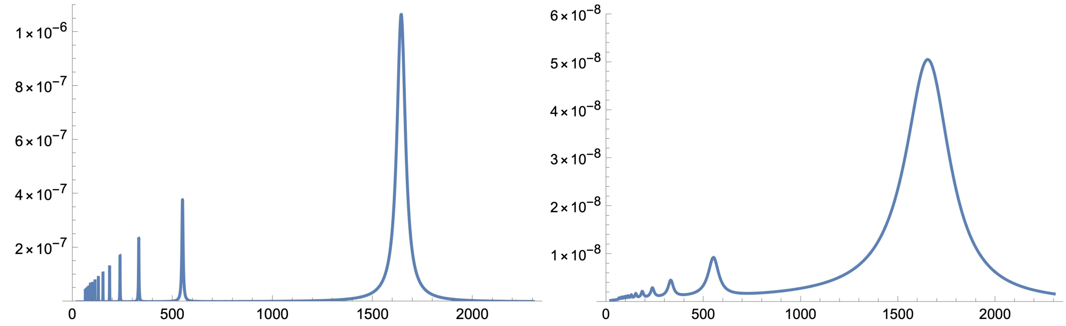


Figure 8: $E(\xi, t)$ for $t = 0$ for the same parameters of Figure 7

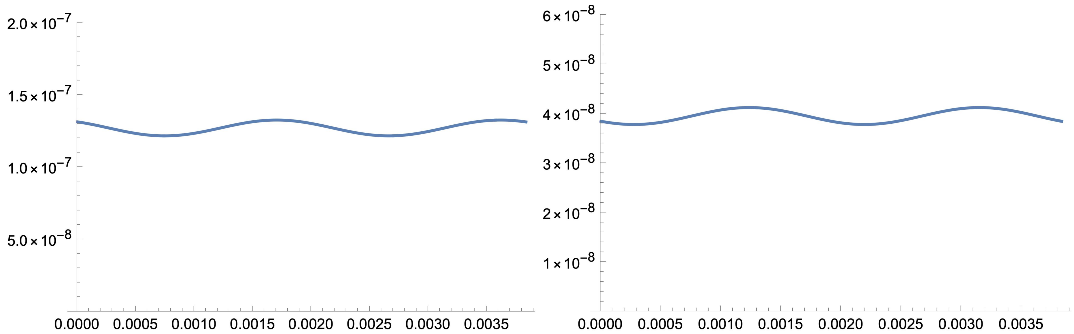


Figure 9: For the same parameters of Figure 7, we depict here $E(\xi, t)$ for $\xi = 2\pi 250 \sim 1570$ (to be close to the main peak but not exactly on it) for $t \in [0, 2\pi/\mathbf{k}]$. Notice that such variations in $t$ are periodic with an angular frequency $2\mathbf{k}$ and that they are small.

to interference phenomena. In other words, *sub-harmonics can be summed, but harmonics cannot*. More details concerning this nonlinearity are given in Section 6.

In Figure 10 we draw $E(\xi, t)$ for a sawtooth corresponding to the $C_4$ for the parameters given in Section 7.1. Notice the emergence of harmonics and sub-harmonics.

# 6 More on the nonlinearity of the response: combination tone

In this section, we study in more detail $E(\xi, t)$ when the input signal is the sum of two sinus functions:

$$F(t) = c_1 \sin(\mathbf{k}_1 t + \varphi_1) + c_0 \sin(\mathbf{k}_0 t + \varphi_2) \tag{30}$$

The main purpose here is to show the emergence of the combination tone when $\mathbf{k}_1$ and $\mathbf{k}_0$ are in a certain relation. It is also a starting point to study the response to a genuine sound with several Fourier components as in Section 2.2.4.

Using the linearity of $p_n$ w.r.t. $F(t)$ in (18), we have that (cf. (21))

$$p_n(t) = c_1 \mathcal{R}_n(\xi, \mathbf{k}_1) \sin(\mathbf{k}_1 t + \varphi_1 + \phi_n(\xi, \mathbf{k}_1)) + c_0 \mathcal{R}_n(\xi, \mathbf{k}_0) \sin(\mathbf{k}_0 t + \varphi_2 + \phi_n(\xi, \mathbf{k}_0)).$$

Then $E(\xi, t) = \sum_{n=1,2,3,\ldots} E_n(\xi, t)$ where $E_n(\xi, t)$ is given by Formula (27) that we report here by convenience

$$E_n(\xi, t) = \frac{1}{2} \frac{\pi^2}{\ell(\xi)^2} T(\xi) n^2 \Big( p_n(t)^2 + \frac{1}{\xi^2 n^2} \dot{p}_n(t)^2 \Big). \tag{31}$$

We have the following result proved in Appendix B.

**Proposition 1.** *Let $\mathbf{k}_1 > \mathbf{k}_0 > 0$. Assume that $\frac{\mathbf{k}_1 + \mathbf{k}_0}{\mathbf{k}_1 - \mathbf{k}_0} = \frac{p}{q}$ with $p > q$ integers and $\frac{p}{q}$ an irreducible fraction. Then, for every $n = 1, 3, 5, \ldots$ and $\xi > 0$, $E_n(\xi, t)$ (and as a consequence $E(\xi, t) = \sum_{n=1,3,5,\ldots} E_n(\xi, t)$) is periodic with period*

$$T_{\text{combination}} = \frac{2\pi}{\mathbf{k}_1 - \mathbf{k}_0} q, \tag{32}$$

