ICONEA 2012-2015, XXX-XXX

EXPRESSION OF THE CONSTITUTION OF ARISTOXENUS' MUSICAL SCALES IN TERMS OF MATHEMATICAL FUNCTIONS

Gilles de Rosny[1]

Our present knowledge of musical theories of Ancient Grece is based on an attentive scan of texts written by Greek musicographers. These texts are available in several manuscripts which sometimes differ in significant manners; sometimes only a portion of the work has survived, which is the case with Aristoxenus. Their study has been regularly renewed since the nineteenth century. Among the most recent scholars who have undertaken this task, a special mention must be made of Professor Andrew Barker for his translations and subtle analysis of a large set of musicographers, including Aristoxenus[2]. The present contribution proposes an interpretation, within the framework of the mathematical theory of functions, of some of the arguments developed by Aristoxenus to establish his scales of notes.

Before going to detailed argumentation, let us summarize briefly how the musical scales of the Ancient Greeks are formed: the basic blocks are the concord of the fourth and the concord of the fifth between two pitches. They add up to form the concord of the octave. The fifth is split into a fourth plus a tone. The fourth is split in various ways into three parts so that it contains a set a four notes that form a tetrachord. Then the octave scale is made of two tetrachords plus a tone; it contains eight notes. When the two tetrachords are conjunct they share a note, the tone being added at one extremity; when the tetrachords are disjunct, the tone is inserted between the two tetrachords. The diversity of Ancient Greek scales is largely due to the variety of ways by which the fourth is split into three parts.

Let us focus our attention on the *Elementa Harmonica* of Aristoxenus[3] to see how he defines a note and an interval. "To put it briefly, a note is the incidence of the voice on one pitch: for it is when the voice appears to rest at one pitch that there seems to be a note capable of being put in a position in a harmonically attuned melody [*melos hērmosmenon*]. That, then, is the sort of thing a note is.

An interval [*diastēma*] is that which is bound by two notes which do not have the same pitch, since an interval appears, roughly speaking, to be a difference between pitches, and a space capable of receiving notes higher than the lower of the pitches which bind it, and lower than the higher of them. Difference between pitches lies in their having been subjected to greater or lower tension" (20.20 B136)[4].

Although Aristoxenus does not assign any numerical value to a pitch, the unique way to identify it being the ear, the definitions given above assume that the space of notes is 'totally ordered', i.e. that the pitches of two notes, whatever the sources are (voice, string, wind or percussion), can be recognized as being equal, or the first note being higher than the other one or the reverse. This procedure of classification may be extended to an arbitrary number of notes. If one indicates notes by italic letters, and the property "*higher than*" by the symbol $>$, "*lower than*" by $<$, and "*of the same pitch*" by =, the ordering properties of pitches may be summarized as follows: two notes *a* and *b* are either at the same pitch: $a=b$, or *a* is higher than *b*: $a>b$, or a is lower than *b*: $a<b$, equivalently *b* is higher than *a*: $b>a$. Let *a* and *b* be two notes such as $a<b$, a third note *c* belongs to the interval limited by *a* and *b* if and only if $a \leq c \leq b$. More precisely, a note represents all the sounds sharing a common pitch, which is called the equivalence class of these sounds.

About the intervals: "The first division of intervals is in respect of their difference in magnitude" (21.17 B137).

The magnitudes of two intervals may be compared if for instance the extreme notes of the first interval are included in the second interval, one can then state that the magnitude of the first interval is smaller than the magnitude of the second one.

Let us assume that there are eight magnitudes of concords. The smallest is the fourth: that it is the smallest is determined by the nature of melody itself. This is shown by the fact that we sing many intervals smaller than the fourth, but all of them are discordant. The second is the fifth: every magnitude which there may be between these two will be discordant. The third is the sum of the two concords mentioned, the octave, all magnitudes between the fifth and the octave being discordant..." (55.12-56.8 B159-160).

One may interpret this excerpt as follows:

1- Starting from a note, if one produces a second note whose pitch is progressively increased, a first concord with the starting note, called the fourth, is encountered at some precise pitch; if one continues to increase the pitch, a second concord, the fifth is encountered, and next a third, the octave.

2- If one tunes the note higher than a first note by a fourth and then the note higher than the second one by a fifth, it is observed that the first and the third notes are in concord of the octave.