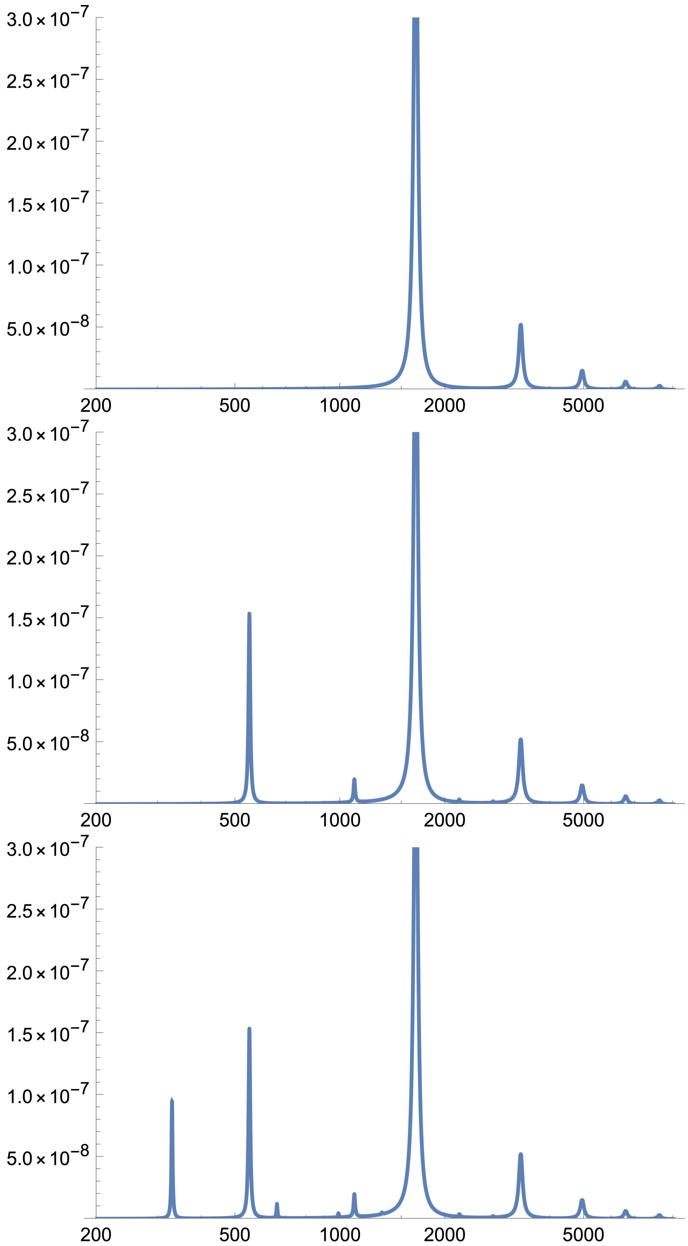


Figure 10: $E(\xi, t)$ for a sawtooth corresponding to the $C_4$ for the parameters given in Section 7.1. For better visibility, in the picture we have fixed $t = 0$ (being the variations in $t$ small w.r.t. variations in $\xi$). **First picture:** $E_1(\xi, t)$. Here, one sees the peaks corresponding to the harmonics of the original sound (the highest is the fundamental angular frequency of the sound). **Second picture:** $E_1(\xi, t) + E_2(\xi, t)$. Here, one sees the peaks corresponding to the harmonics of the original sound and of the second sub-harmonics (of the original sound and of its harmonics). **Third picture:** $E_1(\xi, t) + E_2(\xi, t) + E_3(\xi, t)$. Here, one sees the peaks corresponding to the harmonics of the original sound and of the second and third sub-harmonics. Notice the presence of sub-harmonics of the different harmonics.

**Discussion** When a signal of the form (30) is forcing our model (11), due to linearity, the strings whose undamped resonance frequency have a period close to (32) are not activated more than the other. However, the full energy stored in the system is oscillating with a period (32). One could consider a nonlinear model in which the oscillations of the total energy are transmitted to the endolymph and then back to the forcing term. Although this is out of the purpose of this paper, it would produce an additional peak around the angular frequency

$$\mathbf{k}_{\text{combination}} = \frac{\mathbf{k}_1 - \mathbf{k}_0}{q}.$$

Notice that in the psychoacoustic literature (see for instance [Moo12, CMC18]), the integer $q$ is not present.

Let us see what happens what happens with $\mathbf{k}_1 = \frac{3}{2}\mathbf{k}_0$ – an interval of a **fifth** – which are known to produce an easily perceivable combination tone. [13] In this case $\frac{\mathbf{k}_1+\mathbf{k}_0}{\mathbf{k}_1-\mathbf{k}_0} = 5$ hence $q = 1$ and the perceived angular frequency of the combination tone is the one found in the literature namely $\mathbf{k}_1 - \mathbf{k}_0$.

In table 12 the combination tones generated by natural intervals in between one octave are analyzed.[14] For the columns related to Helmohlz and Lagrange see Section 6.1. Notice that for most intervals $q = 1$.

## 6.1 Comparison with Helmholz's and Lagrange's theory

Let us assume that $\mathbf{k}_1 = \frac{u}{w}\mathbf{k}_0$ with $u > w$ integers and coprime. We have that

$$\frac{\mathbf{k}_1 + \mathbf{k}_0}{\mathbf{k}_1 - \mathbf{k}_0} = \frac{u+w}{u-w} =: \frac{p}{q}, \text{ with } p \text{ and } q \text{ coprime.}$$

Recall that the angular frequencies of the combination tone given by Helmoltz's theory, Lagrange's theory and our theory, are (here GCD indicates the greatest common divisor):

$$\mathbf{k}^H_{\text{combination}} = \mathbf{k}_1 - \mathbf{k}_0 = \frac{u-w}{w}\mathbf{k}_0, \tag{33}$$

$$\mathbf{k}^L_{\text{combination}} = \text{GCD}(\mathbf{k}_0, \mathbf{k}_1) := \frac{\mathbf{k}_0}{w}\text{GCD}(w, u) = \frac{1}{w}\mathbf{k}_0, \tag{34}$$

$$\mathbf{k}_{\text{combination}} = \frac{\mathbf{k}_1 - \mathbf{k}_0}{q} = \frac{u-w}{w}\frac{1}{q}\mathbf{k}_0. \tag{35}$$

Now, Let us set $h = \text{GCD}(u+w, u-w)$, which implies that $u + w = hp$ and $u - w = hq$. We distinguish two cases.

**Case 1.** *At least one between u and w is even.* In this situation $h = 1$, hence $q = u - w$ and

$$\mathbf{k}^H_{\text{combination}} = \frac{u-w}{w}\mathbf{k}_0, \quad \mathbf{k}^L_{\text{combination}} = \mathbf{k}_{\text{combination}} = \frac{1}{w}\mathbf{k}_0. \tag{36}$$

This holds true, for instance, when $u - w = 1$, in which case we have

$$\mathbf{k}^H_{\text{combination}} = \mathbf{k}^L_{\text{combination}} = \mathbf{k}_{\text{combination}} = \frac{1}{w}\mathbf{k}_0. \tag{37}$$

[13] In the literature, see for instance [Moo12], the combination tone generated by the interval of fifth is considered particularly easy to perceive because $\mathbf{k}_1 - \mathbf{k}_0 = 2\mathbf{k}_0 - \mathbf{k}_1$ meaning that the combination tone generated by the two fundamentals is the same as the one generated by the fundamental of the higher sound and the first harmonic of the lower one. For this to happen, one needs sounds for which the first harmonic of the lower sound is present. See, for instance, [Moo12].