3- If one reverses the order by tuning first a fifth and then a fourth, one gets the same note at the octave. This property is not explicitly stated in what remains of the Aristoxenus' treatise, but we assume that it is true.

One may formalize the above properties by introducing functions acting on the set of notes in the following way.

Let us express that a note b is a fourth higher than a note a by the relation $b=\mathbf{f4}(a)$, where **f4** is the 'fourth function'.

One defines in the same way the fifth function **f5**, where $c=\mathbf{f5}(a)$ is a fifth higher than a. It is observed (see 1-above) that $b<c$, i.e. $\mathbf{f4}(a)<\mathbf{f5}(a)$ whatever a, which can be written $\mathbf{f4}<\mathbf{f5}$.

The concord an octave higher, represented by the octave function **f8** is such that $\mathbf{f4}<\mathbf{f5}<\mathbf{f8}$ (see 1-above).

Let now be $b=\mathbf{f4}(a)$ and $c=\mathbf{f5}(b)$; c is a function of a through b: $c=\mathbf{f5}(\mathbf{f4}(a))$. The successive applications of **f4** and **f5** is called their composition[5], noted $\mathbf{f5}\circ\mathbf{f4}$, then $c=\mathbf{f5}\circ\mathbf{f4}(a)$. c and a are in concord of the octave (see 2-above): $c=\mathbf{f8}(a)=\mathbf{f5}\circ\mathbf{f4}(a)$; so that $\mathbf{f8}=\mathbf{f5}\circ\mathbf{f4}$.

The same note c is reached if one proceeds by tuning first a fifth $d=\mathbf{f5}(a)$ and then a fourth (see 3-above): $c=\mathbf{f4}(d)=\mathbf{f4}\circ\mathbf{f5}(a)$. The property $\mathbf{f5}\circ\mathbf{f4}(a)=\mathbf{f4}\circ\mathbf{f5}(a)$ is expressed by saying that the functions **f4** and **f5** commute. The octave function **f8** is related to **f5** and **f4** by $\mathbf{f8}=\mathbf{f5}\circ\mathbf{f4}=\mathbf{f4}\circ\mathbf{f5}$.

$b=\mathbf{f4}(a)$ means that b and a are in concord of a fourth and that b is higher than a. One defines the inverse function of **f4**, noted $\mathbf{f4}^{-1}$, which links b to a: $a=\mathbf{f4}^{-1}(b)$, expressing that a and b are in concord of a fourth and that a is lower than b. The functions **f4** and $\mathbf{f4}^{-1}$ are called inverse of the other because their successive applications lead back to the initial note: $b=\mathbf{f4}(a)\Rightarrow a=\mathbf{f4}^{-1}(b)=\mathbf{f4}^{-1}\circ\mathbf{f4}(a)$. The function that does not move a note is called the identity, noted **I**: $\mathbf{f4}^{-1}\circ\mathbf{f4}=\mathbf{f4}\circ\mathbf{f4}^{-1}=\mathbf{I}$. One defines similarly inverse functions of **f5** and **f8**.

Next is the definition of the tone: "The tone is that by which the fifth is greater than the fourth; the fourth is two and half tones" (57.1, B160).

Let us consider the first part of the sentence. In terms of functions it may be expressed as follows: let be $b=\mathbf{f4}(a)$ and $c=\mathbf{f5}(a)$, the resulting relation between c and b is by definition expressed by the tone function **t**, such that $c=\mathbf{t}(b)=\mathbf{t}\circ\mathrm{f4}(a)=\mathrm{f5}(a)$. Then $\mathbf{t}\circ\mathbf{f4}=\mathbf{f5}$, or $=\mathbf{f5}\circ\mathrm{f4}^{-1}$, i.e. to tune a note one tone higher than a note, one tunes a note a fourth downwards and from this note, a note a fifth upwards. One may also tune in the reverse order since the functions **f5** and $\mathbf{f4}^{-1}$ commute[6]. That the fifth is greater than the fourth means that $c>b$, or $\mathbf{t}(b)>b$, or $\mathbf{t}>\mathbf{I}$. The second part of the sentence appears as a property of the fourth, which will be discussed later.