[14] The intervals of the natural scale are constructed in the following way. Let us start from a given note $\mathbf{k}_0$ (for instance corresponding to a $C_4$). The octave $\mathbf{k}_1/\mathbf{k}_0 = 2$ is its first harmonic. The fifth $\mathbf{k}_1/\mathbf{k}_0 = 3/2$ is the third harmonic lowered by an octave. The major third $\mathbf{k}_1/\mathbf{k}_0 = 5/4$ is the fifth harmonic lowered by two octaves. The fourth $\mathbf{k}_1/\mathbf{k}_0 = 4/3$ is obtained as the difference between the octave and the fifth: $2/\ (3/2) = 4/3$. The minor third $\mathbf{k}_1/\mathbf{k}_0 = 6/5$ is obtained as the difference between the fifth and the major third $(3/2)\ /\ (5/4) = 6/5$. The major sixth $\mathbf{k}_1/\mathbf{k}_0 = 5/3$ is obtained as the sum of the fourth and the major third: $(4/3) \times (5/4) = 5/3$.

The minor sixth $\mathbf{k}_1/\mathbf{k}_0 = 8/5$ is obtained as the sum of the fourth and the minor third: $(4/3) \times (6/5) = 8/5$. The major seventh $\mathbf{k}_1/\mathbf{k}_0 = 15/8$ is obtained as the sum of the fifth and the major third: $(3/2) \times (5/4) = 15/8$. The minor seventh $\mathbf{k}_1/\mathbf{k}_0 = 9/5$ is obtained as the sum of the fifth and the minor third: $(3/2) \times (6/5) = 9/5$. The major tone $\mathbf{k}_1/\mathbf{k}_0 = 9/8$ is obtained as the difference between the fifth and the fourth $(3/2)\ /\ (4/3) = 9/8$. The minor tone $\mathbf{k}_1/\mathbf{k}_0 = 10/9$ is obtained as the difference between the major sixth and the fifth: $(5/3)/(3/2) = 10/9$ (it can be also obtained as the difference between the major third and major tone $5/4\ /\ (9/8) = 10/9$). The diatonic semitone $\mathbf{k}_1/\mathbf{k}_0 = 16/15$ is obtained as the difference between the fourth and the major third: $(4/3)\ /\ (5/4) = 16/15$ (it can be also obtained as the difference between an octave and the major seventh: $2\ /\ (15/8) = 16/15$). The chromatic semitone $\mathbf{k}_1/\mathbf{k}_0 = 25/24$ is obtained as the difference between the major third and the minor third: $(5/4)\ /\ (6/5) = 25/24$.

The degrees of the major scale are: major second (major tone), major third, fourth, fifth, major sixth, and major seventh. The degrees of the minor scale are: major second (major tone), minor third, fourth, fifth, minor sixth, and minor seventh.

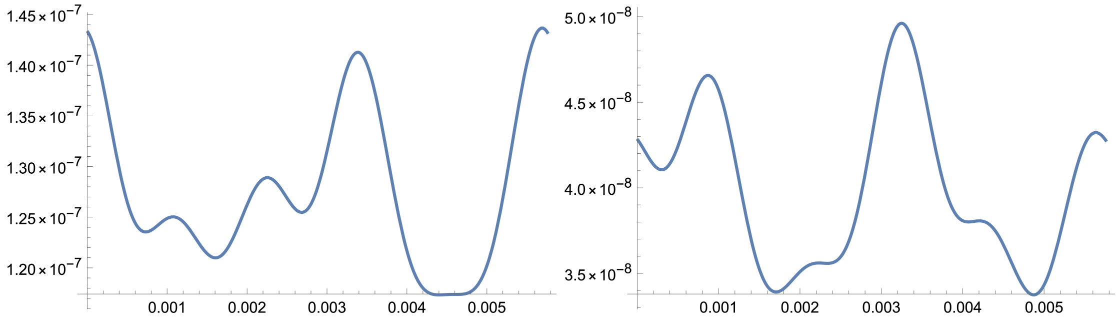


Figure 11: We consider here an input signal $F(t) = \sin(\mathbf{k}_1 t) + \sin(\mathbf{k}_0 t)$ with, as in Case 2 of Section 6.1, $\mathbf{k}_1 = \frac{7}{3}\mathbf{k}_0$. Let us fix $\mathbf{k}_0 = 2\pi 262$ sec$^{-1}$. In this case $E(\xi, t)$ has angular frequency $\frac{2}{3}\mathbf{k}_0$. For $\xi = 2\pi 250 \sim 1570$, we plot $E(\xi, t)$ in one period for the two choices of parameters used previously. Notice that the period of this function is neither the one predicted by Helmholtz $\frac{4}{3}\mathbf{k}_0$ nor the one predicted by Lagrange $\frac{1}{3}\mathbf{k}_0$.

As an example, we are in this case when $\mathbf{k}_0$ and $\mathbf{k}_1$ are two consecutive harmonics of the same root $\bar{\mathbf{k}}$:

$$\mathbf{k}_0 = r\bar{\mathbf{k}}, \quad \mathbf{k}_1 = (r+1)\bar{\mathbf{k}}, \text{ with } r \text{ integer.}$$

In this situation $\mathbf{k}_1 = \frac{r+1}{r}\mathbf{k}_0$, hence $u - w = 1$ and $\frac{1}{w}\mathbf{k}_0$ is the root $\bar{\mathbf{k}}$.

**Case 2.** *Both $u$ and $w$ are odd.* In this case $h = 2$. Then $q = \frac{u-w}{2}$ and

$$\mathbf{k}^H_{\text{combination}} = \frac{u-w}{w}\mathbf{k}_0, \quad \mathbf{k}^L_{\text{combination}} = \frac{1}{w}\mathbf{k}_0, \quad \mathbf{k}_{\text{combination}} = \frac{2}{w}\mathbf{k}_0. \tag{38}$$

These three numbers are all distinct unless $u - w = 2$. For instance taking $u = 7$ and $w = 3$ we have $\mathbf{k}^H_{\text{combination}} = \frac{4}{3}\mathbf{k}_0$, $\mathbf{k}^L_{\text{combination}} = \frac{1}{3}\mathbf{k}_0$, $\mathbf{k}_{\text{combination}} = \frac{2}{3}\mathbf{k}_0$. This situation is described in Figure 11.

# 7 Choice of parameters for numerical simulations

Under (A0), the model depends on several parameters: $\ell$, $L$, $A_\rho, A_T, A_\gamma, k_\rho, k_T, k_\gamma$. Although we are mainly interested in a qualitative analysis we try to use values as close as possible to reality. Unfortunately, to our knowledge, these parameters are not known with great precision and they differ sensibly from one source to another.

Our main references are [MLA24] and [NVTM03]. Let us start with $L$ and $\ell$.

- From [MLA24] we take $L = 35mm$. Notice that in most literature (as for instance in [NVTM03]) $L$ is normalized to 1. We make this normalization in the following as well. Notice that with this choice $k_\rho, k_T, k_\gamma$ become adimensional.
- Concerning $\ell$ we refer to [MLA24], where $\ell = 0.21mm$ at the base $(x = 0)$ and $\ell = 0.36mm$. Since we are working under (A0), we are going to fix $\ell = 0.29$mm$= 2.9 \times 10^{-4}$m independently of $x$.

Concerning the other parameters, we first present (Section 7.1) a set of parameters deduced from [NVTM03] (see Appendix A for the precise comparison between our model and the model of NOBILI). We then slightly modify them to have a more reasonable range of frequencies $[\xi(1)/(2\pi), \xi(0)/(2\pi)]$ and to fit (A03) (see Section 7.2).

| | | | Helmholz | | Lagrange | | our theory | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Interval | $\mathbf{k}_1/\mathbf{k}_0$ | Note corres. to $\mathbf{k}_1$ | $\frac{\mathbf{k}_1-\mathbf{k}_0}{\mathbf{k}_0}$ | Note corres. to $\mathbf{k}_1-\mathbf{k}_0$ | $\frac{\mathrm{GCD}(\mathbf{k}_0,\mathbf{k}_1)}{\mathbf{k}_0}$ | Note corres. to $\mathrm{GCD}(\mathbf{k}_0,\mathbf{k}_1)$ | $\frac{\mathbf{k}_1+\mathbf{k}_0}{\mathbf{k}_1-\mathbf{k}_0}=\frac{p}{q}$ | $q$ | $\frac{\mathbf{k}_1-\mathbf{k}_0}{q\mathbf{k}_0}$ | Note corres. to to $(\mathbf{k}_1-\mathbf{k}_0)/q$ |
| octave | 2 | $C_5$ | 1 | $C_4$ | 1 | $C_4$ | 3 | 1 | 1 | $C_4$ |
| fifth | 3/2 | $G_4$ | 1/2 | $C_3$ | 1/2 | $C_3$ | 5 | 1 | 1/2 | $C_3$ |
| major third | 5/4 | $E_4$ | 1/4 | $C_2$ | 1/4 | $C_2$ | 9 | 1 | 1/4 | $C_2$ |
| fourth | 4/3 | $F_4$ | 1/3 | $F_2$ | 1/3 | $F_2$ | 7 | 1 | 1/3 | $F_2$ |
| minor third | 6/5 | $E\flat_4$ | 1/5 | $A\flat_1$ | 1/5 | $A\flat_1$ | 11 | 1 | 1/5 | $A\flat_1$ |
| major sixth | 5/3 | $A_4$ | 2/3 | $F_3$ | 1/3 | $F_2$ | 4 | 1 | 2/3 | $F_3$ |
| minor sixth | 8/5 | $A\flat_4$ | 3/5 | $E\flat_3$ | 1/5 | $A\flat_1$ | 13/3 | 3 | 1/5 | $A\flat_1$ |
| major seventh | 15/8 | $B_4$ | 7/8 | $B\flat_3$ | 1/8 | $C_1$ | 23/7 | 7 | 1/8 | $C_1$ |
| minor seventh | 9/5 | $B\flat_4$ | 4/5 | $A\flat_3$ | 1/5 | $A\flat_1$ | 7/2 | 2 | 2/5 | $A\flat_2$ |
| major tone | 9/8 | $D_4$ | 1/8 | $C_1$ | 1/8 | $C_1$ | 17 | 1 | 1/8 | $C_1$ |
| minor tone | 10/9 | $D^*_4$ | 1/9 | $B\flat_0$ | 1/9 | $B\flat_0$ | 19 | 1 | 1/9 | $B\flat_0$ |
| diatonic semitone | 16/15 | $D\flat_4$ | 1/15 | $D\flat_0$ | 1/15 | $D\flat_0$ | 31 | 1 | 1/15 | $D\flat_0$ |
| chromatic semitone | 25/24 | $C\sharp_4$ | 1/24 | $F_{-1}$ | 1/24 | $F_{-1}$ | 49 | 1 | 1/24 | $F_{-1}$ |

Figure 12: Tartini's third tone (according to Helmholtz's theory, Lagrange's theory and our approach) within the natural intervals spanning one octave. Here $\mathbf{k}_0$ corresponds to the $C_4$. The $D_4$ is sharper with respect to the $D^*_4$.

### 7.1 Parameters directly deduced from [NVTM03]

From [NVTM03] one deduces (See appendix A) the following values of the parameters

$$\begin{aligned}
A_\rho &= 3.4 \times 10^{-3}\mathrm{Kg/m^2} \\
k_\rho &= 2.9 \\
A_T &= 1.5\ \mathrm{Kg/sec^2} \\
k_T &= 8.5 \\
A_\gamma &= 8.3\ \mathrm{Kg/(m^2\ sec)} \\
k_\gamma &= -1.6
\end{aligned}$$

From which we deduce that

$$\begin{aligned}
\alpha &= 0.78 \\
B_\mu &= 0.16\ \mathrm{sec}^{(\alpha-1)} \\
B_T &= 1.5 \times 10^{-8}\frac{\mathrm{N}}{\mathrm{m}}\mathrm{sec}^{\frac{2k_T}{k_\rho+k_T}} \\
B_\nu &= 8.7 \times 10^{-3}\frac{\mathrm{m}^{\frac{5}{2}}}{\mathrm{Kg}}\mathrm{sec}^{\frac{2k_\rho}{k_\rho+k_T}}
\end{aligned} \tag{39}$$

### 7.2 Modified parameters from [NVTM03] with hypothesis (A03)

Although the parameters given above permit with no difficulties to see the emergence of sub-harmonic series, they do not predict the right range of audible frequencies. Referring to Remark 1, with the values computed above we get

$$\begin{aligned}
\tilde{A} &= 2.2 \times 10^5\mathrm{sec}^{-1} \\
\tilde{k} &= 5.7
\end{aligned}$$

The corresponding frequency range is $[\xi(1)/(2\pi), \xi(0)/(2\pi)] = [120, 35 \times 10^3]$ Hz which a bit shifted toward high frequencies w.r.t. the classical range $[20, 20 \times 10^3]$.

In the following, we propose modified values of $A_\rho, k_\rho, A_T, k_T$ which provide a reasonable range of audible frequencies. See Appendix A for the deatiles on how these modifications have been done.

We also take $k_\gamma = 0$ to fit hypothesis (A03). See Section 2.1. Notice that the presence of higher damping in this case, renders the pics less pronounced. They can be rendered more pronounced by lowering $A_\gamma$ which directly influences $B_\mu$ through (12). Notice that the chosen value of $A_\gamma$ used here produces a bifurcation between the underdamped and the overdamped case (i.e. at $\xi = \mu(\xi)/2 = B_\mu\xi^\alpha/2$) around 18Hz) which is close to the lower threshold of the audible frequencies.