More on concords: "Of the magnitude of intervals, those of the concords appear to have either no range of variation [*topos*] at all, being determined to a single magnitude, or else a range which is quite indiscernible, whereas those of the discords possess this quality to a much smaller degree. Hence perception relies much more confidently on the magnitude of concords than on those of discords. The most accurate way of constructing a discordant interval will therefore be by means of concords. Thus, if we have the task of constructing from a given note a discord such as a ditone downwards (or any of those which can be constructed by means of concords) we should construct from the given note [a][7] a fourth upwards [$b=\mathbf{f4}(a)$], from there a fifth downwards [$c=\mathbf{f5}^{-1}(b)$], then another fourth upwards [$d=\mathbf{f4}(c)$], and then another fifth downwards [$e=\mathbf{f5}^{-1}(d)$]. In this way the ditone downwards from the given note will have been constructed [$e=\mathbf{dt}^{-1}(a)$, where $\mathbf{dt}$ is the ditone function]. And if the task is to construct the discord in the opposite direction, the concords should be constructed the opposite way round." (68.10-69.6 B168).

In terms of functions, Aristoxenus states that $\mathbf{dt}^{-1}=\mathbf{f5}^{-1}\circ\mathbf{f4}\circ\mathbf{f5}^{-1}\circ\mathbf{f4}$, which may be written $\mathbf{dt}^{-1}=\mathbf{t}^{-1}\circ\mathbf{t}^{-1}$, since $\mathbf{t}=\mathbf{f5}\circ\mathbf{f4}^{-1}$. The composition $\mathbf{t}^{-1}\circ\mathbf{t}^{-1}$ may be written more concisely $\mathbf{t}^{-2}$ using a 'functional power' notation; then $\mathbf{dt}^{-1}=\mathbf{t}^{-2}$. In the opposite direction we write $\mathbf{dt}=\mathbf{t}^{2}$.

Aristoxenus continues: "Further if a discord is subtracted from a concordant interval by means of concords, the remainder will also have been found by means of concords. For instance let a ditone be subtracted by means of concords from a fourth. It is clear that the notes bounding the remainder by which the fourth exceeds the ditone have also been constructed, by means of concords, in their relation to one another." (69.6 B168)

In terms of functions the remainder is expressed as $\mathbf{r}=\mathbf{f4}\circ\mathbf{dt}^{-1}=\mathbf{f4}^{3}\circ\mathbf{f5}^{-2}$, where $\mathbf{r}$ is the remainder function. Aristoxenus states that the fourth exceeds the ditone, which we write $\mathbf{f4}>\mathbf{dt}$, or $\mathbf{dt}\circ\mathbf{r}>\mathbf{dt}$ i.e. $\mathbf{r}>\mathbf{I}$. The last inequality means that for any note a, $\mathbf{r}(a)>a$.

Then: "The question whether we were correct in the assumption, which we made in our introduction, that the fourth consists of two and a half tones, can be investigated most accurately in the following way. Take a fourth [a, $b=\mathbf{f4}(a)$], and starting from each of its extremes in turn, mark off a ditone by means of concords [$c=\mathbf{dt}(a)$, $d=\mathbf{dt}^{-1}(b)$]. It is clear that the remainders are equal, since equals have been taken from equals[8] [$d=\mathbf{dt}^{-1}\circ\mathbf{f4}(a)=\mathbf{r}(a)$, $c=\mathbf{dt}\circ\mathbf{f4}^{-1}(b)=\mathbf{r}^{-1}(b)$].