Finally the modified choice of parameters that we propose is the following

$$\begin{aligned}
A_\rho &= 6.9 \times 10^{-3}\ \mathrm{Kg/m^2} \\
k_\rho &= 2.3 \\
A_T &= 0.94\ \mathrm{Kg/sec^2} \\
k_T &= 11.5 \\
A_\gamma &= 8.3\ \mathrm{Kg/(m^2\ sec)} \\
k_\gamma &= 0
\end{aligned}$$

From which we deduce that

$$\begin{aligned}
\alpha &= 0.33 \\
B_\mu &= 24\ \mathrm{sec}^{(\alpha-1)} \\
B_T &= 2.9 \times 10^{-9}\frac{\mathrm{N}}{\mathrm{m}}\mathrm{sec}^{\frac{2k_T}{k_\rho+k_T}} \\
B_\nu &= 44 \times 10^{-3}\frac{\mathrm{m}^{\frac{5}{2}}}{\mathrm{Kg}}\mathrm{sec}^{\frac{2k_\rho}{k_\rho+k_T}}
\end{aligned} \tag{40}$$

With these values we get

$$\tilde{A} = 1.3 \times 10^5 \text{sec}^{-1}$$
$$\tilde{k} = 6.9$$

The corresponding frequency range is $[\xi(1)/(2\pi), \xi(0)/(2\pi)] = [20, 20 \times 10^3]$ Hz.

# A Parameters used in numerical simulations

## A.1 The parameters of [NVTM03]

The values of the parameters $A_\rho, A_T, A_\gamma$ and $k_\rho, k_T, k_\gamma$ are obtained from [NVTM03] by comparing their model with our model.

Before starting let us remark that in equation (7) $\rho$ is the density of mass by unit of area (unit of $x$-length and unit of $z$-length). As a consequence, the equation is written in units of mass/length$^2\times$ acceleration namely in Kg/(m sec$^2$)=N/m$^2$.

Consider our equation (7)

$$\ddot{p}_n = \frac{T}{\rho}\left(-\frac{\pi^2}{\ell^2}n^2\right)p_n - \frac{\gamma}{\rho}\dot{p}_n + \frac{c}{\rho}f_n(t), \quad n = 1, 2, 3, \ldots. \tag{41}$$

Take the first mode $n = 1$ and multiply it by $\ell\rho$

$$\ell\rho\,\ddot{p}_1 = T\left(-\frac{\pi^2}{\ell}\right)p_1 - \ell\gamma\,\dot{p}_1 + c\,\ell f_n(t), \tag{42}$$

and compare it with equation (A1) of [NVTM03]

$$m(x)\ddot{\xi}(x,t) = -k(x)\xi(x,t) - [h(x) - \partial_x s(x)\partial_x]\dot{\xi}(x,t) + \text{external force by unit of length} \tag{43}$$

Notice that in equation (42) $[\ell\rho]$=Kg/m which is the same as $[m(x)]$. Moreover, notice that in [NVTM03] $L$ is normalized to 1, normalization that we make in (42) as well.

Let us fix $\ell = 0.29$mm$= 2.9 \times 10^{-4}$m. Comparing (42)and (43) we get

$$\rho(x) = m(x)/\ell, \quad T(x) = \frac{\ell}{\pi^2}k(x), \quad \gamma(x) = \frac{h(x)}{\ell}$$

Notice that in our model we do not have interaction among the different strings, hence in equation (A1) in [NVTM03], we are assuming $s(x) = 0$.

Now

- In [NVTM03], Fig.2. $m(x)$ varies between $0.1 \times 10^{-5}$ Kg/m (base) to $1.8 \times 10^{-5}$ Kg/m (apex). Hence $\rho$ varies between $3.4 \times 10^{-3}$ Kg/m$^2$ to $62 \times 10^{-3}$ Kg/m$^2$. If we write $\rho(x) = A_\rho e^{k_\rho x}$ (with $x = 0$ at the base and $x = 1$ at the apex) we get
$$A_\rho = 3.4 \times 10^{-3}\text{Kg/m}^2, \quad k_\rho = 2.9$$
- In [NVTM03], Fig.2. $k(x)$ varies between $5 \times 10^4$Kg/(m sec$^2$) (base) to $10^1$ Kg/(m sec$^2$) (apex). Hence $T$ varies between $1.5$ Kg/sec$^2$ (base) to $2.9 \times 10^{-4}$ Kg/sec$^2$. Notice that Kg/sec$^2$ =N/m.

  If we write $T(x) = A_T e^{-k_T x}$ (with $x = 0$ at the base and $x = 1$ at the apex) we get
$$A_T = 1.5\,\text{Kg/sec}^2, \quad k_T = 8.5$$
- In [NVTM03], Fig.2. $h(x)$ varies between $2.4 \times 10^{-3}$ Kg/(m sec ) (base) to $0.5 \times 10^{-3}$ Kg/(m sec ) (apex). Hence $\gamma$ varies between $8.3$ Kg/(m$^2$ sec) (base) to $1.7$ Kg/(m$^2$ sec) (apex).

  If we write $\gamma(x) = A_\gamma e^{k_\gamma x}$ (with $x = 0$ at the base and $x = 1$ at the apex) we get
$$A_\gamma = 8.3\,\text{Kg/(m}^2\text{ sec)}, \quad k_\gamma = -1.6.$$

As a consequence

$$\alpha := \frac{2(k_\rho - k_\gamma)}{k_\rho + k_T} = 0.78,$$

$$B_\mu \underset{\substack{\uparrow \\ \text{under (A0)}}}{=} \frac{A_\gamma}{A_\rho}\left(\sqrt{\frac{A_\rho}{A_T}}\frac{\ell}{\pi}\right)^{\frac{2(k_\rho - k_\gamma)}{k_\rho + k_T}} = 0.16\,\sec^{(\alpha-1)},$$

$$B_T \underset{\substack{\uparrow \\ \text{under (A0)}}}{=} A_T\left(\sqrt{\frac{A_\rho}{A_T}}\frac{\ell}{\pi}\right)^{\frac{2k_T}{k_\rho + k_T}} = 1.5 \times 10^{-8}\,\frac{\mathrm{N}}{\mathrm{m}}\sec^{\frac{2k_T}{k_\rho + k_T}},$$

$$B_\nu = \frac{1}{A_\rho}\left(\sqrt{\frac{A_\rho}{A_T}}\frac{\ell}{\pi}\right)^{\frac{2k_\rho}{k_\rho + k_T}} \frac{2\sqrt{2}}{\pi}\sqrt{\ell} = 8.7 \times 10^{-3}\,\frac{\mathrm{m}^{\frac{5}{2}}}{\mathrm{Kg}}\sec^{\frac{2k_\rho}{k_\rho + k_T}}.$$

**Remark 9.** *Notice that*

- $[B_\mu]$ *are obtained considering that from* (17) $[\mu] = [\xi]$ *=1/sec. Hence from* $\mu = B_\mu \xi^\alpha$ *we get* $[B_\mu]$ *=sec*$^{(\alpha-1)}$.
- $[B_T]$ *are obtained considering that* $T(\xi) = B_T \xi^{\frac{2k_T}{k_\rho + k_T}}$ *and that* $[T]$ *=N/m and* $[\xi]$ *=1/sec which implies that* $[B_T] = \frac{N}{m}(sec)^{\frac{2k_T}{k_\rho + k_T}}$
- $[B_\nu]$ *are obtained considering that* $\nu(x) = \frac{c(x)}{\rho(x)}\frac{2\sqrt{2}}{\pi}\sqrt{\ell(x)}$*. Hence* $[\nu] = \left[\frac{\sqrt{\ell}}{\rho}\right] = \frac{\sqrt{\mathrm{m}}}{\mathrm{Kg/m^2}} = \frac{\mathrm{m}^{\frac{5}{2}}}{\mathrm{Kg}}$*. From* $\nu(\xi) = B_\nu \xi^{\frac{2k_\rho}{k_\rho + k_T}}$ *we get* $[B_\nu] = \frac{[\nu]}{\left[\xi^{\frac{2k_\rho}{k_\rho + k_T}}\right]} = \frac{\mathrm{m}^{\frac{5}{2}}}{\mathrm{Kg}}\sec^{\frac{2k_\rho}{k_\rho + k_T}}$

### A.2 Modified parameters

As explained in Section 7.2 the range of frequencies covered by the model, using the parameters of Appendix A.1 is $[120, 35 \times 10^3]$ Hz which is shifted toward high frequencies w.r.t. the classical range. This could be due to the fact that it is not easy to extract precise values of the parameters from Fig.2 of [NVTM03]. The range of frequencies covered by the model is adjusted by considering that

- $m(x)$ varies between $0.2 \times 10^{-5}$ Kg/m (base) to $2 \times 10^{-5}$ Kg/m (apex).
- $k(x)$ varies between $3.2 \times 10^4$Kg/(m sec$^2$) (base) to .32 Kg/(m sec$^2$) (apex).

Moreover, to fit hypothesis (A03), we consider $h(x) = 2.4 \times 10^{-3}$ Kg/(m sec ) (independently of $x$). We then get the parameters given at the end of Section 7.2.

## B Proof of Proposition 1

$E_n(t)$ has the form

$$E_n(\xi, t) = \Big((a_1 \sin(\mathbf{k}_1 t + \alpha_1) + a_0 \sin(\mathbf{k}_0 t + \alpha_0))^2 + (b_1 \sin(\mathbf{k}_1 t + \beta_1) + b_0 \sin(\mathbf{k}_0 t + \beta_0))^2\Big) \tag{44}$$

where $a_1, a_0, b_1, b_0, \alpha_1, \alpha_0, \beta_1, \beta_0$ are suitable functions depending on $\xi$ and $n$. Notice that the terms in $b_1$ and $b_0$ come from the term $\frac{1}{\xi^2 n^2}\dot{p}_n(t)^2$ in formula (27). Moreover the term $\beta_1$ and $\beta_0$ contains a factor $\pi/2$ coming from the fact that $\frac{d}{dt}\sin(kt) = k\cos(kt) = k\sin(kt + \pi/2)$. Hence we can rewrite $E_n(\xi, t)$ in the form

$$\begin{aligned} E_n(\xi, t) &= f_0(t) + f_1(t) + f_2(t) + g_0(t) + g_1(t) + g_2(t), \text{ where} \\ f_0(t) &= a_0^2 \sin^2(\mathbf{k}_0 t + \alpha_0), \\ f_1(t) &= a_1^2 \sin^2(\mathbf{k}_1 t + \alpha_1), \\ f_2(t) &= 2a_1 a_0 \sin(\mathbf{k}_1 t + \alpha_1)\sin(\mathbf{k}_0 t + \alpha_0) \end{aligned}$$

and similarly for $g_1, g_2, g_3$.

We have the following:

**Claim** Let $h_A$ and $h_B$ be two periodic functions from **R** to **R** with angular frequencies $\omega_A$ and $\omega_B$ and periods $T_A = \frac{2\pi}{\omega_A}$ and $T_B = \frac{2\pi}{\omega_B}$. Assume that $T_A/T_B = m_A/m_B$ with $m_A, m_B$ integers and $m_A/m_B$ an irreducible fraction. Then $h_A + h_B$ is a periodic function with period

$$\frac{2\pi}{\omega_A} m_B = T_A m_B = T_B m_A = \frac{2\pi}{\omega_B} m_A$$

**STEP 1 (periodicity of $f_2$)** Using the formula $2\sin(a)\sin(b) = \cos(a-b) - \cos(a+b)$ we have that $f_2$ is the sum of two periodic functions: one with angular frequency $\mathbf{k}_1 + \mathbf{k}_0$ and the other with angular frequency $\mathbf{k}_1 - \mathbf{k}_0$. Moreover by hypothesis we have that $\frac{\mathbf{k}_1+\mathbf{k}_0}{\mathbf{k}_1-\mathbf{k}_0} = \frac{p}{q}$ with $p > q$ integers and $\frac{p}{q}$ an irreducible fraction. Applying the Claim with $\omega_A = \mathbf{k}_1 + \mathbf{k}_0$, $\omega_B = \mathbf{k}_1 - \mathbf{k}_0$, $T_A/T_B = \frac{\mathbf{k}_1-\mathbf{k}_0}{\mathbf{k}_1+\mathbf{k}_0} = q/p$ and hence $m_A = q$, $m_B = p$ we have that $f_2$ is periodic with period

$$T_2 = T_B m_A = \frac{2\pi}{\mathbf{k}_1 - \mathbf{k}_0} q$$

**STEP 2 (periodicity of $f_0 + f_2$).** Using the formula $\sin^2(a) = \frac{1}{2}(1 - \cos(2a))$ we have that $f_0$ has angular frequency $2\mathbf{k}_0$.