Next take a fourth upwards from the lower bounding note of the lower ditone [$e=\mathbf{f4}(d)$], and take another fourth downwards from the upper bounding note of the lower ditone [$\int=\mathbf{f4}^{-1}(c)$]. It is clear that next to each of the notes which bound the resulting *systēma* there will be two consecutive remainders which must be equal, for the reason stated before[$e=\mathbf{f4}(d)=\mathbf{f4}\circ\mathbf{dt}^{-1}(b)=\mathbf{r}(b)$,$\int=\mathbf{f4}^{-1}(c)=\mathbf{f4}^{-1}\circ\mathrm{dt}(a)=\mathbf{r}^{-1}(a)$]. When this construction has been set up, we must bring the judgement of perception of the outermost of the notes that have been located. If they appear to perception as discordant, it will be evident that the fourth is not two and a half tones: but if they sound the concord of the fifth [if $e=\mathbf{f5}(\int)$], it will be evident that the fourth is two and a half tones [$e=\mathbf{r}(b)=\mathbf{r}\circ\mathbf{f4}(a)$, $a=\mathbf{r}(\int)$ $\Rightarrow$ $e=\mathbf{r}^{2}\circ\mathbf{f4}(\int)$, if $\mathbf{r}^{2}\circ\mathbf{f4}=\mathbf{f5}=\mathbf{t}\circ\mathbf{f4}$, then $\mathbf{r}^{2}=\mathbf{t}$, which may be expressed by saying that the remainder is half a tone or a semitone]. For the lowest of the notes constructed [$\int$] was tuned to make the concord of a fourth with the upper bounding note of the lower ditone [c], and the highest of the notes constructed [e] has turned out to form the concord of the fifth with the lowest [$\int$]; so that since the difference is a tone, and since it is divided into equal parts, each of which is a semitone [$\mathbf{r}$] and is also the excess of a fourth over a ditone, it is clear that the fourth consists of five semitones [$\mathbf{f4}=\mathbf{t}^{2}\circ\mathbf{r}=\mathbf{r}^{5}$]. It is easy to see that the extremes of the systēma constructed will not form any concord but the fifth. In the first place it must be understood that they do not form the concord of the fourth, since there is an excess added in both directions to the fourth which was originally taken. Next we must state that it cannot accommodate the concord of an octave. The magnitude formed by the sum of the remainders is less than a ditone [$\mathbf{r}^{2}<\mathbf{dt}$], since the fourth exceeds the ditone by less than a tone [$\mathbf{dt}<\mathbf{f4}<\mathbf{dt}\circ\mathbf{t}$]: for everyone agrees that the fourth is greater than two tones and smaller than three [$\mathbf{t}^{2}<\mathbf{f4}<\mathbf{t}^{3} \Leftrightarrow \mathbf{dt}<\mathbf{dt}\circ\mathbf{r}<\mathbf{dt}\circ\mathbf{t} \Leftrightarrow \mathbf{I}<\mathbf{r}<\mathbf{t} \Leftrightarrow \mathbf{I}<\mathbf{r}^{2}<\mathbf{dt}$]. Thus the total added to the fourth is less than a fifth [$\mathbf{r}^{2}<\mathbf{dt}<\mathbf{f5}$], which makes it clear that that the combination of them cannot be an octave [$\mathbf{f4}\circ\mathbf{r}^2<\mathbf{f4}\circ\mathbf{f5}=\mathbf{f8}$]. But if the extremes of the note constructed form a concord greater than the fourth but smaller than the octave, the concord which they form must be the fifth: for this is the only concordant magnitude between the fourth and the octave." (70.3-72.6 B169).

The notes *e* and *f* are linked by the relation $e=\mathbf{r2}\circ\mathbf{f4}(f)$ established above; since it has also been established that $\mathbf{r}=\mathbf{f4}^3\circ\mathbf{f5}^{-2}$, then $e=\mathrm{f4}^7\circ\mathbf{f5}^{-4}(f)$. This last relation expresses that to get the note *e* from the note *f*, one has to tune in any order seven fourth upwards and four fifth downwards, totalling eleven tunings of concords. Each tuning needs to be performed with a precision allowing checking if indeed *e* and *f* are in concord of a fifth.

The notes that may be taken by means of concords, i.e. by successive applications of fourths and fifths from a given note are linked to it by functions of the form $\mathbf{c}_{m,n}=\mathbf{f4}^{\mathbf{m}}\circ\mathbf{f5}^{\mathbf{n}}$, where 'm' and 'n' are any positive, negative, or null integers. The set of all $\mathbf{c}_{m,n}$ constitutes a commutative or Abelian group[9] , generated by **f4** and **f5**. In fact, only a subset of these notes can actually be tuned since each intermediate note resulting from the application of a fourth or a fifth needs to be emitted by the voice or by an instrument and also to be heard.

One may express the fourth and the fifth in terms of the semitone **r**: we have already seen that $\mathbf{f4}=\mathbf{r}^5$, then $\mathbf{f5}=\mathbf{f4}\circ\mathbf{t}=\mathbf{r}^7$. Then $\mathbf{c}_{m,n}=\mathbf{r5}^{\mathbf{m+7n}}$; from this writing it may be shown that the notes that can be tuned by concords from a given note are obtained upwards by successive applications of the semitone r, and downwards by successive applications of $\mathbf{r}^{-1}$. The other notes have to be tuned by other means. The octave is made of twelve semitones or six tones: $\mathbf{f8}=\mathbf{f4}\circ\mathbf{f5}=\mathbf{r}^{12}=\mathbf{t}^6$.