Let us apply the claim with $h_A = f_2,\ h_B = f_0,\ \omega_A = (\mathbf{k}_1 - \mathbf{k}_0)/q,\ \omega_B = 2\mathbf{k}_0$. Notice moreover that

$$\frac{T_A}{T_B} = \frac{\frac{2\pi q}{\mathbf{k}1-\mathbf{k}_0}}{\frac{2\pi}{2\mathbf{k}_0}} = \frac{2\mathbf{k}_0}{\mathbf{k}_1 - \mathbf{k}_0} q = p - q$$

(where in the last equality we have used the fact that $\frac{\mathbf{k}_1+\mathbf{k}_0}{\mathbf{k}_1-\mathbf{k}_0} = \frac{p}{q}$ implies $2\mathbf{k}_0 = \frac{p}{q}(\mathbf{k}_1 - \mathbf{k}_0) - \frac{q}{p}(\mathbf{k}_1 + \mathbf{k}_0)$)). Hence we can take $m_A = p - q$ and $m_B = 1$. Then $f_0 + f_2$ is periodic with period

$$T_A m_B = \frac{2\pi q}{\mathbf{k}_1 - \mathbf{k}_0} 1.$$

With a similar argument we conclude that $f_0 + f_1 + f_2$ is periodic with period $\frac{2\pi q}{\mathbf{k}1-\mathbf{k}_0}$. The same argument applies to $g_0 + g_1 + g_2$ and the conclusion follows.

# C Notation

- $x \in [0, L]$ length parameter along the cochlea, parametrizing the different strings. $L$ is the total length of the cochlea which is normalized to 1 all along the paper. See Section 2.1.
- $z \in [0, \ell(x)]$ length parameter along the string $x$. $\ell(x)$ is the length of the string $x$. See Section 2.1.
- $u^x(t, z)$ the displacement of the string $x$ at position $z$ at time $t$. See Section 2.1 and Figure 5.
- $\rho(x)$, mass of the string $x$ (by unit of $x$ and $z$ length). Under (A0) we use the model $\rho(x) = A_\rho e^{k_\rho x}$. See Section 2.1.
- $T(x)$, tension of the string $x$ (by unit of $x$ length). Under (A0) we use the model $T(x) = A_T e^{-k_T x}$. See Section 2.1.
- $\gamma(x)$, damping of the string $x$. Under (A0) we use the model $\gamma(x) = A_\gamma e^{k_\gamma x}$. See Section 2.1.
- $c(x)$ coupling of the string $x$ with the external signal. Under (A0) we fix $c(x) = 1$. See Section 2.1.
- $F(t, z)$ external force (i.e., the sound). Under (A1) there is no dependence on $z$ i.e. $F(t, z) = F(t)$. See Section 2.1.
- $\xi$ undamped resonance frequency of the different strings. See Formula 2. The different strings are parametrized either with $x$ or with $\xi$.
- $p_n(t)$ $n$-Fourier component of $u^x(t, z)$. See Section 2.2

- $\mu(x) = \frac{\gamma(x)}{\rho(x)}$. See Formula 9.
- $\mu(\xi) = \mu(x(\xi))$ (with abuse of notation).
- $\nu_n(x) = \begin{cases} \frac{c(x)}{\rho(x)} \frac{2\sqrt{2}}{\pi} \sqrt{\ell(x)} \frac{1}{n} & \text{if } n \text{ is odd,} \\ 0 \quad \text{if } n \text{ is even} \end{cases}$, See Formula 10.
- $\nu_n(\xi) = \nu_n(x(\xi))$ (with abuse of notation).
- $\alpha := \frac{2(k_\rho - k_\gamma)}{k_\rho + k_T}$, see Formula (12).
- $B_\mu, B_\nu, B_T$ see Formulas (12),(13), (14).
- $\mathbf{k}$ angular frequency of the external sinusoidal signal. See Formula 19.
- $\mathcal{R}_n(\xi, \mathbf{k})$ asymptotic amplitude of oscillations under the action of a sinusoidal signal. See Formula (22).
- $\phi_n(\xi, \mathbf{k})$ asymptotic phase of oscillations under the action of a sinusoidal signal. See Formula (23).
- $n$ index for sub-harmonics, $j$ index for harmonics. See Formula (2.2.4).
- $E(\xi, t)$ (resp. $E_n(\xi, t)$) energy stored in the string (resp. in the $n$-mode of the string) with undamped resonance frequency $\xi$. See Section 3.

**Acknowledgments**

This work has been supported by CNRS through the MITI interdisciplinary programs and partly supported by the ANR-DFG project "CoRoMo" ANR22-CE92-0077-01. Ugo Boscain is grateful to Prof. Bernardino Streito from whom he learned about the problem of the existence of sub-harmonics at the beginning of the 80s and that pointed out the references of Michelangelo Abbado and of Ettore Dabbene. The authors are grateful to Pascal Martin, Gilles Monfort and Pietro Polotti for very inspiring discussions.

# References

[Abb64] Michelangelo Abbado, *Sull'esistenza dei suoni armonici inferiori*, Acta Musicologica **36** (1964), no. 4, 234–237.

[Abb65] ———, *Presenza e udibilità degli armonici inferiori e conseguente spiegazione del terzo e quarto suono*, Musica d'oggi **8** (1965), no. 3, 76, p. 76 ss.

[Abb72] Michelangelo Abbado, *Terzo e quarto suono*, Rivista italiana di musicologia **V** (1972), 99–147.

[BBWB20] Andres Becher, Gustavo Patricio Bernardi, Thomas S. Wennekers, and Clemens Bruns, *Binaural beats through the auditory pathway: From brainstem to connectivity patterns*, eNeuro **7** (2020), no. 3, ENEURO.0084–20.2020.

[Bék47] Georg V Békésy, *The variation of phase along the basilar membrane with sinusoidal vibrations*, The Journal of the Acoustical Society of America **19** (1947), no. 3, 452–460.

[BPST21] Ugo Boscain, Dario Prandi, Ludovic Sacchelli, and Giuseppina Turco, *A bio-inspired geometric model for sound reconstruction*, Journal of Mathematical Neuroscience **11** (2021), no. 1, 2.

[Cas15] Michèle Castellengo, *Écoute musicale et acoustique*, Éditions Eyrolles, Paris, 2015.

[CMC18] Gabriele Caselli, Giulio Masetti, and Giovanni Cecchi, *Tartini, the third tone and the cochlea*, Musica & Fisiologia – Music & Physiology (2018), Open Access, PH (Milano).

[Des50] René Descartes, *Compendium musicae*, Johannes Maire, Utrecht, 1650.

[DJ03] Thomas Duke and Frank Jülicher, *Active traveling wave in the cochlea*, Physical review letters **90** (2003), no. 15, 158101.

[DP73] Ettore Dabbene and Raffaele Pisani, *Generazione di frequenze subarmoniche negli strumenti musicali*, Zanibon, December 1973.

[DP22] Gianluca Dai Prà, *A genealogic-structural hypotesis of harmonic dualism*, Nauchnyy vestnik Moskovskoy konservatorii / Journal of Moscow Conservatory (2022) (English), Independent researcher; published in Moscow Conservatory journal.

[Dui11] Hendrikus Duifhuis, *Modeling the nonlinear cochlea*, Cochlear Mechanics: Introduction to a Time Domain Analysis of the Nonlinear Cochlea, Springer, 2011, pp. 95–146.

[ES12] Stephen J Elliott and Christopher A Shera, *The cochlea as a smart structure*, Smart Materials and Structures **21** (2012), no. 6, 064001.

[Hal88] Jack K. Hale, *Asymptotic behavior of dissipative systems*, Mathematical Surveys and Monographs, vol. 25, American Mathematical Society, Providence, RI, 1988.

[Hau53] Moritz Hauptmann, *Die natur der harmonik und der metrik: Zur theorie der musik*, Breitkopf und Härtel, Leipzig, 1853 (German).

[Hel63] Hermann von Helmholtz, *Die lehre von den tonempfindungen als physiologische grundlage für die theorie der musik*, Friedrich Vieweg und Sohn, Braunschweig, 1863 (German).

[Hul24] Henri Ver Hulst, *Marrying the physics of critical oscillators with traveling-wave models of the cochlea*, Ph.D. thesis, Université Paris Sciences et Lettres, Paris, France, December 2024, Thèse de doctorat en Physique.

[Lag59] Joseph-Louis Lagrange, *Recherches sur la nature et la propagation du son*, Miscellanea Taurinensia (Mélanges de Turin) **1** (1759), 1–112 (French), Reprinted in Œuvres de Lagrange, tome 1, pp. 39–148 (Paris: Gauthier-Villars, 1867).

[Lag62] ———, *Nouvelles recherches sur la nature et la propagation du son*, Miscellanea Taurinensia (Mélanges de Turin) **2** (1762), 11–172 (French), Reprinted in Œuvres de Lagrange, tome 1, pp. 149–316 (Paris: Gauthier-Villars, 1867).

[Lev85] Ernst Levy, *A theory of harmony*, SUNY Series in Cultural Perspectives, State University of New York Press, Albany, NY, 1985.

[Lew82] David B. Lewin, *A formal theory of generalized tonal functions*, Journal of Music Theory **26** (1982), no. 1, 23–60.

[MH99] Pascal Martin and AJ Hudspeth, *Active hair-bundle movements can amplify a hair cell's response to oscillatory mechanical stimuli*, Proceedings of the National Academy of Sciences **96** (1999), no. 25, 14306–14311.

[MLA24] Jahangir Moini, Anthony LoGalbo, and Raheleh Ahangari, *Chapter 9 - the auditory system*, Foundations of the Mind, Brain, and Behavioral Relationships (Jahangir Moini, Anthony LoGalbo, and Raheleh Ahangari, eds.), Academic Press, 2024, pp. 143–160.

[Moo12] Brian C.J. Moore, *An introduction to the psychology of hearing*, 6th ed., Brill, Leiden, The Netherlands, 2012.

[NEAT14] Guangjian Ni, Stephen J Elliott, Mohammad Ayat, and Paul D Teal, *Modelling cochlear mechanics*, BioMed research international **2014** (2014), no. 1, 150637.

[NVTM03] Renato Nobili, Aleš Vetešnik, Lorenzo Turicchia, and Fabio Mammano, *Otoacoustic emissions from residual oscillations of the cochlear basilar membrane in a human ear model*, Journal of the Association for Research in Otolaryngology **4** (2003), no. 4, 478–494.

[Pam89] Richard Pamcutt, *Harmony: A psychoacoustical approach*, Springer Series in lnf011Jlation Sciences [5) RD Patterson, BCJ Moore:(1986)" Auditory filters and excitation patterns...", In BCJ Moore: Frequency selectivity in hearing, London: Academic Press [6) John Pierce:(1966)," Attaining consonance in arbitrary scales", J. Acoustical Society. Am **40** (1989), no. 7, 1061–1069.

[Ram22] Jean-Philippe Rameau, *Traité de l'harmonie réduite à ses principes naturels*, Jean-Baptiste-Christophe Ballard, Paris, 1722.

[RH14] Tobias Reichenbach and AJ Hudspeth, *The physics of hearing: fluid mechanics and the active process of the inner ear*, Reports on Progress in Physics **77** (2014), no. 7, 076601.

[Rie93] Hugo Riemann, *Vereinfachte harmonielehre oder die lehre von den tonalen funktionen der akkorde*, Breitkopf & Härtel, 1893.

[RR01] Luis Robles and Mario A. Ruggero, *Mechanics of the mammalian cochlea*, Physiological Reviews **81** (2001), no. 3, 1305–1352.

[SCH15] Robert Szalai, Alan R Champneys, and Martin Homer, *Nonlinear dynamics of the mammalian inner ear*, arXiv preprint arXiv:1505.04667 (2015).

[Ste12] Benjamin Steege, *Helmholtz and the modern listener*, Cambridge University Press, Cambridge and New York, 2012.

[Tar54] Giuseppe Tartini, *Trattato di musica secondo la vera scienza dell'armonia*, Stamperia del Seminario (Giovanni Manfré), Padova, 1754, Opera teorica in cui Tartini discute del fenomeno del "terzo suono".

[Tar67] ———, *De' principj dell'armonia musicale contenuta nel diatonico genere: Dissertazione*, Stamperia del Seminario, Padova, 1767, Estende e chiarisce le idee armoniche, incluso il terzo suono.

[vO66] Arthur von Oettingen, *Harmoniesystem in dualer entwickelung*, H. Laakmann, Dorpat, 1866 (German).

[VW60] Bekesy George Von and EG Wever, *Experiments in hearing*, 1960.

[Wie59] Robert W. Wienpahl, *Zarlino, the senario, and tonality*, Journal of the American Musicological Society **12** (1959), no. 1, 27–41.

[You39] Robert W. Young, *Terminology for logarithmic frequency units*, Journal of the Acoustical Society of America **11** (1939), no. 1, 134–139.

[Zar58] Gioseffo Zarlino, *Le istitutioni harmoniche*, Francesco Senese, Venezia, 1558 (Italian).

[ZF99] Eberhard Zwicker and Hugo Fastl, *Psychoacoustics: Facts and models*, 2nd ed., Springer, Berlin, Germany, 1999.

[ZLP76] George Zweig, R Lipes, and JR Pierce, *The cochlear compromise*, The Journal of the Acoustical Society of America **59** (1976), no. 4, 975–982.