A summary of the partitions of the fourth gathered in various parts of the *Elementa Harmonica* is presented in Table 1. The note at the lowest pitch is the *hypatē*, it is followed by two mobile notes the *parhypatē* and the *lichanos* the positions of which, expressed in terms of fractional numbers of semitones, may vary in definite zones, the last note, the *mesē*, is located five semitones higher than the *hypatē*. The two extreme notes are fixed. According to the regions where the mobile notes are located, the tetrachords are separated in three genera: enharmonic, chromatic, and diatonic. Moreover, six usual species of tetrachords have been distinguished by Aristoxenus[10].

Within the context of the functions that have been introduced, the fractional number of semitones characterizing the position of a mobile note relative to the *hypatē* is to be understood as a fractional functional power of the semitone function **r**: let p/q be such a fraction, where p and q are positive integers, let us introduce the function rp/q such that, when applied to the *hypatē*, the result is the desired note that we write *p/q: p/q*=**rp/q** (*hypatē*). To give a meaning to the fractional functional power, we postulate that **rp/q** composed q times is equal to **rp**. Only the notes at positions characterised by integer numbers can be tuned by means of concords. In table 1, the notes of the six usual tetrachords that cannot be tuned by means of concord from the *hypatē* are surrounded by a circle.

In order to express the pitches in terms of sound frequencies, one establishes a mapping of the set of functions acting on the Aristoxenus' notes into the set of multipliers acting on the frequencies, namely the octave operator **f8** is mapped on the factor **2** that links the fundamental frequencies of two notes one octave apart. The mapping is such that it preserves the composition laws of each set (homomorphism property); in particular since **f8**=**t6**, the multiplier $\mathbf{t^{f}}$ that relates the frequencies of two notes one tone apart satisfies the relation $\mathbf{2}=\mathbf{t^{f}6}$, so that $\mathbf{t^{f}}=\mathbf{21/6}$, which is the tone of the equal temperament scale. The multiplier associated to the function **rp/q** is **2p/12q**.[11]

In table 2 are presented the sound frequencies, rounded to 1Hz, of octave scales made of two identical disjunct tetrachords for the six species quoted by Aristoxenus. Only the frequencies of the mobile notes, written in italics, differ from one scale to the other. The underlined frequencies are those of the notes that cannot be tuned by means of concords. One expects that these notes would not be tuned accurately by ear only. The tense diatonic scale, whose tetrachord mobile notes *parhypatē* and *lichanos* are respectively at 1 and 3 semitones from the *hypatē* is actually identical to the equal temperament diatonic scale: the lowest note E4 is chosen in order to allow a labelling of the scale that avoids sharps or flats.

The method of calculation of these frequencies is detailed as follows for the enharmonic scale: one sets the value of the lowest frequency f1 to 329.63Hz (E4), the next frequency is $f_2=2^{1/24*}f_1$, since in the enharmonic tetrachord, the *parhypatē* is half of a semitone higher than the *hypatē*. Next $f_3=2^{1/12*}f_1$, and $f_4=2^{5/12*}f_1$, then the disjunctive tone, $f_5=2^{1/6*}f_4$, and finally the upper tetrachord, $f_6=2^{1/24*}f_5$, $f_7=2^{1/12*}f_5$, and $f_8=2^{5/12*}f_5$.

To conclude, it has been shown that the functions introduced to describe the relations between pairs of notes constitute an efficient and faithful interpretation of the intervals used by Aristoxenus to establish his scales. The arithmetic of Aristoxenus about sizes and combinations of intervals is well understood as arithmetic of functional powers; although this concept was not available to him, he anticipated its might.

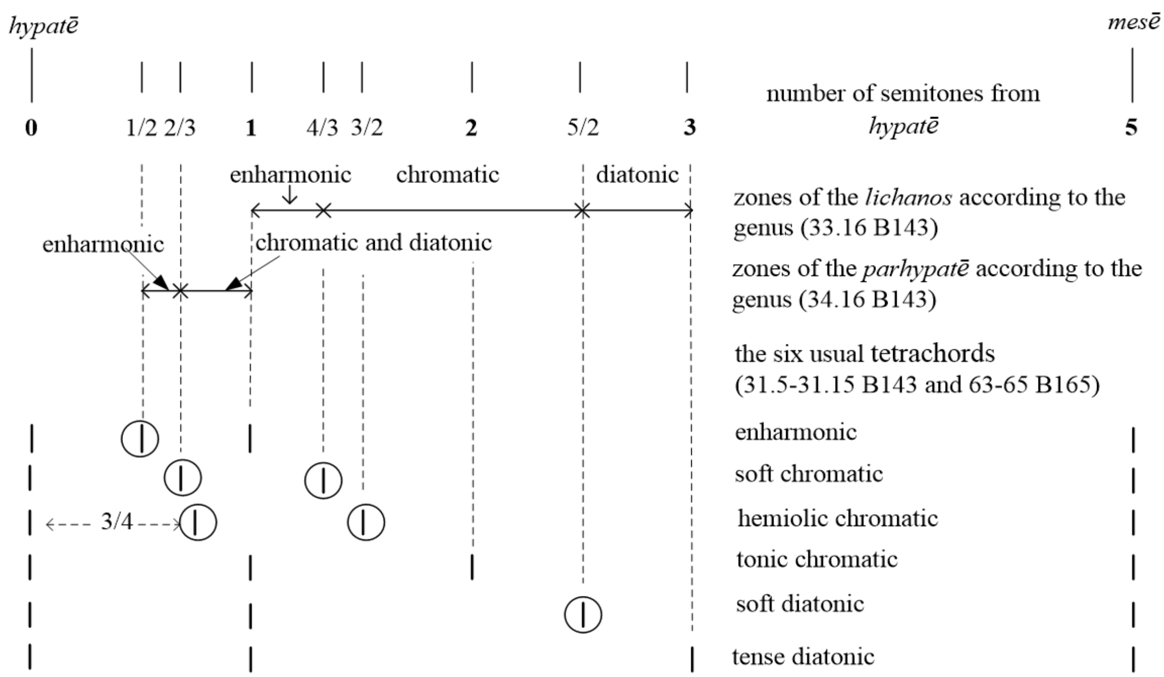

Table 1. Positions of the mobile notes parhypatē and lichanos

	enharmonic	chromatic			diatonic		
		soft	hemiolic	tonic	soft	tense	
f_8	659	659	659	659	659	659	E5
f_7	*523*	*533*	*539*	*554*	*570*	*587*	D5
f_6	*508*	*513*	*516*	*523*	*523*	*523*	C5
f_5	494	494	494	494	494	494	B4
f_4	440	440	440	440	440	440	A4
f_3	*349*	*356*	*359*	*370*	*381*	*392*	G4
f_2	*339*	*342*	*344*	*349*	*349*	*349*	F4
f_1	330	330	330	330	330	330	E4

Octave scales corresponding to the six usual tetrachords

Table 2.

Notes

1 gilles.derosny@gmail.com
2 A. BARKER, Greek Musical Writings (GMW), Cambridge University Press, 1989
3 R. DA RIOS editor, Rome 1954. A. BARKER, GMW, tome II for the translation
4 Page.line in Da Rios, followed by page in GMW t II.
5 http://en.wikipedia.org/wiki/Function_composition. The property of associativity of function composition is often tacitly used in the following.
6 It can be shown that if **f4** and **f5** commute $\mathbf{f4}^{-1}$ and **f5** also commute.
7 The terms between square brackets are added by me.
8 Aristoxenus states that similar intervals (fourths, fifths, remainders) bounded by different couples of notes are equal. In terms of functions, this is expressed by the fact that the notes bounding equal intervals are related the one to the other by the same function.
9 See for instance http://en.wikipedia.org/wiki/Abelian_group
10 Aristoxenus expresses the interval widths in various units from various origins. In table 1 they have been uniformly expressed as numbers of semitones relatively to the *hypatē*.
11 Other mappings are possible, for instance: one may require that the fourth function is mapped on the 'perfect fourth' multiplier 4/3. Then the tone function is mapped on $(4/3)^{2/5}$, and the octave function on $(4/3)^{12/5} \cong 1.995$

This presentation would not have emerged without the support of my friends of the working group on Ancient Greek Music in Paris: Philippe Brunet, François Cam, Jean Boutan, Einar Már Jónsson, Yves Lenoble, and Eliane Poulvet. Let them be acknowledged.